Isthmus progenitor cells contribute to homeostatic cellular turnover and support regeneration following intestinal injury

Malagola E.[1,#], Vasciaveo A.[2,#], Ochiai Y.[1], Kim W.[1], Middelhoff M.[3], Nienhüser H.[4], Deng L.[5], Belin B.[1], LaBella J.[1], Zamechek LB.[1], Wong M.H.[6], Li L.[5], Guha C.[7], Cheng CW.[8], Yan KS.[1,8], Califano A.[2,9,10,11,12], and Wang T.C.[1,12].

[1] Division of Digestive and Liver Diseases, Department of Medicine and Irving Cancer Research Center, Columbia University Medical Center, New York, NY 10032, USA.

[2] Department of Systems Biology, Columbia University, New York, NY 10032, USA.

[3] Klinik und Poliklinik für Innere Medizin II, Klinikum rechts der Isar, Technische Universität München, Munich, Germany.

[4] Department of General, Visceral and Transplant Surgery University Hospital Heidelberg Im Neuenheimer Feld 420 69120, Heidelberg, Germany.

[5] Stowers Institute for Medical Research, Kansas City, MO 64110, USA; Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66107, USA.

[6] Department of Cell, Developmental & Cancer Biology, Oregon Health & Sciences University 3181 SW Sam Jackson Park Road, L215, Portland, USA.

[7] Department of Radiation Oncology, Albert Einstein Cancer Center, Albert Einstein College of Medicine,1300 Morris Park Avenue, Bronx, NY 10461, USA.

[8] Columbia Stem Cell Initiative, Department of Genetics and Development, Columbia University Irving Medical Center, New York, New York.

[9] Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY 10032, USA.

[10] Department of Biochemistry & Molecular Biophysics, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY 10032, USA.

[11] Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, NY 10032, USA.

[12] Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY 10032, USA.

[#] These authors contributed equally

Correspondence to: A. Califano (ac2248@cumc.columbia.edu) and Timothy C. Wang (tcw21@cumc.columbia.edu)

Summary

The currently accepted intestinal epithelial cell organization model proposes that crypt base columnar (CBC) cells marked by high levels of *Lgr5* expression represent the sole intestinal stem cell (ISC) compartment. However, previous studies have indicated that Lgr5^{+} cells are dispensable for intestinal regeneration, leading to two major hypotheses: one favoring the presence of a quiescent reserve stem cell population, the other calling for differentiated cell plasticity. To investigate these possibilities, we studied crypt epithelial cell organization, during homeostasis and regeneration, in unbiased fashion, via high-resolution single-cell profiling. These studies, combined with *in vivo* lineage tracing, show that *Lgr5* is not a specific ISC marker and that stemness potential exists beyond the crypt base in the isthmus region, whose cells, contrary to differentiated cells, participate in tissue homeostasis and support intestinal regeneration. Our results provide a novel model of organization for the intestinal crypt epithelium in which stemness potential is not restricted to CBC cells and suggesting that neither de-differentiation nor reserve stem cell populations are drivers of intestinal regeneration.

Introduction

The intestinal epithelium is characterized by a high cellular turnover rate, making it an attractive model to study adult stem cell biology. The intestinal stem cell (ISC) is defined by its ability to self-renew and to give rise to all intestinal epithelial lineages. Lineage tracing studies have uncovered a multitude of genes expressed in ISCs[1-7]; among others, *Lgr5* has been widely accepted as the primary ISC marker[8] due to its proposed high degree of specificity. In particular, *Lgr5* was reported to be selectively expressed in crypt base columnar (CBC) cells, located at the very bottom of the intestinal crypts, between Paneth cells, typically at positions +1 to +3[8]. However, recent studies on intestinal regeneration following high doses of radiation[5,6,8], or selective *Lgr5*-expressing cell ablation[9], have challenged this view by showing that other crypt epithelial cells can compensate for the loss of Lgr5$^+$ cells. These observations suggest that the 'Lgr5-CBC' model may not fully represent the regenerative ability of the intestinal epithelium. To account for these findings, two major hypotheses have emerged: the first proposing that, in addition to the Lgr5$^+$ ISCs, a quiescent '+4' stem cell population may act as a 'reserve' to replenish the ISC pool following intestinal injury[5,7,9], the second asserting that, following ISC loss, multiple differentiated cell types may undergo de-differentiation and serve as *bona fide* ISCs[10-13].

Single cell transcriptomics has emerged as a powerful tool to study tissue heterogeneity and to uncover the identity and dynamics of rare cell populations[14,15]. With respect to the intestine, this approach has been instrumental in detailing the specific differentiation states of diverse intestinal lineages, as well as to investigate the cellular composition of the regenerating intestine[7,16-18]. However, previous studies and descriptions of the intestinal epithelium hierarchical organization have relied heavily on established markers, thus forcing an interpretation compatible with prior ISC models and limiting the generation of alternative hypotheses[19].

Recently, a number of studies have shown that computational cell potency inference is effective in reconstructing cell hierarchies within tissues[20-23]. These novel approaches can infer the

41 hierarchical organization of intestinal epithelial cells *de novo*, in unbiased fashion, *i.e.*, without
42 having to rely on prior knowledge. Such an approach, coupled with relevant experimental assays,
43 can thus help assess whether the current ISC models are consistent with actual tissue
44 organization. Furthermore, these methodologies allow the investigation of putative changes in
45 intestinal epithelial cell potency following intestinal injury, as well as the possible emergence of a
46 reserve stem cell population.

47 To address these questions, we studied intestinal crypt epithelial cells by combining single cell
48 profiling with novel lineage tracing mouse and then analyzed their organization using an unbiased,
49 data-driven approach that does not rely on previously established markers. Our data show that
50 *Lgr5* overlaps with but is not uniquely expressed in stem and progenitor cells. More critically, they
51 indicate that stemness potential exists beyond the CBC cells' compartment in the isthmus region.
52 Based on these results, we propose a novel model of intestinal organization in which additional
53 actively cycling cells have the potential to act as the multipotent, self-renewing ISCs and can
54 contribute to homeostatic cellular turnover alongside CBC cells. Following intestinal injury,
55 surviving isthmus cells guide intestinal regeneration, identifying them as the postulated reserve
56 ISCs, and suggesting that de-differentiation constitutes an exceptional event rather than the main
57 driver of intestinal regeneration.

Results

An unbiased approach to elucidate the organization of intestinal crypt epithelial cells: To study intestinal crypt epithelial cell composition, we developed an integrative, experimental and computational pipeline designed to recover the transcriptional identity of individual intestinal cells, as well as their hierarchical relationships by transcriptomic and epigenetic single cell profile analysis (Fig. 1a). A high level of enrichment of crypt epithelial cells was achieved via nearly complete villus epithelial cell exclusion, using a villi-cell specific antibody (B6A6[24]) (Extended Data Fig. 1a).

To avoid critical gene dropout issues common to single cell RNA-seq (scRNA-seq) profile analyses—where, on average, ≥ 80% of the genes are undetectable in each individual cell[25]—we relied on VIPER, an established network-based algorithm designed to measure the transcriptional activity of regulatory and signaling proteins from scRNA-seq profiles[26]. Specifically, akin to highly multiplexed gene reporter assay, VIPER measures a protein's activity based on the differential expression of its 50 highest-confidence transcriptional targets (*regulon*), as identified by the ARACNe algorithm[27,28]. Critically, we have shown that this methodology allows identification of subpopulations that are essentially undetectable by gene expression analysis[14,15], as well as robust, highly reproducible quantification of protein activity, which compares favorably with antibody-based single cell measurements[14], as well as removal of technical artifacts and batch effects[29].

To identify intestinal stem and progenitor cells, we combined analyses of crypt epithelial cell chromatin accessibility via single-cell Assay for Transposase-Accessible Chromatin scATAC-seq[30] and cell potency inference via scRNA-seq analysis[20,22,23] (*See Methods*). Three major subpopulations emerged initially from cluster analyses of >3,500 crypt epithelial cells, corresponding to two established intestinal lineages, secretory (*Hepacam2*^{hi}, *Tff3*^{hi}, etc.) and absorptive (*Ckmt1*^{hi}, *Ccl25*^{hi}, etc.) cells, with a third subpopulation mostly comprised of stem and

83 progenitor cells (*Stmn1*^{hi}, *Dek*^{hi}, etc.)[16,18,19] (Extended Data Fig. 1b-f, Extended Data Table 3-4).
84 Importantly, inference of cell potency, using independent algorithms[20,21,23], showed a high degree
85 of similarity and correlated well with the degree of chromatin accessibility observed, supporting
86 the effectiveness of these methodologies.

87 To further reveal the lineage-specific nature of intestinal sub-populations, we performed
88 independent sub-clustering of these three major populations (Fig. 1b-c, Extended Data Fig. 1g,
89 Extended Data Table 5-6). This refined analysis revealed two absorptive clusters, corresponding
90 to subpopulations representing early specified and more committed cells, respectively, as well as
91 three clusters corresponding to established secretory subpopulations, including tuft (*Dclk1*+),
92 enteroendocrine (*NeuroD1*+), and goblet (*Atoh1*+) cells[19]. Notably, consistent with previous single
93 cell, VIPER-based, studies[14,15], the vast majority of subpopulation-specific proteins—including
94 established intestinal lineage markers such as Fev[31], and Gfi1b[19]—could only be detected by
95 VIPER-based protein activity analysis but not based on the differential expression of their
96 encoding genes (Extended Data Fig. 1h). Due to their relative small number, *Lyz1*^{hi} Paneth cells
97 did not form an independent cluster but rather were intermixed within the goblet cluster (Extended
98 Data Fig. 2a-b). To overcome this limitation, we further identified Paneth cells based on
99 expression of their established markers[16] thus allowing for their inclusion in further analyses.
100 Atoh1 and Spdef, both previously established regulators of Paneth cell's identity, were among the
101 most activated proteins in these cells, explaining their initial co-segregation within the goblet cells
102 cluster.

103 Finally, stem and progenitor cells co-segregated within two clusters corresponding to
104 subpopulations presenting similar levels of inferred potency, suggesting that this subdivision does
105 not reflect a hierarchical partition of these cells (Fig. 1d, Extended Data Fig. 2c-e). Rather, the
106 highest level of stemness potential was detected at the boundary between these two clusters
107 (Extended Data Fig. 2f), suggesting that these reflect a gradual yet progressive differentiation of

108 progenitor cells towards a secretory (Sec-Progenitor) or absorptive (Abs-Progenitor) fate,
109 respectively. Consistent with these findings, discrimination of candidate ISCs from proliferative
110 progenitors, by joint CytoTRACE and RNA content analysis, as previously described[20], indicated
111 that these candidate were well included in the top decile of inferred cell potency corresponding to
112 virtually all boundary cells (Extended Data Fig. 2g). Analogous results were obtained when
113 analyzing previously published datasets of crypt epithelial cells[7,17], which however presented a
114 lesser degree of enrichment for stem and progenitor cells, corroborating the advantage of our
115 isolation strategy (Extended Data Fig. 2h-i). Furthermore, lineage-specific enrichment analyses
116 of crypt epithelial cells further highlighted the observed bias of the progenitor populations towards
117 either a secretory or absorptive fate, respectively (Extended Data Fig. 2j).

118 Notably, these analyses provide a unique opportunity to study the molecular features that best
119 describe stemness within the intestine. To this aim, we characterized the regulatory factors with
120 the most statistically significant association with inferred cell potency, by correlating the protein
121 activity profiles, as assessed by VIPER, with inferred potency, as assessed by CytoTRACE[20], on
122 a cell by cell basis (Fig. 1e, Extended Data Fig. 3, Extended Data Table 7). Regulatory proteins,
123 whose activity presented the most significant correlation with CytoTRACE scores, were highly
124 enriched for factors known to regulate stemness across different tissues, including the intestine[32-
125 38]. Consistently, known regulators of intestinal differentiation[19] had negative correlation scores.

126 Among others, the transcription factor Yy1, was previously reported to play an essential role in
127 preventing ISC differentiation and to impair epithelial cells organoids forming capacities upon
128 genetic knock-out[36]. Several chromatin remodeler enzymes also emerged from the analysis—
129 including Smarca5[33] and Atad2[34], which have been previously associated with stemness—thus
130 highlighting the crucial role of chromatin remodeling in cell potency. Several splicing factors also
131 emerged within the broad group of transcriptional regulators associated with cell potency,

132 including members of the Srsf family, which were recently shown to play an important role in
133 regulating the intestinal epithelium[37].

134 Focusing on genes whose differential expression, rather than activity, correlated with inferred cell
135 potency (Extended Data Table 8), we identified *Smarca5*, *Stmn1* and *Dek*, all representing
136 previously proposed markers of adult stem cells[32,39]. Birc5a[40] and Ncl[41], both described as highly
137 expressed in embryonic stem cells, also emerged among the most correlated genes. Finally,
138 *Fgfbp1* was also positively correlated with inferred cell potency (*Capdevila C et al., accompanying*
139 *paper*). These analyses provide an unbiased characterization of intestinal crypt epithelial cells
140 using mechanism-based dissection of protein activity in single-cell data, via regulatory network
141 analysis, and highlight novel factors that best recapitulate intestinal cell potency.

142 ***Lgr5 expression is not restricted to intestinal stem and progenitor cells:*** Lgr5 has been
143 described as a highly specific ISC marker[8]. Based on Lgr5 reporter allele and CreER lineage
144 tracing assays, Lgr5^{+} cells were first postulated to correspond to crypt base columnar (CBC) cells
145 in the +1 to +3 region[8]. Surprisingly, however, our analyses reveal no statistically significant
146 correlation between *Lgr5* expression and inferred cell potency (Extended Data Fig. 4a-b).
147 Specifically, cells presenting the highest *Lgr5* expression levels were not associated with highest
148 stemness potential; rather, the highest potency was detected in some Lgr5^{low} cells (Fig. 2a,
149 Extended Data Fig. 4c).

150 To better characterize the *Lgr5* expression pattern, we generated additional scRNA-seq profiles
151 from highly purified crypt epithelial cells harvested from an Lgr5$^{DTR\text{-}eGFP}$ mouse[9] (Fig. 2b). By
152 mapping the *DTR-eGFP* allele on sequencing data, we were able to study expression profiles of
153 both wild type (Wt) and transgenic (Tg) Lgr5 alleles in purified crypt epithelial cells (Fig. 2c).
154 Quantification of the number of cells expressing Wt or Tg alleles revealed virtually complete
155 overlap, confirming that nearly all *Lgr5* expressing cells were also positive for *DTR-eGFP* mRNA

156 (Fig. 2c). Intriguingly, *Lgr5* expression was detected in about 20% of all sequenced crypt cells in
157 our datasets (20.495±1.8, % of total), suggesting a much broader domain of expression then
158 previously appreciated[8]. Flow cytometry analysis of the crypt epithelium consistently revealed an
159 average of 20% Lgr5^{eGFP+} cells (19.96±3.96 % of Epcam^{+} cells), thus confirming the faithful activity
160 of the reporter allele, which was not confined to the CBC cells population (Fig. 2d).

161 We next analyzed the identity of *Lgr5* expressing cells within our datasets. Similar to other
162 proposed ISC-specific markers[1-4,6], Lgr5 expression was broadly detected in multiple intestinal
163 cell types (Extended Data Fig. 4d). Most *Lgr5^{hi}* cells were detected in the Sec-Progenitor
164 subpopulation (Extended Data Fig. 4e). In line with this, *Lgr5* expression was found to be
165 associated with the secretory lineage (Fig. 2e, Extended Data Fig. 4f). Moreover, we detected a
166 small fraction of *Lgr5*-expressing cells that were also positive for known secretory markers, such
167 as *Atoh1* and *Dll1*, in agreement with previous studies that have observed direct specification of
168 secretory cells from Lgr5^{+} ones[17,19] (Extended Data Fig. 4g-h). Surprisingly, we also observed
169 high levels of *Lgr5* expression in differentiated tuft cells marked by *Dclk1* expression[19,42]
170 (Extended Data Fig. 4e,g). Dclk1 staining validated its overlap with Lgr5^{eGFP+} cells, both within the
171 crypt as well as in the villus region (39.2±9.5 % of Dclk1^{+} cells; Fig. 2f**,** Extended Data Fig. 4i).
172 Moreover, sorted Dclk1$^{ZsGreen+}$ cells[43] confirmed expression of *Lgr5* and *Ascl2*[44], both proposed
173 ISC markers, in differentiated tuft cells (Extended Data Fig. 4j-k). Taken together, these data show
174 that *Lgr5* expression is not restricted to intestinal stem and progenitor cells but also spans
175 differentiated secretory subpopulations, including tuft cells.

176 ***Stemness potential exists beyond Lgr5^{+} cells:*** Lgr5$^{eGFP-CreERT}$ mice show clear labelling of a
177 subset of crypt epithelial cells exhibiting long term lineage tracing, consistent with an ISC[8].
178 Nevertheless, our analyses indicate that *Lgr5* expression *per se* does not distinguish cells with
179 potency from differentiated ones; rather stemness potential characterize broadly stem and
180 progenitors cells, with highest levels of cell potency detected in some Lgr5^{low} cells (Fig. 2a). In

181 order to investigate the ISC potential of cells other than Lgr5^{+} CBC ones, we studied their capacity
182 to generate organoids *in vitro*[45]. Lgr5^{neg} cells could form organoids (Extended Data Fig. 5a) and
183 enrichment for Ki67$^{RFP+;}$Lgr5$^{DTReGFP-(neg)}$ proliferative cells[46] increased their efficiency (Fig. 3a);
184 thus, in agreement with our analysis, stemness potential is not restricted to *Lgr5* expressing cells.

