Isthmal stem cells sustain intestinal homeostasis and regeneration

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

[1] Division of Digestive and Liver Diseases Medicine and Irving Cancer Research Center, Department of Medicine, 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] Department of Cell, Developmental & Cancer Biology, Oregon Health & Sciences University 3181 SW Sam Jackson Park Road, L215, Portland, USA.

[6] 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.

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

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

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

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

[11] 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 equates crypt base columnar (CBC) cells, marked by high levels of *Lgr5* expression, with the intestinal stem cell (ISC)[1]. However, recent intestinal regeneration studies have uncovered limitations of the 'Lgr5-CBC' model[2,3], leading to two major views: one favoring the presence of a quiescent reserve stem cell population[4-7], the other calling for differentiated cell plasticity[8-11]. To test if an alternative model may help reconcile these perspectives, we studied the hierarchical organization of crypt epithelial cells in an unbiased fashion, by combining high-resolution, single-cell profiling and lineage tracing in multiple transgenic mouse models. These show that *Lgr5* is not a specific ISC marker; rather, cells located in the crypt isthmus, which include Lgr5^{low} cells, comprise the ISCs that sustain tissue homeostasis. Following irradiation or intestinal injury, surviving ISCs and progenitors, but not differentiated cells, participate in intestinal regeneration, suggesting that neither de-differentiation nor reserve stem cell populations are drivers of intestinal regeneration. Our results provide a novel viewpoint for the intestinal crypt epithelium, in which ISCs localize to the crypt isthmus, and ISC potential is restricted to stem and progenitor cells.

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[3-7,12,13]; among others, *Lgr5* has been widely accepted as the primary ISC marker[*14*], 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[14]. However, recent studies on intestinal regeneration following high doses of radiation[3,7,14], with selective *Lgr5*-expressing cell ablation[2], have shown that other crypt epithelial cells can compensate for the loss of Lgr5$^+$ cells. These observations suggest that the ‘Lgr5-CBC’ model may not effectively 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[2-4], the second asserting that, following ISC loss, multiple differentiated cell types may undergo de-differentiation and serve as *bona fide* ISCs[8-11].

Single cell transcriptomic has emerged as a powerful tool to study tissue heterogeneity and to uncover the identity of rare cell populations[15,16]. With respect to the intestine, this approach has been instrumental in inventorying the specific differentiation states of diverse intestinal lineages, as well as to investigate the cellular composition of the regenerating intestine[4,17-19]. 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[1].

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

To address these questions, we combined single cell profiling with novel lineage tracing mouse models to study intestinal crypt cells and then analyzed their hierarchical organization using an unbiased, data-driven approach that does not rely on established markers. We find that *Lgr5* overlaps with but is not uniquely expressed in stem and progenitor cells. Moreover, we provide evidence that the source of intestinal stemness responsible for homeostasis and regeneration lies within the isthmus compartment. Based on these results, we propose a novel model of intestinal organization in which many actively cycling cells in the crypt isthmus are uncommitted and include multipotent, self-renewing ISCs. Following intestinal injury, surviving isthmal stem cells guide intestinal regeneration, linking them to the postulated reserve ISCs, and suggesting that de-differentiation constitutes an exceptional event rather than a main regenerating intestine driver.

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). High 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 an established network-based computational pipeline designed to measure the transcriptional activity regulatory and signaling proteins from scRNA-seq profiles, using the VIPER algorithm[26]. Specifically, VIPER measures a protein's activity based on the differential expression of its transcriptional targets (*regulon*), with the latter identified by the ARACNe algorithm[27,28]. This allows quantitative activity assessment of > 6,500 proteins—including transcription factors, co-factors, signaling proteins, and surface markers—with virtually no dropouts. Critically, we have shown that this methodology allows identification of subpopulations that are essentially undetectable by gene expression analysis[15,16], robust, highly reproducible quantification of protein activity, comparing favorably with antibody-based single cell measurements[15], and optimal 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 from protein activity-based 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 progenitor cells (*Stmn1*^{hi}, *Dek*^{hi}, etc.)[1,17,19] (Extended Data Fig.1b-e, Table 3-4).

To uncover lineage-specific sub-populations, we performed independent, protein activity-based cluster analysis of each one of these three major subpopulations (Fig.1b-c, Extended Data Table 5-6). Notably at the single cell level, the vast majority of population-specific proteins—including established intestinal lineage markers such as Fev[31], and Gfi1b[1]—could be detected only by VIPER-based protein activity analysis but not based on the expression of their encoding genes (Extended Data Fig.2). The refined cluster analysis revealed two absorptive clusters, corresponding to subpopulations of early specified and more committed cells, respectively; moreover, it revealed three established secretory subpopulations, including Tuft (*Dclk1*+), Enteroendocrine (*NeuroD1*+), and Goblet (*Atoh1*+) cells[1], with *Lyz1*^{hi} Paneth cells found mainly within the Goblet cluster, possibly due to low representation in this dataset (Extended Data Fig.3a). Finally, stem and progenitor cells segregated into two clusters corresponding to subpopulations that, intriguingly, presented similar levels of inferred potency, suggesting that this subdivision does not reflect a hierarchical partition of these cells (Fig.1d, Extended Data Fig.3b). Rather, the highest level of stemness potential was detected at the boundary between these two (Fig.1d), suggesting that these two clusters reflect a gradual yet progressive differentiation of progenitor cells towards a secretory (Sec-Progenitor) or absorptive (Abs-Progenitor) fate, respectively.