185 To clarify if organoid forming capacity reflects functional stemness or a common feature of
186 intestinal epithelial cells, by means of de-differentiation, we employed multiple engineered mouse
187 models that allow discriminating stem and progenitor cells from differentiated cells and evaluated
188 their capacity to generate organoids *in vitro*. To enrich for progenitors[47] we generated a novel
189 mouse model (Lgr4^{CreERT}) that targets inducible Cre to *Lgr4* expressing cells. Consistent with our
190 expectation, following TAM induction, a majority of labelled cells were actively cycling and
191 localized to the crypt isthmus (+4/+12 positions), where proliferation is highest (Fig. 3b, Extended
192 Data Fig. 5b). Similarly, the Ki67^{CreERT} mouse model[48] confirmed efficient labelling of cycling cells
193 within the crypts and presented a similar spatial distribution (Extended Data Fig. 4b). Over time,
194 both models gave rise to tracing ribbons, which persisted for up to six months, thus demonstrating
195 labelling of some ISCs (Fig. 3c, Extended Data Fig. 5c-d). Conversely, to selectively label
196 differentiated cells we generated a novel mouse model that targeted inducible Cre to *Dll1*-
197 expressing cells (Dll1^{CreERT}). Dll1^{CreERT} mice efficiently label committed secretory cells that overlap
198 only minimally with proliferative progenitors and do not give rise to long term lineage tracing (Fig.
199 3b-c). Moreover, we studied our previously described Dclk1$^{-ZsGreen-DTR}$ mouse line that efficiently
200 labels tuft cells within the intestine (Extended Data Fig. 4j-k)[43]. Notably, as previously reported,
201 differentiated tuft cells do not expand and the majority are lost within one week from labelling
202 (Extended Data Fig. 5e)[49]. In agreement with cell potency inference (Fig. 1d), only stem/progenitor
203 cells could form organoids *in vitro* (Fig. 3d), suggesting that organoids forming capacity reflects
204 stemness competence rather than plasticity, and that exposure to growth factors alone is not
205 sufficient to promote epithelial cells de-differentiation.

206 These results show that cells other than Lgr5^{+} CBC cells recapitulate functional stemness
207 properties *in vitro.* Intriguingly, recent intravital studies have reported downward migration of cells
208 at the base[50], suggesting that positional advantage does not constitute a limiting factor for niche
209 competition[51]. Together, these and our observations raise the possibility that cells outside the
210 base may contribute to normal tissue turnover. To determine if isthmus cells can exert downward
211 migration and give rise to CBC cells, we sought to carefully analyze the tracing behavior of
212 Lgr4^{CreERT} and Ki67^{CreERT} mouse lines. Albeit not specific to the crypt isthmus, these mouse models
213 show minimal labelling of the CBC cells' compartment while including cells capable of long-term
214 lineage tracing (Extended Data Fig. 5b,e), making them an attractive model to test this hypothesis.
215 Quantification of positional labelling over time indicated bi-directional clonal expansion of traced
216 cells from the crypt isthmus, with some invading the CBC cells compartment (Fig. 3e-f, Extended
217 Data Fig. 5f-g). This was further supported by Cre-independent analysis of proliferative cells'
218 migration (Extended Data Fig. 5h-j). Additionally, analysis of the overlap between labelled cells
219 and Lgr5^{DTReGFP} indicated enrichment for Lgr5^{low} cells in the double positive –
220 tdTomato^{+};Lgr5$^{DTReGFP+}$ - population (Fig. 3g, Extended Data Fig. 5k), supporting the isthmus
221 origin of the expanded clones.

222 Altogether these analyses indicate that stemness potential is not restricted to *Lgr5*-expressing
223 cells but rather covers broadly the stem and progenitor compartments. Furthermore, they provide
224 evidence that isthmus cells can serve as an ISC compartment and contribute to cellular turnover
225 during intestinal homeostasis.

226 ***Lgr5^{neg} isthmus cells compensate for loss of Lgr5 expressing cells:*** Our data indicate that
227 alongside Lgr5^{+} cells, other crypt epithelial cells retain functional stem cell properties. This is
228 consistent with multiple studies showing that selective ablation of *Lgr5* expressing cells does not
229 impact normal tissue turnover. In particular, two models of interpretation have been proposed to

230 explain such observations, whereby Lgr5^{+} cells repopulation is driven by either (a) de-
231 differentiation of terminally differentiated cells[10-13], or (b) activation of a reserve quiescent ISC[7,9].

232 To clarify the dynamics of this process we analyzed the diphtheria toxin (DT) ablation model using
233 the Lgr5^{DTReGFP} mouse line[9]. Flow cytometry analysis confirmed consistent, complete elimination
234 of Lgr5$^{DTReGFP+}$ cells after two doses of DT without compromising the intestinal mucosa; this was
235 followed by the later reappearance of *Lgr5* expressing cells, returning to control values within 10
236 days (Fig. 4a, Extended Data Fig. 6a). Of note, the disappearance of Lgr5^{+} cells coincided with a
237 marked reduction of Dclk1^{+} tuft cells (Extended Data Fig. 6b), corroborating their expression of
238 *Lgr5* and aligning well with previous reports[52].

239 To gain additional insights into the response to Lgr5^{+} cell ablation, we profiled the transcriptome
240 of purified crypt epithelial cells at 24h after two doses of DT; this represents a time point when
241 most Lgr5^{+} cells are ablated (% of Lgr5-eGFP^{+} cells: 0.44±0.85 vs 20.52±1.94 - Fig. 4b) and
242 regeneration has begun. ScRNA-seq profile analysis revealed that, despite marked reduction in
243 *Lgr5* expressing cells, the overall cellular composition remained intact (Extended Data Fig. 6c-d),
244 consistent with the observation that Lgr5^{+} cells are largely dispensable for crypt epithelium
245 regeneration[9]. Importantly, these analyses did not reveal expansion of any novel cell populations
246 that may be ascribed to activation of a reserve ISC. Analysis of inferred cell potency of DT-treated
247 crypt epithelial cells, relative to control, revealed dramatic reduction of cells that comprised the
248 top decile of high potency scored cells (Fig. 4c, Extended Data Fig. 6e), in line with the notion that
249 *Lgr5* is expressed in some stem and progenitor cells. Nevertheless, a consistent fraction of high
250 inferred stemness cells persisted at this time point, further supporting the hypothesis that some
251 Lgr5^{neg} cells possess the potential to serve as ISCs to support tissue regeneration.

252 Similar to control, following DT treatment high cell potency was confined to stem and progenitor
253 cells, suggesting that ISC potential is also restricted to these populations during regeneration. To

254 test this hypothesis, we first analyzed the behavior of Lgr4^{CreERT} labelled cells following DT
255 treatment. Surviving tdTomato^{+}Lgr5$^{DTReGFP-}$ cells could be observed scattered across the crypt
256 epithelium and over time expanded from the isthmus giving rise to new tdTomato^{+}Lgr5$^{DTReGFP+}$
257 cells at day 10 (Fig. 4d-e, Extended Data Fig. 6f-g). Similar findings were obtained when exposing
258 mice to prolonged DT treatment, which revealed consistent lineage tracing of Lgr5^{neg} proliferating
259 cells (Extended Data Fig. 6h). Notably, mice had to be sacrificed by day 9 due to toxicity of DT
260 that was independent of Lgr5^{DTReGFP} expression (Extended Data Fig. 6i).

261 Next, we tested whether differentiated cells may also participate in Lgr5^{+} cell repopulation
262 alongside proliferating cells, thus reflecting the proposed broad plasticity of intestinal epithelial
263 cells[11]. Upon crosses to Lgr5^{DTReGFP} mice and DT treatment, Dll1^{CreERT} labelled cells did not
264 expand but rather appeared as Paneth cells at day 10, suggesting that committed and/or
265 differentiated secretory cells do not contribute to Lgr5^{+} cell repopulation (Fig. 4f, Extended Data
266 Fig. 7a-b). A possible explanation for the discrepancy with previous reports[10,12] is the higher
267 specificity of our newly generated mouse line for committed secretory cells relative to the
268 previously reported Dll1^{eGFPCreERT}, which showed lineage tracing events also in homeostatic
269 conditions[12]. Indeed, low *Dll1* expression can be detected broadly within the stem/progenitor
270 compartment, and this appears true also for other proposed markers of differentiation including
271 *Alpi*, *Fabp2*, and *Krt20* (Extended Data Fig. 7c-d).

272 Taken together, these data show that, upon loss of *Lgr5* expressing cells, Lgr5^{neg} isthmus cells
273 have the capacity to act as ISCs and to support tissue turnover. Furthermore, they indicate that
274 regenerative potential is restricted to stem and progenitor cells, and that neither de-differentiation
275 nor activation of a reserve quiescent-ISC are the source of Lgr5^{+} cell repopulation in this model
276 (Fig. 4g).

277 ***Surviving isthmus cells regenerate the intestinal epithelium following IR damage:*** As
278 discussed, analysis of CTRL and DT-ablated crypt epithelial cells revealed neither the presence
279 of a putative reserve ISC population nor any sign of active de-differentiation. Rather, it supports
280 a model where some Lgr5^{neg} cells retain ISC potential and sustain intestinal tissue turnover. To
281 determine whether this may be applicable to other intestinal injury models, we assessed whether
282 intestinal regeneration after lethal irradiation (IR) follows similar cellular dynamics. Damage due
283 to high dose IR exposure can induce the gastrointestinal syndrome, with highly proliferative cells
284 and *Lgr5*$^{+}$ CBCs representing subpopulations more likely to be lost due to irreversible IR damage[8].

285 To pinpoint the identity of cells with regenerative potential, we studied crypt epithelial cell
286 composition after exposure to 12 Gy whole body IR (WBI), by scRNA-seq analysis. First, to
287 identify the earliest stage of intestinal regeneration, we characterized the dynamics of intestinal
288 proliferation following IR damage (Extended Data Fig. 8a). We postulated that at this time point,
289 regenerating stem cells would be undertaking their first or second round of cell divisions, marked
290 by an increase in Ki67 labeling. We identified 60 hours post IR as the earliest time point when
291 proliferation increases, reflecting the first regenerative wave and the focal point for our analysis
292 (Fig. 5a).

293 Following IR damage, surviving cells aligned well with previously identified clusters and no novel
294 subpopulations emerged (Fig. 5a). As expected[8], proliferative progenitors as well as *Lgr5*
295 expressing cells were largely depleted at this time point (Extended Data Fig. 8b). Similar to the
296 ablation model, surviving differentiated cells retained low levels of inferred cell potency (Extended
297 Data Fig. 8c), and evaluation of Dll1^{CreERT} or Dclk1^{CreERT} -driven lineage tracing corroborated the
298 absence of any participation by differentiated secretory cells in intestinal regeneration (Extended
299 Data Fig. 8d).

300 In line with our hypothesis, we observed a small fraction of surviving cells that were characterized
301 by high levels of inferred cell potency and expression of proliferative markers (Fig. 5b), thus
302 identifying them as the most suitable regenerating ISCs candidates. Regenerating and
303 homeostatic cells with high cell potency showed a high degree of overlap (Extended Data Fig.
304 8b), suggesting that regenerating ISCs may correspond to surviving isthmus cells. To clarify the
305 nature of the regenerating ISCs, we lineage traced intestinal proliferating cells following high
306 doses of IR. Sixty hours after IR exposure, surviving tdTomato^{+} cells (Lgr4^{CreERT}) could be
307 observed within the damaged epithelium (Fig. 5c, Extended Data Fig. 8e), thereby demonstrating
308 that not all proliferative cells are lost to IR induced damage. Moreover, Ki67 staining and lineage
309 tracing analyses at 5 days post IR confirmed active participation of surviving isthmus proliferative
310 tdTomato^{+} cells in intestinal regeneration (Fig. 5c-d).

311 We next sought to identify the regulatory programs that characterize regenerating ISCs. For this
312 purpose, we searched for features unique to cells with highest inferred cell potency pre- and post-
313 IR (Fig. 5e-f, Extended Data Fig. 8f, Extended Data Table 9-10). Pathway analysis showed that
314 regenerating ISCs activate cell cycle damage checkpoints and upregulate multiple damage
315 response factors[53-55], reflecting the profound damage generated by IR exposure. This was further
316 corroborated by analysis of the regulatory proteins that define the regenerating ISC compartment
317 (Fig. 5f). Indeed, Apex1 and Baz1a, both known to be involved in DNA damage response[56,57],
318 were among the most statistically significant activated proteins. Moreover, regenerating ISCs
319 expressed high levels of *Ly6a* (Sca1), as well as *Areg* and other proposed intestinal regeneration
320 markers[58] (Fig. 5e). Interestingly, we observed upregulation of *Clu* (Fig. 5e, Extended Data Fig.
321 8g), whose expression has been proposed as specific to a quiescent radio-resistant stem cell
322 population ("revival ISC"[7]).

323 These results raised the possibility that regenerating ISCs may represent an expanding reserve
324 stem cell subpopulation. To clarify the nature of these cells, we analyzed *Clu* expression in traced

325 intestinal proliferating cells (Ki67^{CreERT}) pre- and post- IR damage. Results showed *Clu*
326 upregulation in surviving tdTomato^{+} cells sixty hours post IR exposure (Extended Data Fig. 8h),
327 providing evidence that following IR damage, the proposed 'revival' ISCs correspond to surviving
328 cycling cells rather than a separate quiescent stem cell population. Lastly, although we were
329 unable to discriminate the relative contribution of pre-existing stem cells or early progenitors, our
330 analyses strongly suggest that the potency to serve as regenerating ISCs is restricted to these
331 populations (Fig. 5G).

332 Discussion

333 Here we provide an unbiased characterization of intestinal crypt epithelial cells in homeostasis
334 and regeneration. We find that stemness potential exists beyond the CBC cells compartment in
335 the crypt isthmus, the region previously proposed to accommodate transit-amplifying progenitors.
336 Furthermore, we provide evidence that intestinal regeneration is driven by surviving stem and
337 progenitor cells.

338 Our optimized isolation protocol—using the B6A6 Ab[24] to exclude villi cells—ensured high
339 enrichment of the crypt epithelium and reproducibility across all high-throughput datasets,
340 especially for the injured intestine where loss of tissue integrity greatly affects the isolation
341 strategy. Furthermore, our computational pipeline integrated multiple algorithms to unbiasedly
342 recover cell identities and elucidate their hierarchical organization based on single cell
343 transcriptomic and epigenetic profiles. Compared to previous studies, the results obtained show
344 a high degree of similarity[16,17] yet provide improved molecular resolution of intestinal stem and
345 progenitor cells. Moreover, the proposed regulatory network-based classification schema offers
346 novel insights on mechanistic determinants of intestinal epithelial cell subpopulations. Intriguingly,
347 many of the identified stem cell regulatory factors align well with previous reports[19], confirming
348 their findings and suggesting that such signatures may be useful to study the role of novel factors
349 in defining specific intestinal lineages.

350 Surprisingly, *Lgr5* expression did not correlate with inferred cell potency. *Lgr5*^{hi} cells were largely
351 found in the clusters associated with secretory-biased progenitor cells, as previously suggested
352 for some intestinal Lgr5^{+} cells[17,59], and in line with findings in the stomach, where high levels of
353 *Lgr5* can be detected in differentiated chief[60] and antral basal secretory cells[61].

354 Moreover, Lgr5 expression *per se* does not discriminate between cells with potency from
355 differentiated ones. Stemness potential is found well outside the CBC cell compartment, and
356 isthmus cells can participate in normal tissue homeostasis. Importantly, these observations
357 resonate well with our and others previous reports[1,4-6] and highlight similarities with the gastric
358 mucosa where contribution of cells from the isthmus region in tissue cellular turnover has been
359 described[62]. Thus, suggesting that proposed competition dynamics[51] may involve broadly CBC
360 and isthmus cells compartments.

361 Cells with the highest inferred cell potency were well segregated from Lgr5^{hi} cells and detected at
362 the boundary between the progenitors clusters (Sec-Progenitor and Abs-Progenitor). This
363 suggest a distinction between highest inferred stemness potential and proposed relative
364 contributions to long term lineage tracing. While some of the applied computational methodologies
365 may suffer of intrinsic biases[20], this may indicate that selective lineage engagement biases
366 intestinal epithelial cells for long term retention.

367 Furthermore, the difficulty in segregating cells with highest potency from early progenitors
368 suggests their high-degree of similarity and may indicate their co-existence in closely comparable
369 cellular states. Analysis of the factors that best associate with inferred cell potency revealed
370 multiple chromatin remodelers, including Atad2, which has been recently shown to determine cell
371 potency in the skin[34], as well as other factors previously suggested to regulate intestinal
372 stemness[36,37]. Whereas, similarly to Lgr5, other proposed markers for ISC, such as Ascl2[63], also

373 failed to correlate with inferred cell potency and were rather broadly expressed within the
374 secretory lineage.

375 In addition to characterizing homeostatic crypt cell composition, we also analyzed intestinal
376 regeneration following Lgr5^{+} cell ablation and IR damage. In line with previous observations[7,9],
377 both models mostly eliminated *Lgr5* expressing cells without evoking major composition changes
378 within the intestine, especially for the DT ablation model. Notably, our results indicate that cells
379 with high inferred cell potency persist within the intestine and align well with the proposed
380 homeostatic isthmus compartment. Furthermore, our findings highlight that differentiated cell
381 types retain low cell potency states in both injury models and do not appear to acquire features
382 of stemness. In fact, differentiated secretory cells marked by high levels of *Dll1* or *Dclk1*
383 expression do not participate in intestinal regeneration, whereas surviving isthmus cells do. While
384 the mouse models studied do not allow for the discrimination of the relative contribution between
385 surviving ISCs or early progenitors, they provide evidence that the potential to regenerate is
386 restricted to these populations. In addition, when we analyzed the changes in the transcriptional
387 profile of ISCs following IR damage, we detected a specific signature of regenerating cells[58]. This
388 included upregulation of *Ly6a* (Sca-1) and *Areg* for example, and was accompanied by increased
389 activation of several damage response factors, such as Baz1a and Apex1[56-58,64]. Intriguingly, we
390 noted upregulation of *Clu*, an additional target of YAP signaling[65], indicating that previously
391 proposed 'revival' stem cells[7] correspond to surviving ISCs and early progenitors rather than a
392 distinct quiescent stem cell population.

393 In summary, our unbiased analyses suggest a new regulatory model whereby isthmus cells
394 possess functional stem cells properties and participate both in homeostasis and regeneration.
395 Previous findings suggesting the presence of a reserve stem cell or extensive cellular plasticity
396 likely reflected overlap of inducible Cre driver with the isthmus compartment. Our finding of broad
397 expression by intestinal progenitors of multiple genes used for lineage tracing can help reconcile

398 previous reports[12,13]. Future studies will need to focus on the regulation of these early events and
399 the key niche signals which maintain stem cell renewal within the intestinal isthmus.

400 **Author contribution:** EM, AV, AC, TCW conceived, designed the study and wrote the
401 manuscript. EM performed most of the experiments. AV performed the computational analyses.
402 YO performed *in vitro* experiments. WK generated and characterized Dll1 transgenic mice. MM,
403 HN, BB, JL, LBZ participated in performing the experiments. MHW provided the B6A6 antibody.
404 LL provided critical conceptual input and manuscript revisions. KSY, CWC, and CG participated
405 in the discussion. TCW secured the funding.

406 **Acknowledgments:** This research was funded in part through the NIH/NCI Cancer Center
407 Support Grant P30CA013696 and used the resources of the Herbert Irving Comprehensive
408 Cancer Center Flow Cytometry Shard Resources, Molecular Pathology/MPSR, Genomics and
409 High Throughput Screening, as well as the Genetically Modified Mouse Model Shared Resource
410 (GMMMSR). This research was also supported by the Columbia University Digestive and Liver
411 Disease Research Center (CU-DLDRC) grant 1P30DK132710. This work was supported by
412 grants from the NIH/NCI including UO1DK103155, R35CA210088, RO1NK128195 to TCW; as
413 well as a NCI Outstanding Investigator Award (R35 CA197745) and two NIH Shared
414 Instrumentation Grants (S10 OD012351 and S1 0OD021764), all to AC. AV is supported by a
415 U.S. Department of Defense Early Investigator Research Award (W81XWH19-1-0337) and an
416 Early Career Development Pilot Award NIH/NCI Cancer Center, funded through the Cancer
417 Center Support grant, P30CA013696. We are grateful for the support we received by Columbia
418 University shared resources and we would like to thank Sun Dajiang “Kevin” (Molecular
419 Pathology/MPSR), Kissner, Michael (Columbia Stem Cell Initiative Flow Cytometry core facility),
420 Erin Bush (Genomics and High Throughput Screening), and Lin Chyuan-Sheng “Victor”
421 (Genetically Modified Mouse Model), for their incredible expertise and help in this project. We
422 thank Nicoletta Barolini (Columbia University) for designing the graphical models.

423 **Conflict of interest:** Dr. Califano is founder, equity holder, and consultant of DarwinHealth Inc.,
424 a company that has licensed some of the algorithms used in this manuscript from Columbia
425 University. Columbia University is also an equity holder in DarwinHealth Inc. US patent number
426 10,790,040 has been awarded related to this work, and has been assigned to Columbia University
427 with Dr. Califano as an inventor.

428 **References**

429 1 Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed in vivo in intestinal stem cells. *Nat Genet* **40**,
430 915-920, doi:10.1038/ng.165 (2008).
431 2 Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly
432 cycling intestinal stem cells. *Proc Natl Acad Sci U S A* **108**, 179-184, doi:10.1073/pnas.1013004108
433 (2011).
434 3 Takeda, N. *et al.* Interconversion between intestinal stem cell populations in distinct niches.
435 *Science* **334**, 1420-1424, doi:10.1126/science.1213214 (2011).
436 4 Powell, A. E. *et al.* The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that
437 functions as a tumor suppressor. *Cell* **149**, 146-158, doi:10.1016/j.cell.2012.02.042 (2012).
438 5 Asfaha, S. *et al.* Krt19(+)/Lgr5(-) Cells Are Radioresistant Cancer-Initiating Stem Cells in the Colon
439 and Intestine. *Cell Stem Cell* **16**, 627-638, doi:10.1016/j.stem.2015.04.013 (2015).
440 6 Sheng, X. *et al.* Cycling Stem Cells Are Radioresistant and Regenerate the Intestine. *Cell Rep* **32**,
441 107952, doi:10.1016/j.celrep.2020.107952 (2020).
442 7 Ayyaz, A. *et al.* Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell.
443 *Nature* **569**, 121-125, doi:10.1038/s41586-019-1154-y (2019).
444 8 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5.
445 *Nature* **449**, 1003-1007, doi:10.1038/nature06196 (2007).
446 9 Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells
447 dispensable. *Nature* **478**, 255-259, doi:10.1038/nature10408 (2011).
448 10 Murata, K. *et al.* Ascl2-Dependent Cell Dedifferentiation Drives Regeneration of Ablated Intestinal
449 Stem Cells. *Cell Stem Cell* **26**, 377-390 e376, doi:10.1016/j.stem.2019.12.011 (2020).
450 11 Shivdasani, R. A., Clevers, H. & de Sauvage, F. J. Tissue regeneration: Reserve or reverse? *Science*
451 **371**, 784-786, doi:10.1126/science.abb6848 (2021).
452 12 van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat Cell*
453 *Biol* **14**, 1099-1104, doi:10.1038/ncb2581 (2012).
454 13 Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their
455 Enterocyte-Lineage Daughters. *Cell Stem Cell* **18**, 203-213, doi:10.1016/j.stem.2016.01.001
456 (2016).
457 14 Obradovic, A. *et al.* Single-cell protein activity analysis identifies recurrence-associated renal
458 tumor macrophages. *Cell* **184**, 2988-3005 e2916, doi:10.1016/j.cell.2021.04.038 (2021).
459 15 Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals
460 Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov* **9**, 1102-1123,
461 doi:10.1158/2159-8290.CD-19-0094 (2019).
462 16 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339,
463 doi:10.1038/nature24489 (2017).
464 17 Bottcher, A. *et al.* Non-canonical Wnt/PCP signalling regulates intestinal stem cell lineage priming
465 towards enteroendocrine and Paneth cell fates. *Nat Cell Biol* **23**, 23-31, doi:10.1038/s41556-020-
466 00617-2 (2021).
467 18 Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the
468 Intestinal Villus Axis. *Cell* **175**, 1156-1167 e1115, doi:10.1016/j.cell.2018.08.063 (2018).
469 19 Beumer, J. & Clevers, H. Cell fate specification and differentiation in the adult mammalian
470 intestine. *Nat Rev Mol Cell Biol* **22**, 39-53, doi:10.1038/s41580-020-0278-0 (2021).
471 20 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential.
472 *Science* **367**, 405-411, doi:10.1126/science.aax0249 (2020).

21 Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun* **8**, 15599, doi:10.1038/ncomms15599 (2017).
22 Guo, M., Bao, E. L., Wagner, M., Whitsett, J. A. & Xu, Y. SLICE: determining cell differentiation and lineage based on single cell entropy. *Nucleic Acids Res* **45**, e54, doi:10.1093/nar/gkw1278 (2017).
23 Malta, T. M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354 e315, doi:10.1016/j.cell.2018.03.034 (2018).
24 Wang, F. *et al.* Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* **145**, 383-395 e381-321, doi:10.1053/j.gastro.2013.04.050 (2013).
25 Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat Commun* **11**, 1169, doi:10.1038/s41467-020-14976-9 (2020).
26 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838-847, doi:10.1038/ng.3593 (2016).
27 Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233-2235, doi:10.1093/bioinformatics/btw216 (2016).
28 Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7, doi:10.1186/1471-2105-7-S1-S7 (2006).
29 Ding, H. *et al.* Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat Commun* **9**, 1471, doi:10.1038/s41467-018-03843-3 (2018).
30 Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* **12**, 36-47, doi:10.1038/nrm3036 (2011).
31 Gehart, H. *et al.* Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell* **176**, 1158-1173 e1116, doi:10.1016/j.cell.2018.12.029 (2019).
32 Chen, Z. *et al.* Nuclear DEK preserves hematopoietic stem cells potential via NCoR1/HDAC3-Akt1/2-mTOR axis. *J Exp Med* **218**, doi:10.1084/jem.20201974 (2021).
33 Ding, Y. *et al.* Smarca5-mediated epigenetic programming facilitates fetal HSPC development in vertebrates. *Blood* **137**, 190-202, doi:10.1182/blood.2020005219 (2021).
34 Baggiolini, A. *et al.* Developmental chromatin programs determine oncogenic competence in melanoma. *Science* **373**, eabc1048, doi:10.1126/science.abc1048 (2021).
35 Zhang, Z. *et al.* Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. *Proc Natl Acad Sci U S A* **114**, 3903-3908, doi:10.1073/pnas.1621177114 (2017).
36 Perekatt, A. O. *et al.* YY1 is indispensable for Lgr5+ intestinal stem cell renewal. *Proc Natl Acad Sci U S A* **111**, 7695-7700, doi:10.1073/pnas.1400128111 (2014).
37 Yu, T. *et al.* SRSF1 governs progenitor-specific alternative splicing to maintain adult epithelial tissue homeostasis and renewal. *Dev Cell* **57**, 624-637 e624, doi:10.1016/j.devcel.2022.01.011 (2022).
38 Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* **8**, 285-292, doi:10.1038/ncb1372 (2006).
39 Han, S. *et al.* Defining the Identity and Dynamics of Adult Gastric Isthmus Stem Cells. *Cell Stem Cell* **25**, 342-356 e347, doi:10.1016/j.stem.2019.07.008 (2019).
40 Mull, A. N., Klar, A. & Navara, C. S. Differential localization and high expression of SURVIVIN splice variants in human embryonic stem cells but not in differentiated cells implicate a role for SURVIVIN in pluripotency. *Stem Cell Res* **12**, 539-549, doi:10.1016/j.scr.2014.01.002 (2014).
41 Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405 e319, doi:10.1016/j.cell.2018.05.043 (2018).

42 Bezencon, C. *et al.* Murine intestinal cells expressing Trpm5 are mostly brush cells and express markers of neuronal and inflammatory cells. *J Comp Neurol* **509**, 514-525, doi:10.1002/cne.21768 (2008).
43 Middelhoff, M. *et al.* Prox1-positive cells monitor and sustain the murine intestinal epithelial cholinergic niche. *Nat Commun* **11**, 111, doi:10.1038/s41467-019-13850-7 (2020).
44 van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903-912, doi:10.1016/j.cell.2009.01.031 (2009).
45 Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265, doi:10.1038/nature07935 (2009).
46 Basak, O. *et al.* Mapping early fate determination in Lgr5+ crypt stem cells using a novel Ki67-RFP allele. *EMBO J* **33**, 2057-2068, doi:10.15252/embj.201488017 (2014).
47 de Lau, W., Peng, W. C., Gros, P. & Clevers, H. The R-spondin/Lgr5/Rnf43 module: regulator of Wnt signal strength. *Genes Dev* **28**, 305-316, doi:10.1101/gad.235473.113 (2014).
48 Fafilek, B. *et al.* Troy, a tumor necrosis factor receptor family member, interacts with lgr5 to inhibit wnt signaling in intestinal stem cells. *Gastroenterology* **144**, 381-391, doi:10.1053/j.gastro.2012.10.048 (2013).
49 Westphalen, C. B. *et al.* Long-lived intestinal tuft cells serve as colon cancer-initiating cells. *J Clin Invest* **124**, 1283-1295, doi:10.1172/JCI73434 (2014).
50 Azkanaz, M. *et al.* Retrograde movements determine effective stem cell numbers in the intestine. *Nature* **607**, 548-554, doi:10.1038/s41586-022-04962-0 (2022).
51 Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134-144, doi:10.1016/j.cell.2010.09.016 (2010).
52 Tan, S. H. *et al.* A constant pool of Lgr5(+) intestinal stem cells is required for intestinal homeostasis. *Cell Rep* **34**, 108633, doi:10.1016/j.celrep.2020.108633 (2021).
53 Orre, L. M., Pernemalm, M., Lengqvist, J., Lewensohn, R. & Lehtio, J. Up-regulation, modification, and translocation of S100A6 induced by exposure to ionizing radiation revealed by proteomics profiling. *Mol Cell Proteomics* **6**, 2122-2131, doi:10.1074/mcp.M700202-MCP200 (2007).
54 Madureira, P. A., Hill, R., Lee, P. W. & Waisman, D. M. Genotoxic agents promote the nuclear accumulation of annexin A2: role of annexin A2 in mitigating DNA damage. *PLoS One* **7**, e50591, doi:10.1371/journal.pone.0050591 (2012).
55 Awasthi, Y. C., Ramana, K. V., Chaudhary, P., Srivastava, S. K. & Awasthi, S. Regulatory roles of glutathione-S-transferases and 4-hydroxynonenal in stress-mediated signaling and toxicity. *Free Radic Biol Med* **111**, 235-243, doi:10.1016/j.freeradbiomed.2016.10.493 (2017).
56 Whitaker, A. M., Flynn, T. S. & Freudenthal, B. D. Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. *Nat Commun* **9**, 399, doi:10.1038/s41467-017-02175-y (2018).
57 Oppikofer, M. *et al.* Non-canonical reader modules of BAZ1A promote recovery from DNA damage. *Nat Commun* **8**, 862, doi:10.1038/s41467-017-00866-0 (2017).
58 Yui, S. *et al.* YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* **22**, 35-49 e37, doi:10.1016/j.stem.2017.11.001 (2018).
59 Buczacki, S. J. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* **495**, 65-69, doi:10.1038/nature11965 (2013).
60 Leushacke, M. *et al.* Lgr5-expressing chief cells drive epithelial regeneration and cancer in the oxyntic stomach. *Nat Cell Biol* **19**, 774-786, doi:10.1038/ncb3541 (2017).
61 Sigal, M. *et al.* R-spondin-3 induces secretory, antimicrobial Lgr5(+) cells in the stomach. *Nat Cell Biol* **21**, 812-823, doi:10.1038/s41556-019-0339-9 (2019).

566 62 Hayakawa, Y., Nakagawa, H., Rustgi, A. K., Que, J. & Wang, T. C. Stem cells and origins of cancer in
567 the upper gastrointestinal tract. *Cell Stem Cell* **28**, 1343-1361, doi:10.1016/j.stem.2021.05.012
568 (2021).
569 63 Schuijers, J. *et al.* Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in
570 intestinal crypts. *Cell Stem Cell* **16**, 158-170, doi:10.1016/j.stem.2014.12.006 (2015).
571 64 Shao, J. & Sheng, H. Amphiregulin promotes intestinal epithelial regeneration: roles of intestinal
572 subepithelial myofibroblasts. *Endocrinology* **151**, 3728-3737, doi:10.1210/en.2010-0319 (2010).
573 65 Gregorieff, A., Liu, Y., Inanlou, M. R., Khomchuk, Y. & Wrana, J. L. Yap-dependent reprogramming
574 of Lgr5(+) stem cells drives intestinal regeneration and cancer. *Nature* **526**, 715-718,
575 doi:10.1038/nature15382 (2015).

576

Isthmus progenitor cells contribute to homeostatic cellular turnover and support regeneration following intestinal injury

Malagola E.[1,#], Vasciaveo A.[2,#], Ochiai Y.[1], Kim W.[1], Middelhoff M.[3], Nienhüser H.[4], Deng L.[5], Belin B.[1], LaBella J.[1], Zamechek LB.[1], Wong M.H.[6], Li L.[5], Guha C.[7], Cheng CW.[8], Yan KS.[1,8], Califano A.[2,9,10,11,12], and Wang T.C.[1,12].

Figure panels

Figure 1: An unbiased approach to elucidate the organization of intestinal crypt epithelial cells

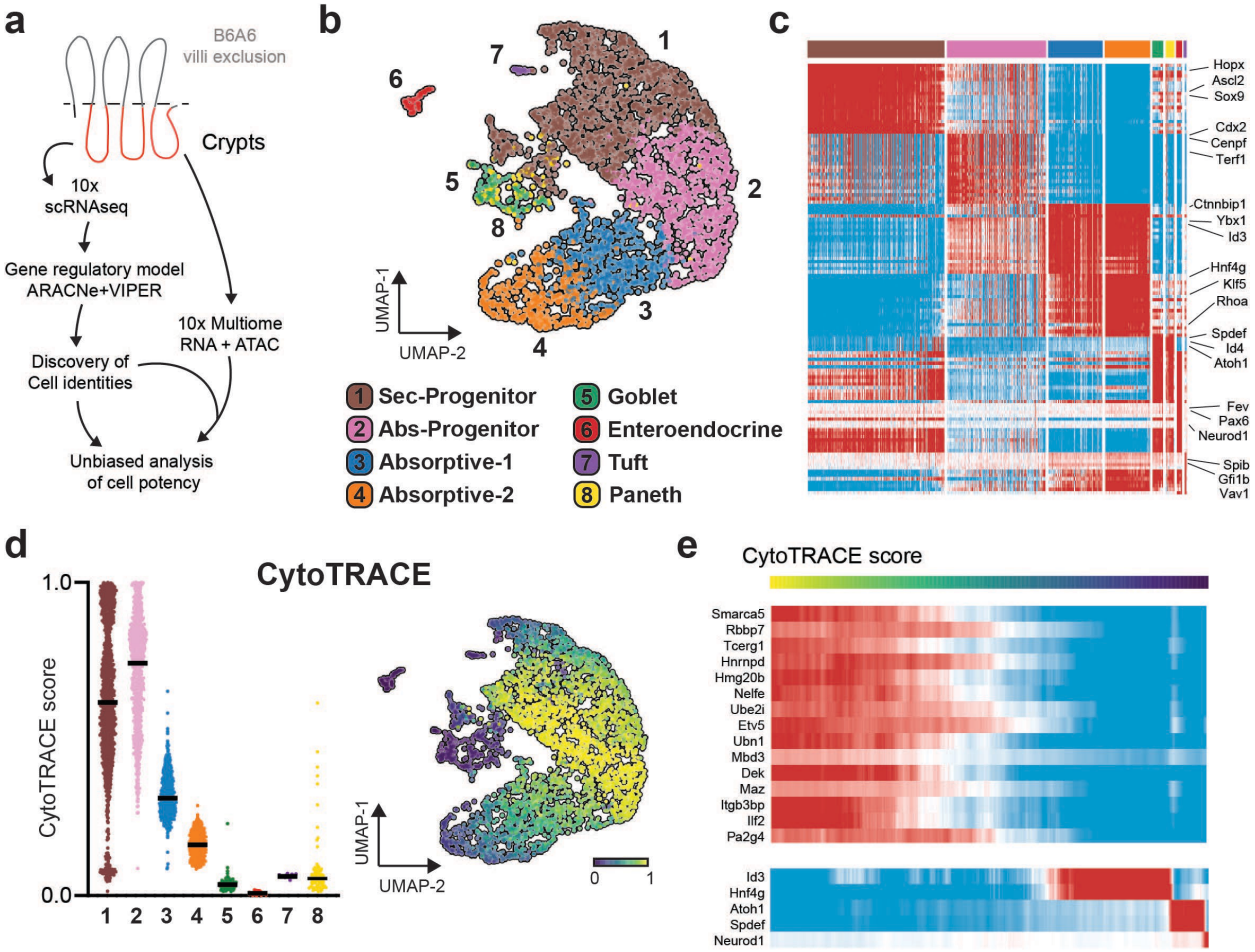
Figure 2: *Lgr5* expression is not restricted to intestinal stem and progenitor cells

Figure 3: Stemness potential exists beyond Lgr5^{+} cells

Figure 4: Lgr5^{neg} isthmus cells compensate for loss of *Lgr5* expressing cells

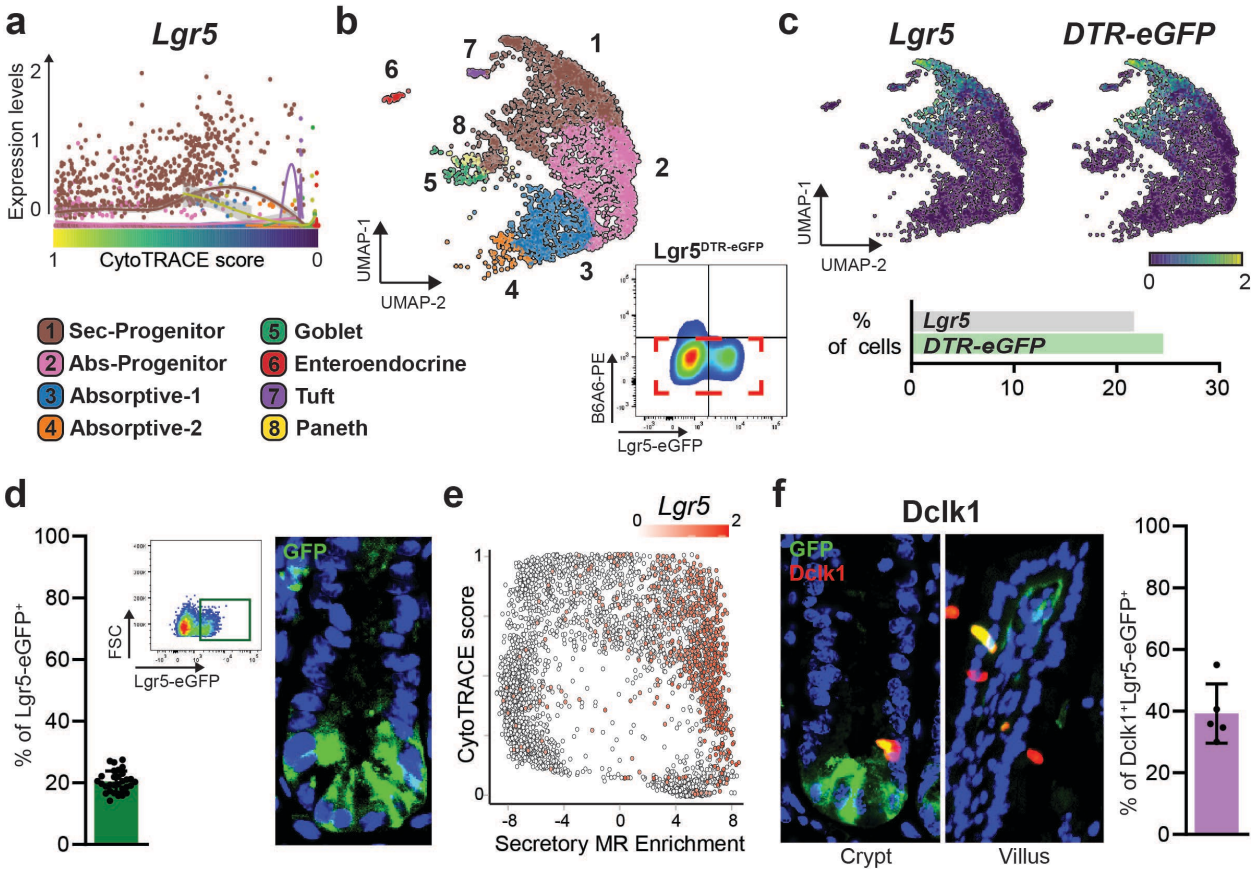
Figure 5: Surviving isthmus cells regenerate the intestinal epithelium following IR damage

Figure 1



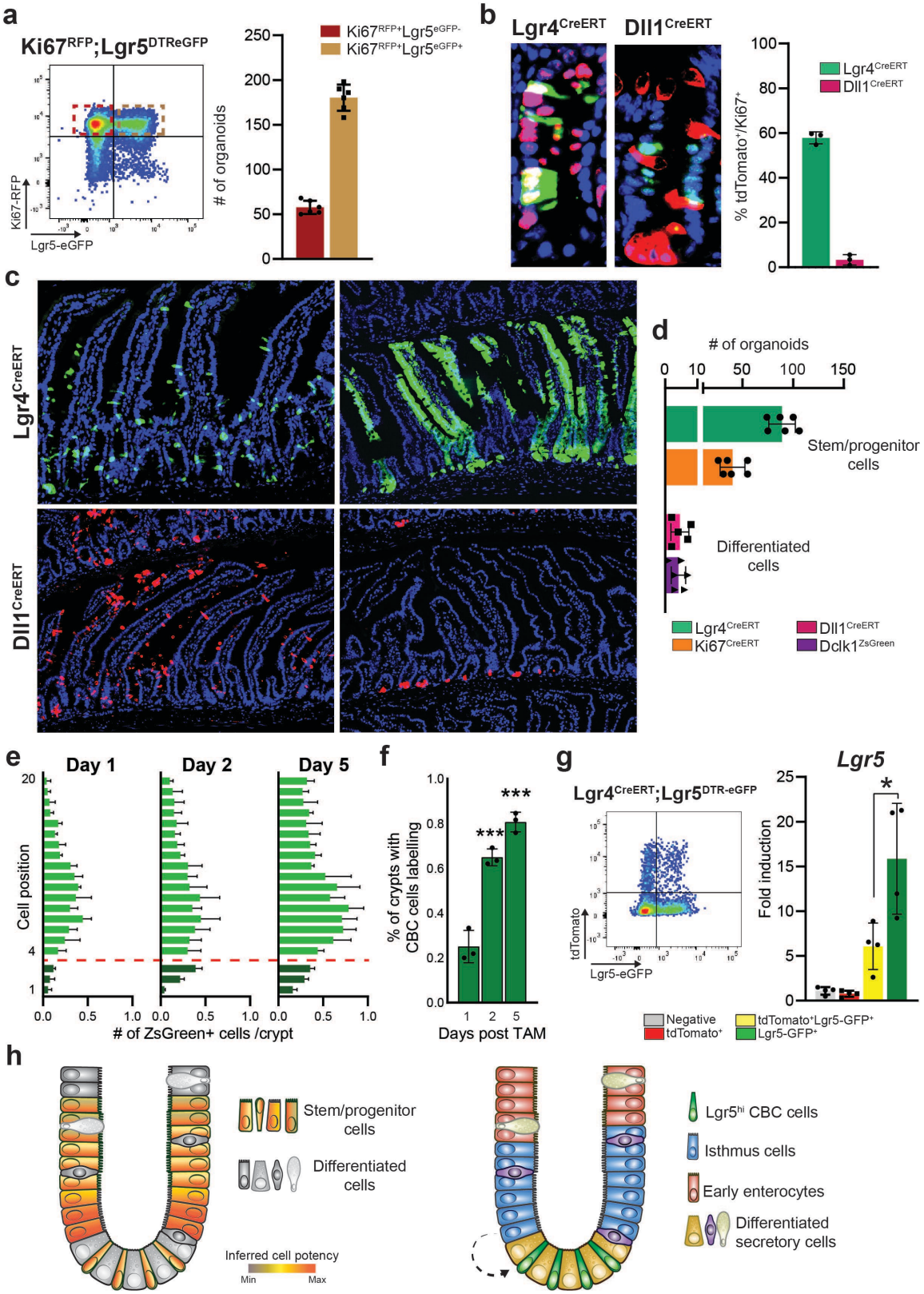
An unbiased approach to elucidate the organization of intestinal crypt epithelial cells: a. Schematic representation of the experimental workflow and computational pipeline for the analysis of crypt epithelial cells. (*See methods*) **b.** UMAP plot showing protein activity -based clustering solution based on gene regulatory network analysis. **c.** Heatmap showing top differentially activated regulatory proteins. (Active scores in red) **d.** Violin- and UMAP plots showing CytoTRACE scores for individual cells; black lines indicate median value per cluster. **e.** Pseudotime analysis using CytoTRACE scores to order cells based on inferred cell potency (high – yellow - to low – blue - potency from left to right), showing protein activity changes of top correlating regulatory proteins. On the bottom, protein activity profile for known transcription factors involved in cellular differentiation. (Active scores in red)

Figure 2



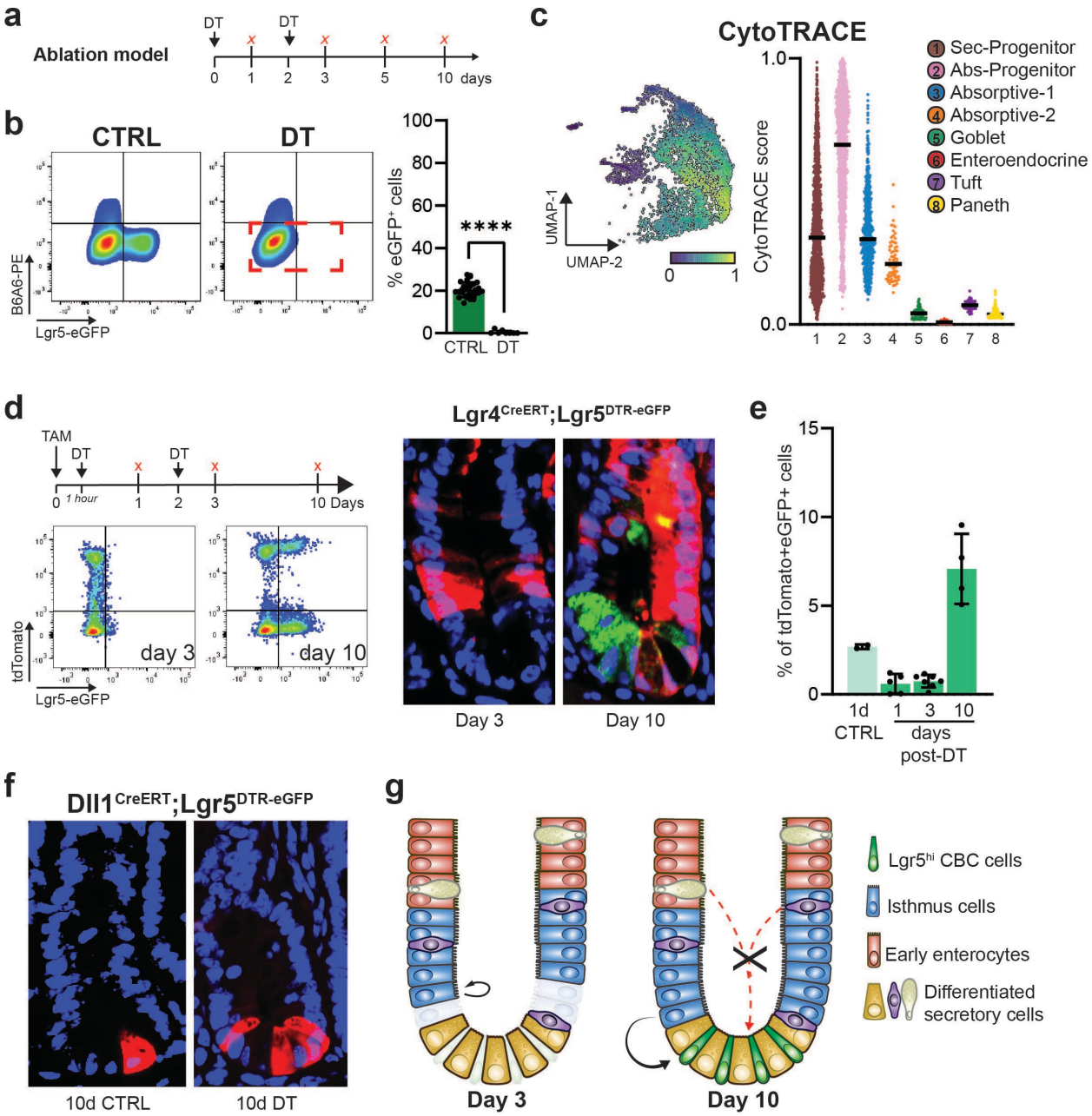
***Lgr5* expression is not restricted to intestinal stem and progenitor cells: a.** Pseudotime analysis using CytoTRACE scores to order cells based on inferred cell potency (high to low potency from left to right), showing expression levels of *Lgr5* on y-axis. Cells are colored based on their cluster. **b.** Sorting strategy and UMAP plot of Lgr5$^{DTR\text{-}eGFP}$ crypt epithelial cells. (Sorted gate highlighted in Red) **c.** UMAP plots showing expression of *Lgr5*^{Wt} and *Lgr5*$^{DTR\text{-}eGFP}$ alleles, below bar-plot showing the number of cells expressing each allele in the dataset (as % of total cells). **d.** Bar-plot and gating strategy for flow cytometry –based quantification of Lgr5-eGFP^{+} cells in purified crypts (n=30); together with representative image of immunofluorescence staining for Lgr5-eGFP in mouse jejunum. **e.** Scatter plot of inferred cell potency (CytoTRACE) on y-axis and enrichment scores for Secretory lineage on x-axis. Cells are colored based on Lgr5 expression. **f.** Immunofluorescence staining for Dclk1 and bar-plot showing quantification of double positive cells (Dclk1^{+}Lgr5-eGFP^{+} over total Dclk1^{+}) (n=5).

Figure 3:


a
Ki67RFP;Lgr5DTReGFP
Ki67-RFP
Lgr5-eGFP
Ki67RFP+Lgr5eGFP-
Ki67RFP+Lgr5eGFP+
# of organoids
b
Lgr4CreERT
Dll1CreERT
% tdTomato+/Ki67+
c
Lgr4CreERT
Dll1CreERT
d
# of organoids
Stem/progenitor cells
Differentiated cells
Lgr4CreERT
Ki67CreERT
Dll1CreERT
Dclk1ZsGreen
e
Day 1
Day 2
Day 5
Cell position
# of ZsGreen+ cells /crypt
f
% of crypts with CBC cells labelling
Days post TAM
g
Lgr4CreERT;Lgr5DTR-eGFP
tdTomato
Lgr5-eGFP
Lgr5
Fold induction
Negative
tdTomato+
tdTomato+Lgr5-GFP+
Lgr5-GFP+
h
Stem/progenitor cells
Differentiated cells
Inferred cell potency
Min
Max
Lgr5hi CBC cells
Isthmus cells
Early enterocytes
Differentiated secretory cells

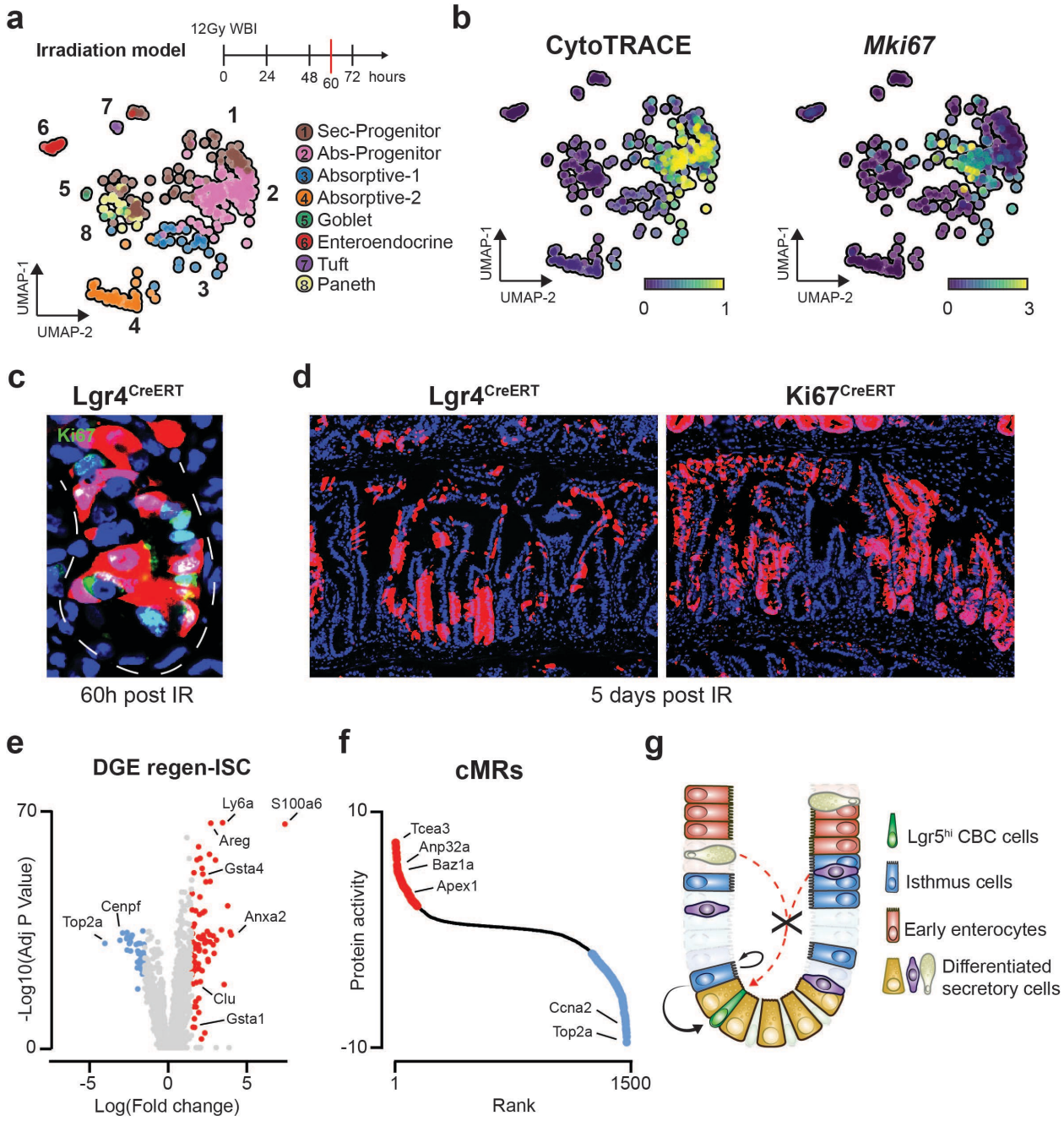
Isthmal stem cells sustain intestinal tissue homeostasis: a. Representative scatter-plot for flow cytometry –based cells sorting (Ki67^{RFP};Lgr5^{DTReGFP}). On the right, bar-plot showing number of organoids generated from the isolated fractions. **b.** Representative image of Ki67 staining combined with lineage tracing in Lgr4^{CreERT} and Dll1^{CreERT} mice (24h post TAM). On the right, bar-plot showing quantification of % of Ki67+/tdTomato+ cells. **c.** Representative images of Lgr4^{CreERT} and Dll1^{CreERT} –derived labelling 1 day and 1 (Dll1^{CreERT}) or 6 (Lgr4^{CreERT}) months post TAM induction. **d.** Bar-plot showing quantification of the number of organoids generated from each sorted population. **e.** Bar-plots showing quantification of traced cells in Lgr4^{CreERT} based on cell position within the crypt. Red dotted line used to mark the boundary between CBC and isthmus cells (n=3). **f.** Bar-plots showing percentage of crypt-units with at least one labelled CBC cell in Lgr4^{CreERT}. **g.** Flow cytometry analysis of overlap between Lgr4^{CreERT} labelled cells (24h post TAM) and Lgr5^{eGFP+} cells; together with bar plot showing expression levels of *Lgr5* in single sorted cells, groups labelled in the panel. (n=4). **h.** (Left) Schematic representation of the intestinal crypts, cells are colored based on inferred cell potency. Highest stemness potential is found in the isthmus region (e.g. Lgr5^{low} cells). (Right) Schematic representation of intestinal crypts, isthmus cells, colored in blue, have the capacity to migrate downward and give rise to Lgr5+ CBCs during homeostasis. Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Figure 4


a
Ablation model
DT
DT
0 1 2 3 5 10 days
b
CTRL
DT
B6A6-PE
Lgr5-eGFP
% eGFP+ cells
****
CTRL DT
c
CytoTRACE
UMAP-1
UMAP-2
0 1
CytoTRACE score
1.0
0.0
1 2 3 4 5 6 7 8
1 Sec-Progenitor
2 Abs-Progenitor
3 Absorptive-1
4 Absorptive-2
5 Goblet
6 Enteroendocrine
7 Tuft
8 Paneth
d
TAM
DT
DT
0 1 hour 1 2 3 10 Days
tdTomato
Lgr5-eGFP
day 3
day 10
Lgr4CreERT;Lgr5DTR-eGFP
Day 3
Day 10
e
% of tdTomato+eGFP+ cells
1d CTRL
1 3 10 days post-DT
f
Dll1CreERT;Lgr5DTR-eGFP
10d CTRL
10d DT
g
Day 3
Day 10
Lgr5hi CBC cells
Isthmus cells
Early enterocytes
Differentiated secretory cells

Lgr5^{neg} proliferating cells compensate for loss of *Lgr5* expressing cells: a. Schematic representation of DT based ablation model. **b.** On the left, representative dot-plot of gating strategy adopted for analysis of crypt epithelial cells in CTRL or DT ablated mice (sorted gate in red). On the right, bar-plot showing flow cytometry quantification of Lgr5$^{DTReGFP+}$ cells in control and DT treated crypts (n≥10). **c.** UMAP and violin plots showing CytoTRACE scores in DT treated epithelial cells. **d.** Schematic representation of the model of study together with representative scatter plots for flow-cytometry -based analysis of lineage tracing (Lgr4^{CreERT}). On the right, representative images for Lgr4^{CreERT};Lgr5^{DTReGFP} mice lineage tracing upon concurrent DT treatment. **e.** Bar-plot showing quantification of double positive (tdTomato^{+}/Lgr5$^{DTReGFP+}$) cells at the indicated time points (Lgr4^{CreERT}). (n≥4) **f.** Representative images of Dll1^{CreERT} mice 10 days post TAM induction with or without DT treatment. **g.** Schematic representation of the model proposed, on the left layout of intestinal crypts 3 days post DT treatment, at day 10 (on the right) cells turn back to homeostatic organization. Isthmus cells (blue) expand following loss of Lgr5+ cells, supporting normal tissue turnover. On the other hand, differentiated cells retain low cell potency and do not serve as facultative stem cells. Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Figure 5


a
Irradiation model
12Gy WBI
0
24
48
60
72
hours
1 Sec-Progenitor
2 Abs-Progenitor
3 Absorptive-1
4 Absorptive-2
5 Goblet
6 Enteroendocrine
7 Tuft
8 Paneth
UMAP-1
UMAP-2
b
CytoTRACE
Mki67
c
Lgr4CreERT
Ki67
60h post IR
d
Lgr4CreERT
Ki67CreERT
5 days post IR
e
DGE regen-ISC
Ly6a
S100a6
Areg
Gsta4
Cenpf
Top2a
Anxa2
Clu
Gsta1
-Log10(Adj P Value)
Log(Fold change)
f
cMRs
Tcea3
Anp32a
Baz1a
Apex1
Ccna2
Top2a
Protein activity
Rank
g
Lgr5hi CBC cells
Isthmus cells
Early enterocytes
Differentiated secretory cells

Surviving isthmal stem cells regenerate the intestinal epithelium following IR damage: a. Schematic representation of irradiation damage model, together with UMAP plot showing protein activity-based clustering solution for irradiated crypt epithelial cells analyzed 60 hours post IR exposure. **b.** UMAP plots showing both inferred cell potency through CytoTRACE and *Mki67* expression in irradiated crypt epithelial cells. **c.** Representative image for Ki67 staining 60h post IR, in red tdTomato labelling derived from lineage tracing of $Lgr4^{CreERT}$. **d.** Representative images of $Lgr4^{CreERT}$ and $Ki67^{CreERT}$ linage tracing 5 days post IR. **e.** Volcano plot of differentially expressed genes in regenerating (positive LogFC) vs homeostatic ISC (negative LogFC). **f.** Signature of differentially activated regulatory proteins in ISC resulting from exposure to IR damage. Proteins are ranked (left to right) based on computed protein activity score (NES, normalized enrichment score). **g.** Schematic representation of the model proposed of surviving intestinal cells after IR exposure. Surviving pre-existing stem and progenitors expand and drive tissue repair while differentiated cells are unable to serve as regenerating ISCs.

Isthmus progenitor cells contribute to homeostatic cellular turnover and support regeneration following intestinal injury

Malagola E., Vasciaveo A., Ochiai Y., Kim W., Middelhoff M., Nienhüser H., Deng L., Belin B., LaBella J., Zamechek LB., Wong M.H., Li L., Guha C., Cheng CW., Yan KS., Califano A., and Wang T.C.

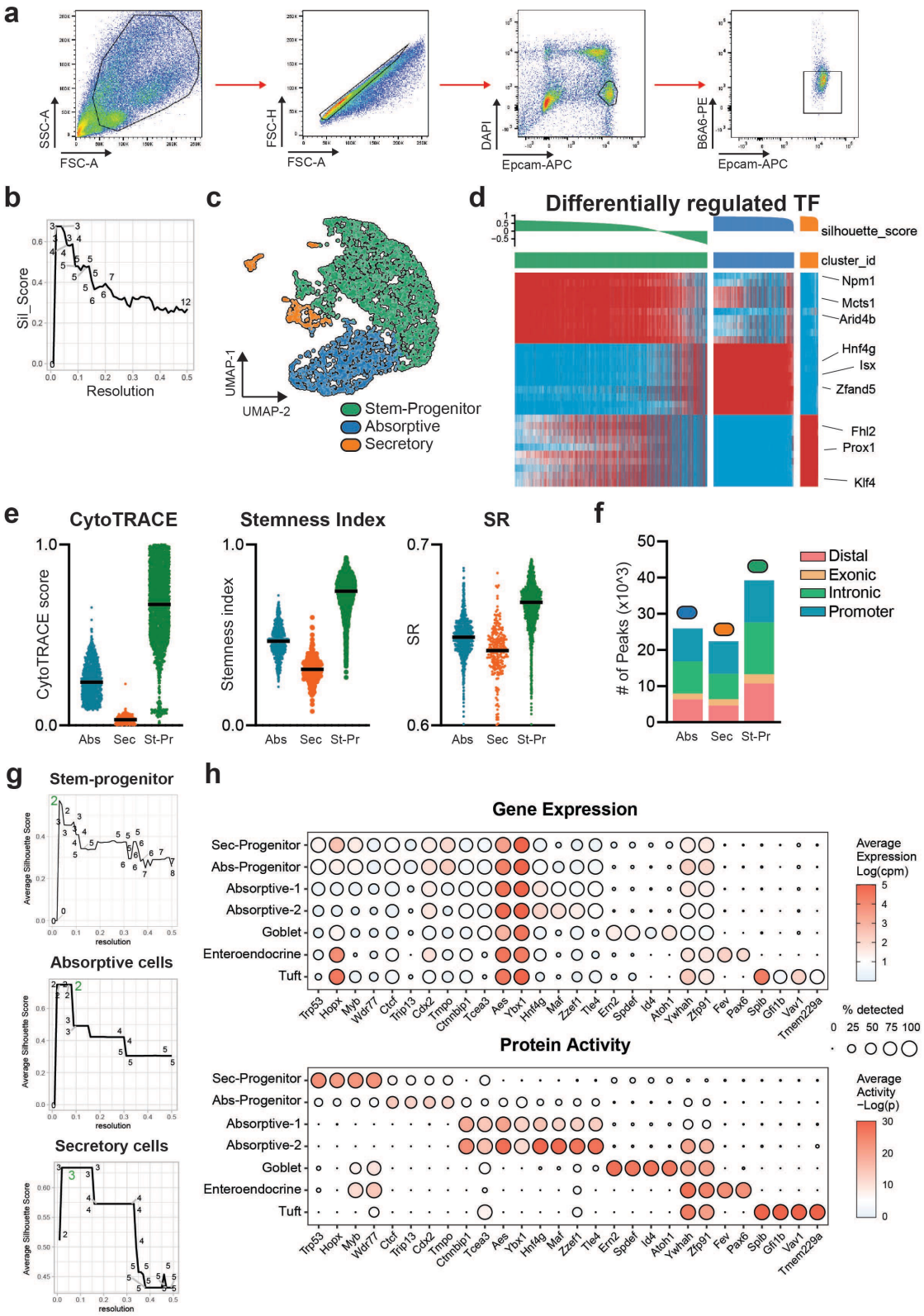
Extended Data Figures and Methods

Extended Data Fig. 1-8

List of Extended Data Table

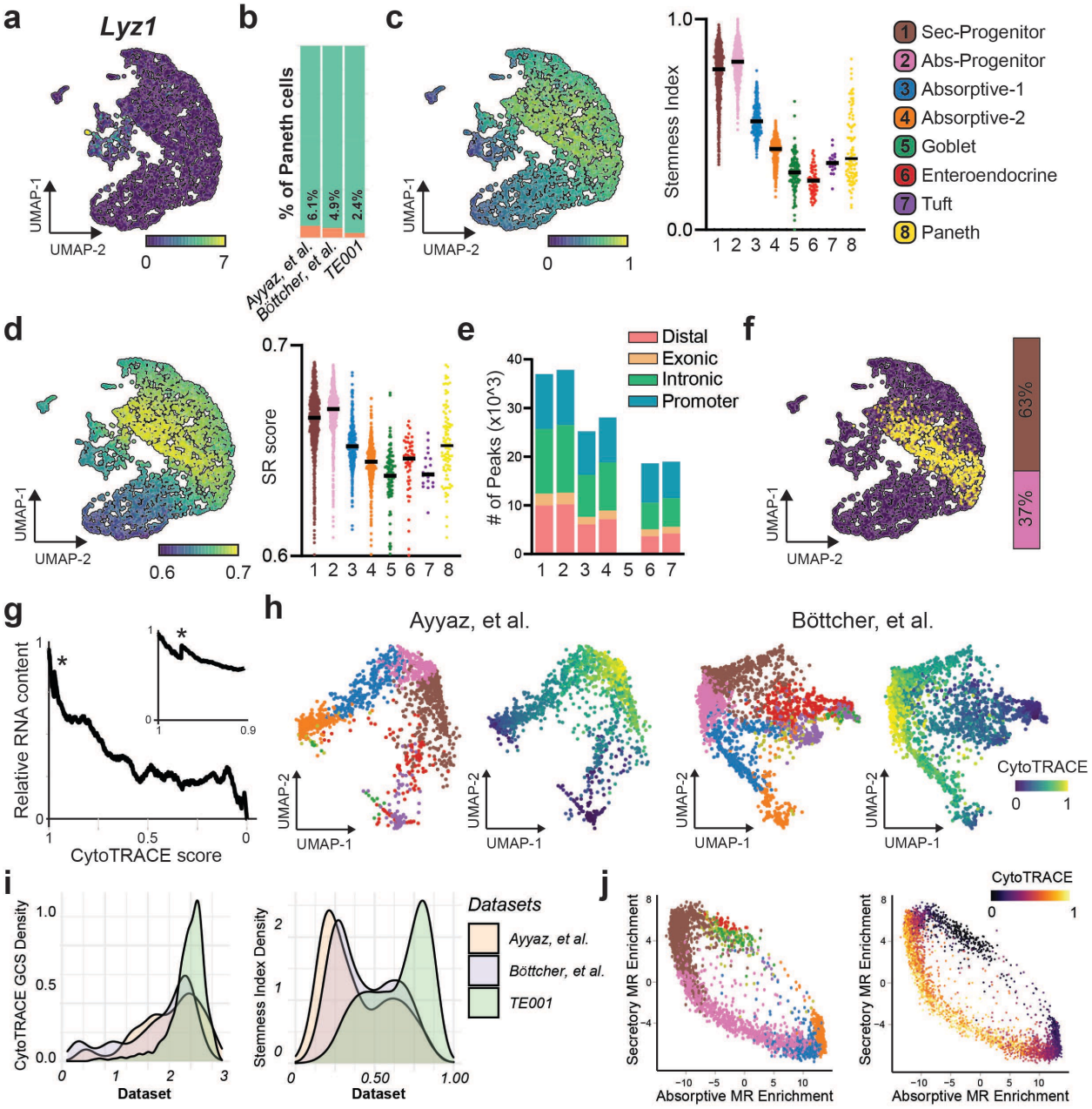
Methods description

Extended Data Figure 1



Extended Data Fig. 1: a. Gating strategy used for flow cytometry-based sorting of crypt epithelial cells. **b.** Showing Louvain's clustering resolution parameter optimization between 0.01 and 0.5. On the y-axis showing the average silhouette score for each clustering solution of crypt epithelial cell regulatory protein activity as identified using the relative resolution parameter shown on the x-axis. This optimization strategy yielded to n=3 clusters. **c.** UMAP plot showing protein activity-based clustering solution. **d.** Heatmap showing top differentially activated regulatory proteins (activated proteins in red color). Silhouette scores for individual cells shown on the top of the heatmap. Cells in each cluster are sorted by silhouette score. **e.** Violin plots showing CytoTRACE, Stemness index, and Signaling Entropy (SR) scores computed for individual cells. **f.** Bar plot showing number of identified accessible peaks per cluster based on scATACseq. **g.** Louvain's clustering resolution parameter optimization as 1-step iterative sub-clusters identification applied to each one of the n=3 clusters identified on crypt epithelial cells (i.e., stem-progenitors, absorptive and secretory cells). In green highlighted the optimal number of clusters for each iteration. **h.** Showing dot-plots of top 4 markers for each cluster (rows) as identified from protein activity-based clustering. Panel on the top shows gene expression as log(cpm), while panel on the bottom shows for the same markers the corresponding protein activity as -log10(p).

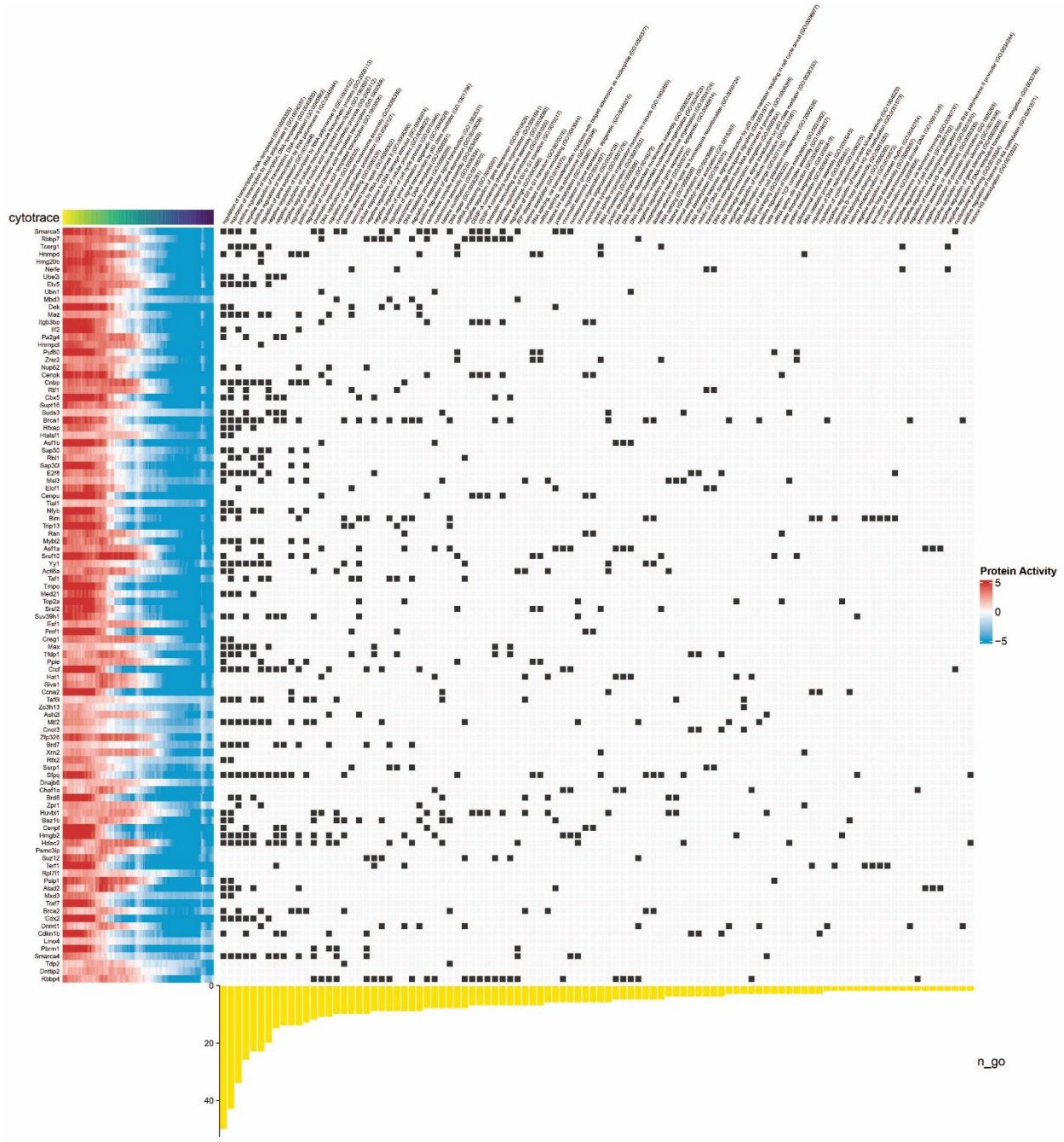
Extended Data Figure 2



Extended Data Fig. 2: a. UMAP plot showing expression domain for *Lyz1* on crypt epithelial cells. **b.** Bar-plot showing quantification of Paneth cells (in orange) across the datasets tested. Candidate Paneth cells were identified based on the concurrent expression of *Lyz1*, *Mmp7*, *Dll4*, and *Nupr1*. **c-d.** UMAP and Violin plots for stemness-index (c.) and entropy (d.) scores computed for individual crypt epithelial cells. **e.** Bar plot showing number of accessible peaks per cluster together; note cluster 5 consist of very few cells and was therefore excluded from further analyses. **f.** UMAP plot showing top 10% cells based on CytoTRACE score, on the right bar-plot showing

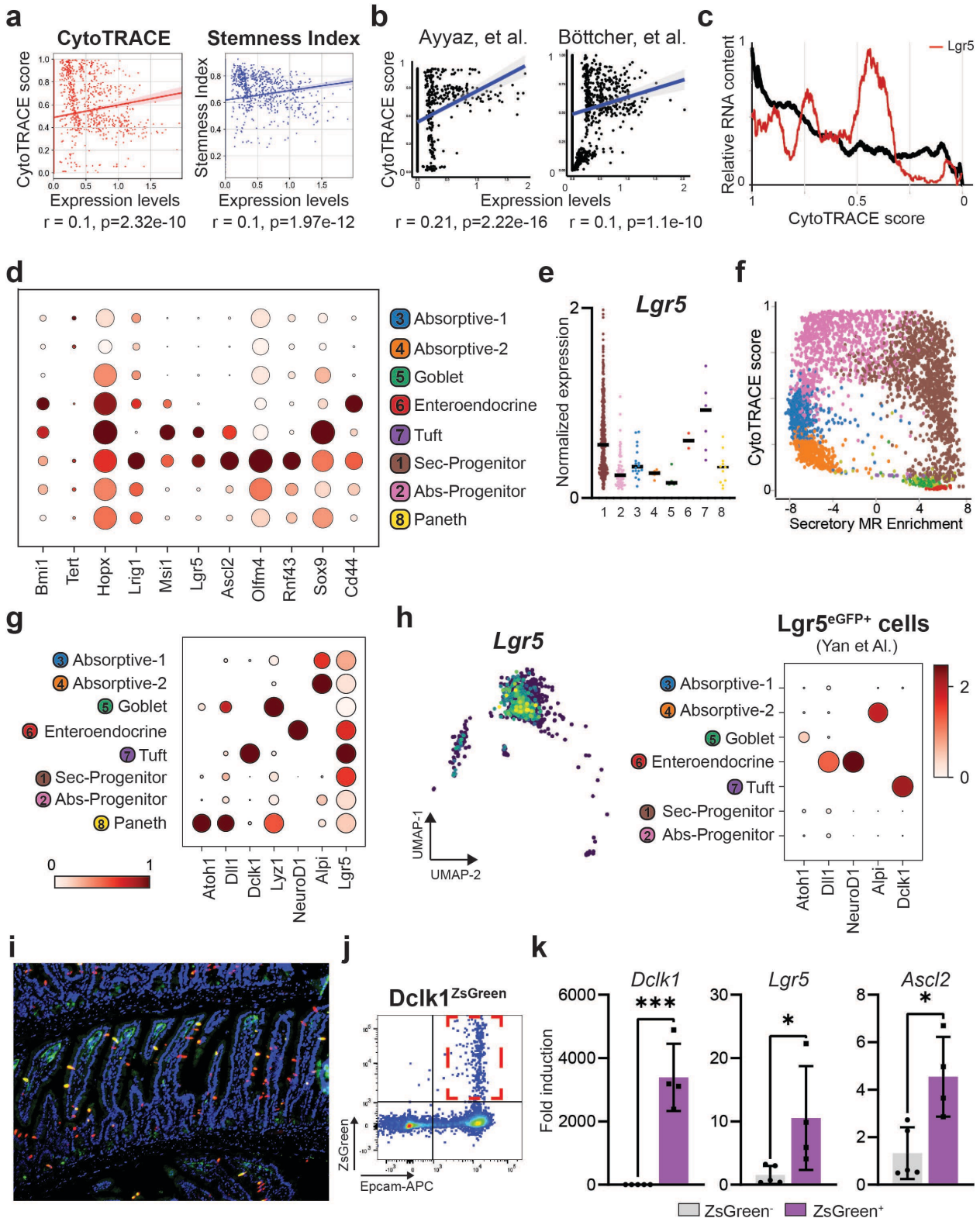
cluster contribution to the top 10% population. **g.** Plot showing individual cells' values of Relative RNA content and Cytotrace scores. In small, showing only the top 10% of cells based on CytoTRACE levels. Black star indicates the 'valley' as previously described[1]. **h.** UMAP plots of homeostatic crypt epithelial cells from Ayyaz et al. and Bottcher et al. analyzed at protein activity, together with CytoTRACE analysis in the two datasets. Clusters are assigned to each cell by performing enrichment analysis on MRs signatures as identified by us. **i.** Density plots showing distribution of CytoTRACE GCS[1] and Stemness-index scores[2] for the three datasets tested. **j.** Showing homeostatic crypt epithelial cells analyzed at protein activity on a scatter plot where each axis represents the extent of enrichment for MRs candidate drivers of absorptive or secretory lineages respectively; on the right, computed CytoTRACE scores are overlaid.

Extended Data Figure 3



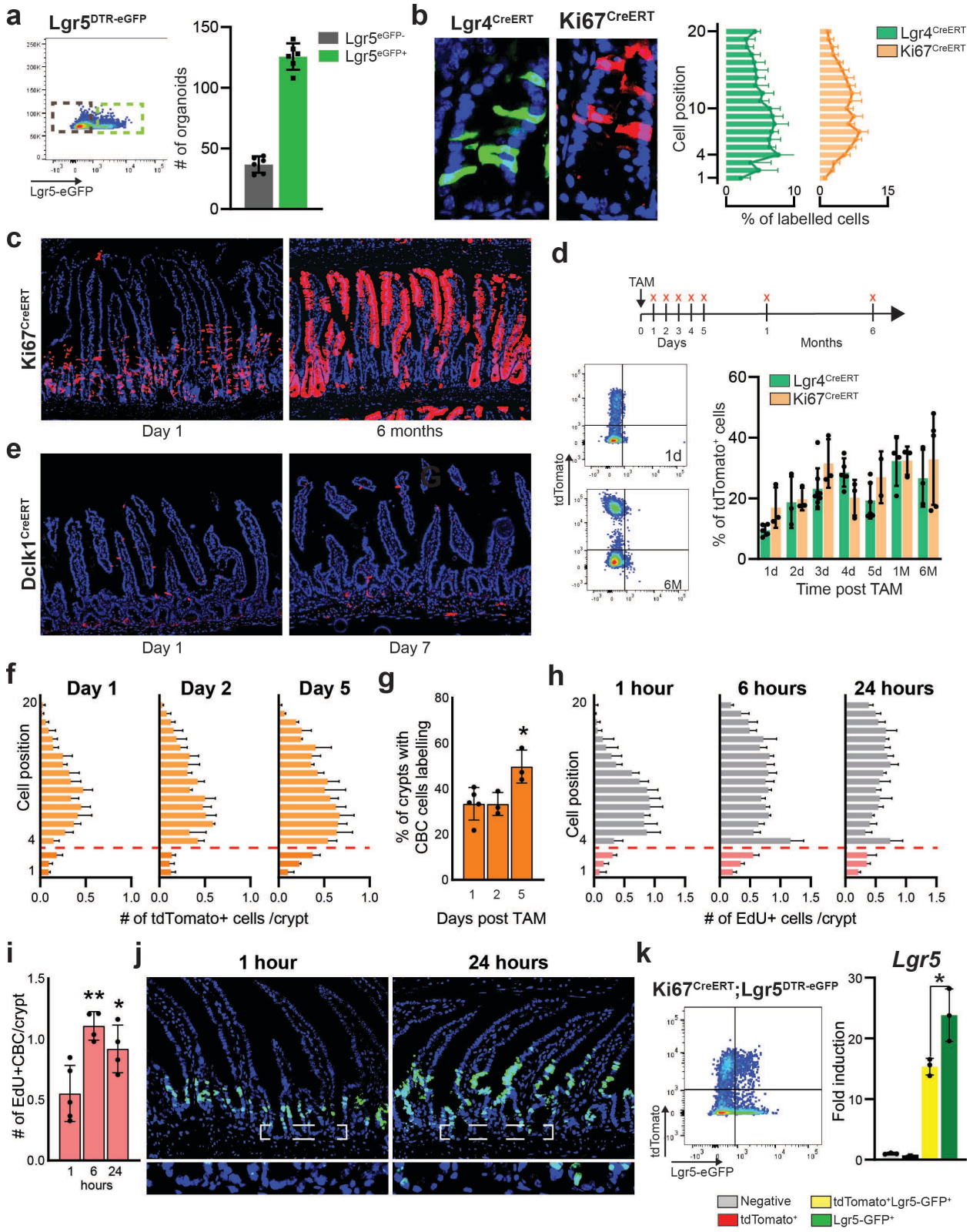
Extended Data Fig. 3: On the left, strip plots showing inferred activity of the 100 CytoTRACE score-based top-correlating transcriptional regulators. On the right, top GO-associated terms for each regulatory protein, using ‘Molecular Function” as query.

Extended Data Figure 4



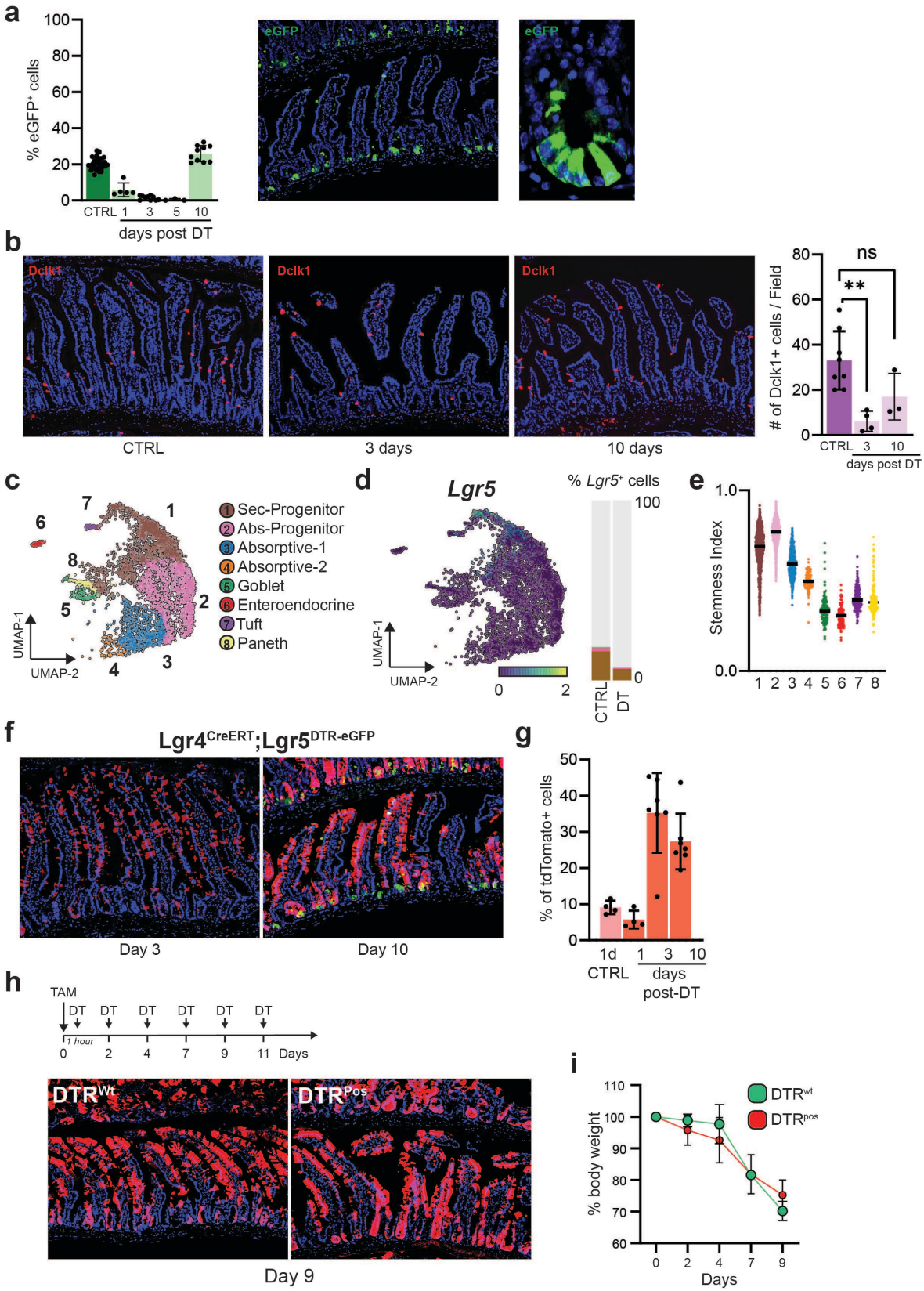
Extended Data Fig. 4: a. Scatter plots showing Pearson's correlation analysis between *Lgr5* expression and inferred cell potency, as assessed by CytoTRACE (left) and Stemness Index (right). **b.** Scatter plots showing Pearson's correlation analysis between *Lgr5* and CytoTRACE scores in the analyzed datasets. **c.** Plot showing individual cells' values of Relative RNA content and Cytotrace scores with Lgr5 expression distribution overlaid in red. **d.** Dot-plot showing expression of selected genes in crypt epithelial cells. **e.** Violin-plot showing expression levels of *Lgr5* in the identified clusters (only $Lgr5^{+}$ cells are shown). **f.** Scatter-plot where inferred cell potency (CytoTRACE, y-axis) is presented together with enrichment scores for secretory (x-axis) lineage. Cells are colored according to clustering solution. **g.** Dot-plot showing expression of known markers of differentiation in Lgr5 expressing cells. (note that only $Lgr5^{+}$ cells are shown) **h.** Analysis of $Lgr5^{eGFP+}$ sorted cells - Source data: Yan KS et al.[3]. Left: UMAP distribution plot showing expression of *Lgr5*; right: dot-plot showing expression of markers of differentiated cells. **i.** Representative image of low power magnification view of Dclk1 (Red) and Lgr5-eGFP (green) in intestinal jejunum of a $Lgr5^{DTR\text{-}eGFP}$ mouse. **j.** Sorting strategy for the isolation of $Dclk1^{ZsGreen+}$ cells (Red gate). **k.** Bar-plots showing expression levels of *Dclk1*, *Lgr5*, and *Ascl2* in sorted $Dclk1^{ZsGreen+}$ cells. (Expression presented as fold induction relative to the negative population) (n=4). Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Extended Data Figure 5:



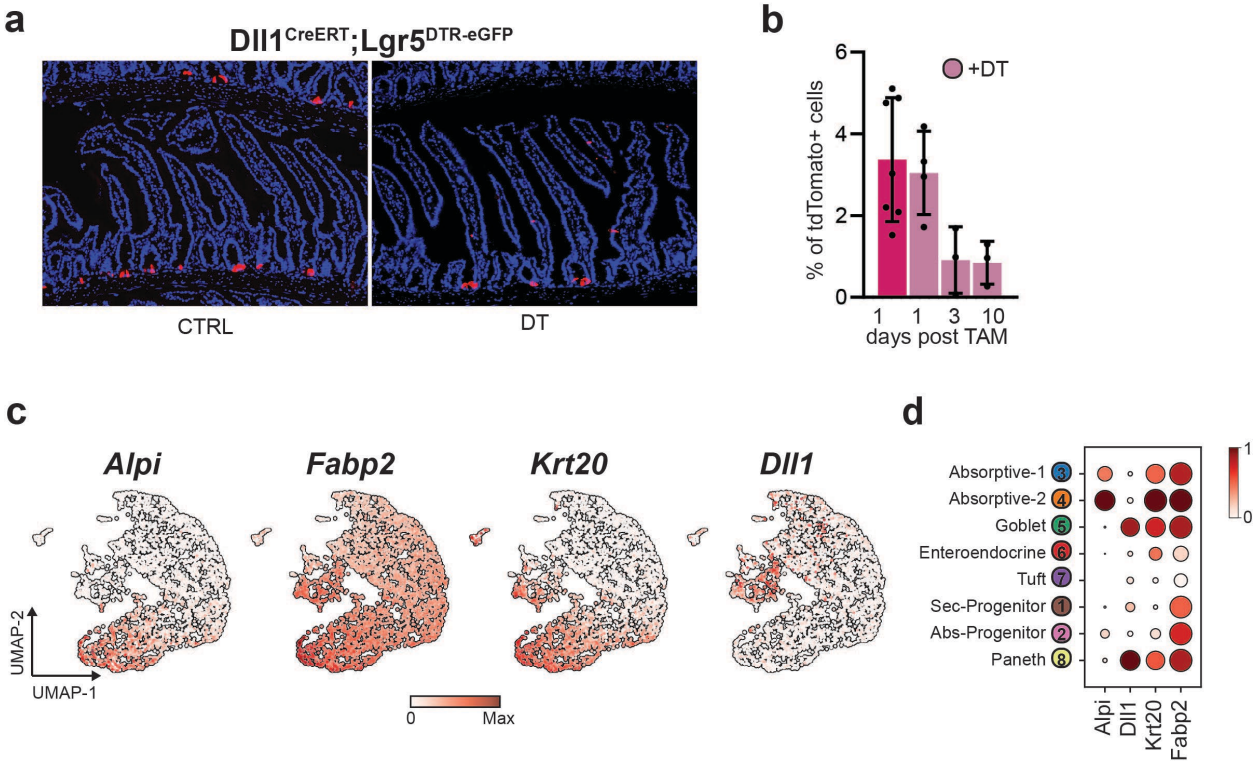
Extended Data Fig. 5: a. Representative scatter-plot for flow cytometry based sorting of Lgr5^{DTReGFP} mice. On the right, bar-plot showing quantification of the number of organoids generated in the isolated cells fractions. **b.** Representative images together with bar-plots showing position based counting of labelled cells 24h post TAM administration in Lgr4^{CreERT} and Ki67^{CreERT} mice. (n=3) **c.** Representative images for Ki67CreERT lineage tracing 1 day and 6 months post TAM induction. **d.** On top, schematic representation of the model of study, red crosses indicate time-points of analysis. Bottom, representative scatter plots for flow-cytometry -based analysis of lineage tracing (Lgr4CreERT) together with bar-plot showing flow cytometry-based quantification of tdTomato+ cells at indicated time points (n≥3). **e.** Lineage tracing analysis of Dclk1^{CreERT} mouse model. At day 1 scattered tdTomato^{+} tuft cells can be observed, at day 7 post TAM most tdTomato^{+} cells are washed out. **f.** Bar-plots showing quantification of traced cells in Ki67^{CreERT} mice based on cell position within the crypt. Red dotted line used to mark the boundary between CBC and isthmus cells (n=3). **g.** Bar-plots showing percentage of crypt-units with at least one labelled CBC cell in Ki67^{CreERT} mice at indicated time points. (n=3) **h.** Bar-plots showing quantification of EdU-labelled cells based on cell position at the indicated time points. Red dotted line used to mark the boundary between CBC and isthmus cells (n=3). **i.** Bar-plot showing average number of EdU labelled-CBC cells at indicated time points. **j.** Representative images of EdU staining in the mouse intestine. Below, enlargement of the highlighted region to show absence/presence of CBC labelling. **k.** Flow cytometry analysis of overlap between Ki67 labelled cells (24h post TAM) and Lgr5^{eGFP+} cells; together with bar plot showing expression levels of *Lgr5* in single sorted cells, groups labelled in the panel. (n=3). Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Extended Data Figure 6



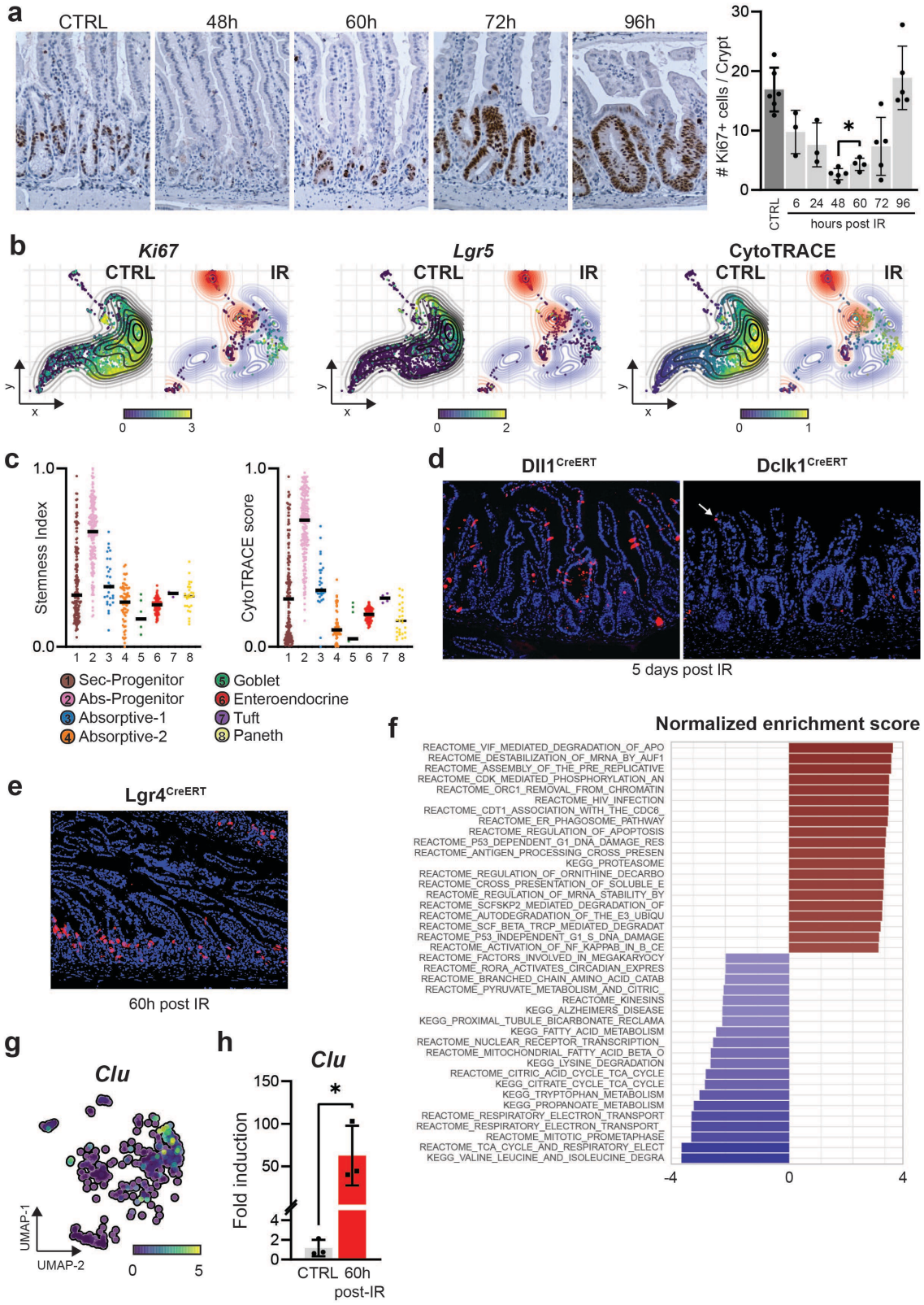
Extended Data Fig. 6: a. Flow cytometry quantification of Lgr5^{eGFP+} cells in purified crypts (n≥3). On the right, representative images of Lgr5^{eGFP} 10 days after the first dose of DT. **b.** Representative images for Dclk1 staining of intestine following DT treatment together with bar plot showing quantification of the number of Dclk1+ cells at indicated time points (n≥4). **c.** UMAP plot showing recovered clusters in DT treated epithelial cells. **d.** UMAP plot showing expression levels of *Lgr5* in the DT treated dataset, on the right bar-plot showing % of detected expressing cells relative to control. Cells are colored based on clusters. **e.** Violin-plot showing computed stemness-index for individual DT-treated crypt epithelial cells. **f.** Representative images for Lgr4^{CreERT};Lgr5^{DTReGFP} lineage tracing at day 3 and 10 post DT exposure. **g.** Bar-plot showing percentage of total tdTomato^{+} cells at indicated time-points (Lgr4^{CreERT}, n≥4). **h.** Schematic representation of the model used. TAM was administered 1 hour prior the first DT injection. After this DT was injected every two days. Below, representative images of intestinal sections for Ki67^{CreERT} lineage tracing at day 9. **i.** Plot showing percentage of body weight loss recorded over the course of the experiment. No significant differences could be detected between the groups (n=3). Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Extended Data Figure 7



Extended Data Fig.7: a. Representative images of Dll1CreERT mice 10 days post TAM induction with or without DT treatment. **b.** Bar plot showing flow-cytometry based quantification of (Dll1^{CreERT}) tdTomato^{+} cells at indicated time points. (n≥3). **c-d.** UMAP and dot- plots showing expression levels of the indicated genes in crypt epithelial cells.

Extended Data Figure 8



Extended Data Fig. 8: a. Immunohistochemistry analysis for Ki67 following IR damage; on the right, bar plot showing quantification of the number of Ki67^{+} cells at indicated time points. (n≥3) **b.** Differential Kernel density estimation (KDE) analysis of irradiated crypt epithelial cells, using a bootstrapped null model for the control. Showing expression for Ki67 (Left) and Lgr5 (middle), as well as computed CytoTRACE scores (right). **c.** Bar-plots showing computed CytoTRACE as well as Stemness Index scores in irradiated crypt epithelial cells. CytoTRACE scores are presented relative to control dataset. **d.** Representative images showing tracing of tdTomato^{+} cells (Dll1^{CreERT} and Dclk1^{CreERT} mouse lines) five days post IR damage. **e.** Representative image of Lgr4^{CreERT} mice 60 hours after IR damage. **f.** Bar-plot showing gene expression-based pathway enrichment analysis**. g.** UMAP plot showing *Clu* expression in irradiated crypt epithelial cells. **h.** qPCR analysis of *Clu* expression in sorted tdTomato^{+} (Ki67^{CreERT} mice) of CTRL and 60h following IR. (n=3). Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$.

Extended data tables:

Table 1: Mouse models used in this study

Table 2: Reagents and antibodies

Table 3: Gene expression 3-cluster Solution

Table 4: Protein activity 3-cluster Solution

Table 5: Gene expression 8-cluster Solution

Table 6: Protein activity 8-cluster Solution

Table 7: Protein activity vs CytoTRACE score correlation

Table 8: Gene expression vs CytoTRACE score correlation

Table 9: Gene expression signature (IR vs CTRL)

Table 10: Protein activity signature (IR vs CTRL)

Methods:

Animal studies: All animal studies were carried out in compliance with the National Institutes of Health guidelines for animal research and approved by the Institutional Animal Care and Use Committee of Columbia University.

All transgenic mouse lines used in this project are listed in (Extended Data Table 1) together with their respective Tamoxifen (TAM) (Sigma, T5648) dosage. Lgr4-DSRED2-CreERT Knock-in and Dll1-ZsGreen-CreERT BAC mouse models were generated here at Irving Cancer Research Center (The Genetically Modified Mouse Models Core) using standard recombineering strategies. Inducible Cre lines were crossed with R26-tdTomato (Strain #:007909), R26-ZsGreen (Strain #:007906), or R26-Confetti (Strain #:017492) reporter allele purchased from Jax (The Jackson Laboratory). TAM was dissolved in corn oil and administered 200uL per oral gavage at indicated time points. For multicolor tracing analysis, using the R26-Confetti allele, a single dose of 6mg/mouse (TAM) was administered. BrdU (Biolegend, 423401) was injected i.p. at 50mg/Kg 6 hours before harvesting. EdU (Sigma, 900584) was injected i.p. at 50mg/Kg, chase time indicated in the figure panels. Diphtheria toxin (DT - EMD Millipore, 322326) was injected i.p. at 50ug/Kg as previously described[4]. For irradiation studies mice received a single dose of 12Gy whole body irradiation using Mark I Cesium-137 based gamma-ray irradiator (J.L. Shepherd & Associates, San Fernando, USA).

Crypt epithelial cell preparation and staining: All reagents and antibodies are listed in Extended Data Table 2. 10 cm of intestinal Jejunum was harvested and cut in small pieces, EDTA (Invitrogen, 15575020) based dissociation was performed as previously described [5] with minor modifications. Following EDTA dissociation crypts were incubated at 37°C for 7 minutes in TrypLE Express (Thermo Fisher Scientific, 12-604-021) and DNase I (Thermo Fisher Scientific, 10104159001) in order to obtain a single cell suspension. Following enzymatic digestion cells were resuspended in Advanced DMEM/F12 (Gibco, 12634-010) supplemented with GlutMAX (Thermo Fisher Scientific, 35050061), HEPES (Thermo Fisher Scientific, 15-630-080), Antibiotic-Antimycotic (A/A - Fisher Scientific, 15240062), and 10% FBS (Gemini Bio-Products, 900-108). Cells were stained for 20 minutes Epcam-APC (1:200, BioLegend 118214), washed and re-suspended in media containing DAPI (1:10'000, BD Biosciences 564907). For B6A6[6] staining, cells were incubated with primary antibody for 30 minutes, following conjugated secondary Ab for 30 minutes (1:200, BioLegend 405406), washed and re-suspended in media containing DAPI.

Flow cytometry analysis and cell sorting: All flow cytometry analyses were performed using a LSRII Fortessa (CCTI Flow core) and results were analyzed using FlowJo. Cell sorting experiments were performed in the Columbia Stem Cell Initiative Flow Cytometry core facility at Columbia University Irving Medical Center under the leadership of Michael Kissner, using a BD FACS Aria II. All flow cytometry quantifications are presented as percentage of Alive/Epcam$^+$ cells following standard gating strategy (Cells/SingleCells/DAPI$^-$Epcam$^+$).

***In vitro* organoids:** Single sorted epithelial cells were resuspended in MEDIA and counted under the microscope. Appropriate volume of cell suspension was mixed to GFR Matrigel (Corning 356231) and plated in a 24-wells plate as 25uL domes, 5000 cells/dome were seeded in all the experiments. ENR media (Advanced DMEM/F12, GlutMAX, HEPES, A/A, B27 (Thermo Fisher Scientific, 17504044), N-2 supplement (Fisher Scientific, 17502048), N-Acetyl-L-cysteine (Sigma-Aldrich, A9165), EGF (Thermo Fisher Scientific, PMG8043), Noggin (PeproTech, 250-38), Rspo1 (R&D Systems, 3474-RS-050) supplemented with CHIR (Sigma-Aldrich, SML1046) was changed every two days. Number of organoids was counted on day 5, each experiment consisted in at least 4 technical replicates and was repeated at least two times.

Tissue processing and Immunostaining: Intestinal jejunal segments were harvested, flushed with DPBS (Fisher Scientific, 14-190-250), cut longitudinally, and rolled around a cotton swab stick (Swissroll orientation). For cryo-preservation, tissue was fixed in 4% PFA (Electron Microscopy Sciences, 15714) overnight following 24h in 30% Sucrose (in PBS), and embedded in OCT. 5uM sections were stained with indicated Ab using standard experimental workflow. For immunohistochemistry tissue was fixed overnight in 10% buffered formalin (VWR, 89370-094) and embedded using standard experimental procedures. All embedding and sectioning was performed in the Molecular Pathology Shared Resource. IHC staining was performed using standard protocols. EdU was imaged using the Click-IT imaging assay (ThermoFisher, C10337).

RNA extraction and qPCR: Single sorted cells (~50’000 cells) were lysed in buffer RLT and stored at -80°C until starting the RNA isolation following manufacturer instructions (Qiagen, 74034). cDNA reaction was performed using qScript™ cDNA SuperMix (QuantaBio, 95048). qPCR analyses were performed using a QuantStudio 3 thermocycler (Applied Biosystems); gene expression is presented using the ΔΔCt method (*18S* or *Gapdh* genes used as housekeeping).

10x scRNAseq and single cell Multiome (ATAC+RNA): Library preparation and sequencing were performed by the JP Sulzberger Columbia Genome Centre (Single cell analysis core), using standard methodologies. For scRNA-seq crypt epithelial cells were sorted to exclude dead cells

and contaminating villi (Alive/Epcam+/B6A6-). Immediately after sorting, cells were counted using an automated cell counter (ThermoFisher Countess II FL) to check viability and processed for library preparation (10x chromium). For single cell Multiome, immediately after sorting, isolated cells (>100'000) were processed to extract nuclei following manufacturer recommendation (protocol CG000365, Rev B). Isolated nuclei were counted and quality control using an automated cell counter (ThermoFisher Countess II FL) and immediately processed for library preparation.

Single-cell analysis of mouse intestinal crypt cells: Single-cell RNA-seq (scRNA-Seq) UMI profiles were processed using Seurat (v.4.1.0)[7]. Cells with >1,000 expressed genes and mitochondrial gene content < 10% were retained for downstream analysis, yielding to 3,656 cells. UMI counts were normalized and scaled using SCTtransform from the Seurat package[7]. Next, a Shared Neighbors Graph (SSN) was built with knn=10 to select cells with most similar transcriptional profiles and merge them to generate high resolution ensembles of cells called metacell: this approach augments the number of detected genes per cells, which usually is very low due to dropout technological bias (<20%), thus increasing the number of targets that can be recovered by reverse-engineering regulatory networks. Metacell profiles were computed on normalized data, but merged into UMI counts and transformed to count per million (cpm) for downstream analysis. Cell doublets were identified using scanpy's implementation of scrublet[8].

Reverse-engineering of mouse intestinal crypt regulatory networks: An intestinal stem cell (ISC)-specific regulatory network (interactome) was reverse engineered from the resulting metacell cpm profiles (n = 1,218) using ARACNe-AP[9], the most recent implementation of the ARACNe algorithm[10], with 200 bootstraps, a Mutual Information (MI) P-value threshold P ≤ 10-8, and Data Processing Inequality (DPI) enabled. A total of n = 2305 regulatory proteins (RP) were selected into manually curated protein sets, including n = 1465 Transcription Factors (TF) and n = 840 co-Transcription Factors or chromatin remodeling enzymes, using the following Gene Ontology (GO) identifiers: GO:0003700 and GO:0003712[11,12]. The resulting network includes 1,797 regulators, 14,935 targets and 548,442 interactions. Epithelial crypts scRNA-Seq profiles were transformed to protein activity profiles using the metacell-derived regulatory network and the VIPER algorithm [13]. To avoid bias due to different regulon sizes, regulons were pruned to include only the 50 highest likelihood targets, as recommended in[13], and regulons with < 50 targets were excluded from the analysis. Next, we sought to recover cell identities by identifying clusters of cells that share the same regulatory program using the Louvain clustering algorithm applied on the protein activity profiles of cells. Using these profiles, a SSN was built with knn=15 using the first 6 Principal Component (PC) as identified by the Elbow method. We performed a grid search

analysis to tune Louvain's resolution parameter to maximize the average of within-cluster Silhouette scores across each candidate optimal clustering solution. A high Silhouette score is an indication that clustered cells have homogenous profiles, hence as sampled from the same cell population. The optimal solution yielded 3 major clusters and differential markers analysis identified the two major intestinal lineages, secretory and absorptive, with the third population appearing to correspond to stem/progenitors, based on cell potency analysis (See Cell Potency Inference paragraph) For each one of the 3 clusters, a lineage-specific regulatory network was reverse-engineered as explained above and used to recover ISC cell identities.

Lineage-specific protein activity analysis and identification of cell identities: Lineage-specific protein activity and clustering analysis was performed as follows. Crypt epithelial cells were computationally isolated in three distinct datasets based on their inferred lineage as identified at the previous step. For each lineage, a specific gene expression signature (GES) based on the relative subset of cells was computed by scaling the expression data after having normalized and variance-stabilized the UMI counts matrix using *sctransform* as implemented in the Seurat R package. Next, VIPER analysis was performed using cluster-specific regulatory networks and the lineage-specific GES as computed at the previous step. Clustering analysis was performed in a lineage-specific manner by re-running the grid search analysis on the isolated cells protein activity matrix. The absorptive lineage was divided in two clusters, the secretory lineage in three clusters and the stem/progenitor one in two clusters (see Extended Data). The cells from the whole crypt epithelial cell dataset were then labeled based on these seven clusters. To generate the final matrix of protein activity, a GES was generated by comparing each crypt epithelial cell to the average of all the cells independently of their lineage, and VIPER was run on each of the three lineage-specific regulatory networks for all the cell. For each cell, only the VIPER run with the best matching network was retained, as selected by enrichment of top 50 cell-specific candidate MRs. This ensures that each cell, especially those in transition, has been analyzed with its best matching lineage-specific network, and, that each cell is comparable with each other across the dataset because the reference distribution for the GES has been computed across all the crypt epithelial cells present in the dataset. For the RPs that were lineage specific, we used the VIPER score from the lineage-specific regulatory network analysis, while protein activity of RPs that were not present in one of the networks (e.g., due to their lineage specific activity) were computed using metaVIPER across all the networks [14].

Display of crypt epithelial cells as scatter plot: For the purpose of displaying crypt epithelial cells on a 2-axis plane, we performed the following operations on the final VIPER matrix.

Specifically, a KNN graph of the data was built with k=15 using n=30 Principal Components (PCs). Next, we ran the partition-based graph abstraction (PAGA)[15] algorithm on the data using the clustering solution we identified as input partition, to provide an informative initialization for the UMAP algorithm to increase its accuracy by initializing its embedding[16]. The UMAP embedding as computed with this protocol on protein activity data was then transferred to the gene expression matrix, along with information of clustering labels to generate the plots in the manuscript. Specifically, for DT-ablated and IR datasets, we performed asymmetric data integration with the *ingest* methodology included in the *scanpy* framework[17], using CTRL dataset as reference.

Analysis of publicly available intestinal epithelial crypt cells datasets: Dataset as UMI counts from *Ayyaz, et al.* (2019) and *Böttcher, et al.* (2021) where downloaded from GEO. For each dataset, GES was computed using an internal reference in each dataset independently, and metaVIPER was run using all the networks from the three lineages identified from the CTRL dataset. Cluster assignment on each cell was performed by enrichment analysis using top 50 candidate MR proteins for each of the 7 clusters. Specifically, a cell was assigned to the cluster for which it had the highest enrichment score.

Cell Potency Inference: Cell potency inference was performed using three distinct approaches. The first one, the CytoTRACE algorithm, relies on gene expression data and leverages information embedded in the number of detected genes per cell[1], to compute the CytoTrace Score (CT) that we used to sort cells from the least differentiated to the terminally differentiated ones. The second approach we used builds a one-class logistic regression (OCLR) using pluripotent stem cell bulk samples (ESC and iPSC) as a predictive model to compute a Stemness Index (SI) on new samples, that we used on protein activity profiles[2]. The third approach, SCENT, uses protein-protein interaction networks (PPIs) and single-cell transcriptomic data to infer signaling entropy scores (SR) to use as proxy of cell differentiation potency, with the hypothesis that a cell with high signaling entropy should be endowed with higher differentiation potential[18]. PPIs were modeled using PrePPI[19], a large-scale database of human PPI, by retaining the top 5% of high confidence interactions. To convert murine gene products to human, for PrePPI network, gene identifiers were mapped to their human-mouse orthologs using R biomaRt services. To compute CT scores across pre- and post-radiation cells, we performed bootstrap analysis by subsampling 100 times the sham dataset with the same number of cells we recovered after irradiation, in order to make them comparable. The mean CT score per cell was used for the final analysis to compare cell potency in the post-radiation sample.

Stem Cell Markers Discovery: Markers associated with high cell potency were inferred by correlating regulatory protein activity with CT scores. RP were prioritized based on the Pearson's correlation coefficient p-value after correction for multiple hypothesis testing using the Benjamini-Hochberg method.

Joint analysis of CTRL and IR-exposed crypt epithelial cells: To compare CTRL and IR datasets in a quantitative manner we build a null model of cell density estimates with a bootstrap approach and subsampling, to take also into account the difference of about an order of magnitude in sample size between CTRL (n~3,500) and IR (n~600). Specifically, differential Kernel Density Estimate (KDE) was performed in the following way. Pre- and post-radiation samples were normalized independently using sctransorm with method glmGamPoi. Next, the two datasets were joined together by identifying common anchors using reciprocal PCA as implemented in using Seurat (v.4.1.0) [7]. metaVIPER analysis was performed using the three lineage-specific networks and the scaled integrated gene expression matrix as cell specific expression signature. Next, UMAP was performed on the protein activity matrix of the joined datasets using Euclidean as distance metric and 30 PCs. The first two UMAP dimensions were used to lay pre- and post-radiation cells on a bi-dimensional pane of 1e4 equally-sized tiles addressed by 100 intervals per UMAP dimension. The pre-radiation dataset was used to bootstrap 100 times a subset of 500 cells that were uniformly sampled to create a distribution of density estimates using a Gaussian kernel. To assess cell depletion or enrichment for each tile in the post-radiation sample, we used the positive or negative z-score computed using mean and standard deviation of the control (pre-radiation sample).

Archetype analysis: Archetype analysis was performed in R using the *archetypes* package. Briefly, the VIPER matrix was analyzed using classical archetype analysis for every number of *k* archetypes between 2 and 12. The elbow method allowed to determine that k=5 is the optimal number of archetypes (see Extended Data).

Paneth cells analysis: Paneth cells were identified in crypt epithelial cells by looking at the concurrent expression of all the following markers: *Lyz1*, *Mmp7*, *Dll4*, and *Nupr1,* calling it Paneth cell when all these markers were found expressed at any degree. This is a lower bound estimation of Paneth cells, given technical dropouts. Paneth cell specific candidate MR proteins were identified by selecting Paneth cells, generating a GES using as reference all the other non-Paneth cells, and running msVIPER using the secretory lineage regulatory network.

scATAC-Seq analysis: scATAC-Seq data was processed using ArchR[20]. Peak calling was performed using MACS2 with default parameters[21]. Cluster analysis was performed using the RNA data modality by projecting the clustering solution identified on the CTRL sample using *ingest* over scanpy[17]. Known motifs enrichment analysis was performed using the *cisbp* motifs dataset and the 7-cluster solution identified over the VIPER-transformed data.

Data availability: scRNAseq datasets together with single cell Multiome (scRNA+scATAC seq) of crypt epithelial cells will be made available before publishing these results.

Code availability: All code will be available on Github or upon request at the moment of publication.

Extended data references

1 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411, doi:10.1126/science.aax0249 (2020).

2 Malta, T. M. *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354 e315, doi:10.1016/j.cell.2018.03.034 (2018).

3 Yan, K. S. *et al.* Non-equivalence of Wnt and R-spondin ligands during Lgr5(+) intestinal stem-cell self-renewal. *Nature* **545**, 238-242, doi:10.1038/nature22313 (2017).

4 Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255-259, doi:10.1038/nature10408 (2011).

5 Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265, doi:10.1038/nature07935 (2009).

6 Wang, F. *et al.* Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* **145**, 383-395 e381-321, doi:10.1053/j.gastro.2013.04.050 (2013).

7 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

8 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289, doi:10.1016/j.cels.2018.11.005 (2019).

9 Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233-2235, doi:10.1093/bioinformatics/btw216 (2016).

10 Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382-390, doi:10.1038/ng1532 (2005).

11 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

12 Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7, doi:10.1186/1471-2105-7-S1-S7 (2006).

13 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838-847, doi:10.1038/ng.3593 (2016).

14 Ding, H. *et al.* Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat Commun* **9**, 1471, doi:10.1038/s41467-018-03843-3 (2018).

15 Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59, doi:10.1186/s13059-019-1663-x (2019).

16 Kobak, D. & Linderman, G. C. Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat Biotechnol* **39**, 156-157, doi:10.1038/s41587-020-00809-z (2021).

17 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

18 Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun* **8**, 15599, doi:10.1038/ncomms15599 (2017).

19 Zhang, Q. C. *et al.* Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* **490**, 556-+, doi:10.1038/nature11503 (2012).

20 Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411, doi:10.1038/s41588-021-00790-6 (2021).

21 Gaspar, J. M. Improved peak-calling with MACS2. *BioRxiv*, 496521 (2018).