Although the presence of a postulated rare ISC subpopulation remains elusive[32], these analyses provide a unique opportunity to study the molecular features that best describe the ISC compartment, as defined by the highest cell potency level. To this aim, we characterized the regulatory factors that best associate with cell potency by correlating the activity profiles of regulatory proteins, as assessed by VIPER, with CytoTRACE-based analysis[20] of individual cells (Fig.1e, Extended Data Fig.4, Table 7). Regulatory proteins, whose activity presented the most significant correlation with CytoTRACE scores were highly enriched for factors known to regulate stemness across different tissues, including the intestine[33-39]. Consistently, known regulators of intestinal differentiation[1] had negative correlation scores.

Among others, the transcription factor Yy1, was previously reported to play an essential role in preventing ISC differentiation and to impair epithelial cells organoids forming capacities upon genetic knock-out[37]. Several chromatin remodeler enzymes also emerged from the analysis—including Smarca5[34] and Atad2[35], which have been previously associated with stemness—thus highlighting the crucial role of chromatin remodeling in cell potency. Interestingly, several splicing factors also emerged within the broad group of transcriptional regulators associated with cell potency, including members of the Srsf family, which were recently shown to play an important role in regulating the intestinal epithelium[38].

Focusing on genes, whose differential expression correlated with inferred cell potency (Extended Data Table 8), we identified *Smarca5*, *Stmn1* and *Dek*, all representing previously proposed markers of adult stem cells[33,40]. Birc5a[41] and Ncl[42], both described as highly expressed in embryonic stem cells, also emerged among the most correlated genes. Finally, *Fgfbp1*, which is expressed in cells capable to lineage trace the intestine (*Personal communication*), was also positively correlated with inferred cell potency.

These analyses provide the first unbiased characterization of intestinal crypt epithelial cells using mechanism-based dissection of protein activity in single-cell data, via regulatory network analysis. Moreover, they provide novel insights on the factors that implement and maintain the transcriptional state of individual intestinal populations, via their regulatory targets, and highlight those that best recapitulate intestinal cell potency.

Lgr5 expression is not restricted to intestinal stem and progenitor cells: Lgr5 has been described as a highly specific ISC marker[14]. Based on Lgr5 reporter allele and CreER lineage tracing assays, Lgr5$^+$ cells were postulated to correspond to crypt base columnar (CBC) cells in the +1 to +3 region[14]. Surprisingly, however, our analyses did not reveal a statistically significant correlation between Lgr5 expression and inferred cell potency (Extended Data Fig.5a). Specifically, cells presenting the highest *Lgr5* expression levels were not associated with the highest stemness potential, as assessed from CytoTRACE analysis; rather, the highest potency was detected in some Lgr5^{low} cells (Fig.2a). Importantly, similar considerations could be drawn for other genes previously proposed as ISC-specific markers (Extended Data Fig.5b-c)[5-7,12,13].

To better characterize the *Lgr5* expression pattern, we generated additional scRNA-seq profiles from highly purified crypt epithelial cells, harvested from an Lgr5$^{DTR\text{-}eGFP}$ mouse[2] (Fig.2b). By mapping the *DTR-eGFP* allele on sequencing data, we were able to study expression profiles of both Wild type (Wt) and Transgenic (Tg) Lgr5 alleles in purified crypt epithelial cells (Fig.2c). Quantification of the number of cells expressing Wt or Tg alleles revealed virtually complete overlap, confirming that nearly all *Lgr5* expressing cells were also positive for *DTR-eGFP* (Fig.2c). Intriguingly, *Lgr5* expression was detected in about 20% of all sequenced crypt cells in our datasets (20.495±1.8, % of total), suggesting a much broader domain of expression then previously appreciated[14]. Flow cytometry analysis of the crypt epithelium consistently revealed an average of 20% Lgr5^{eGFP+} cells (19.96±3.96 % of Epcam$^+$ cells), thus confirming the faithful activity of the reporter allele, which was not confined to the CBC population (Fig.2d).

We next analyzed the identity of *Lgr5* expressing cells within our datasets. Most of *Lgr5*^{hi} cells were detected in the Sec-Progenitor-1 subpopulation (Fig. 2e), which we annotated as secretory-biased progenitors. Interestingly, previous studies have observed direct specification of secretory cells from Lgr5$^+$ ones[1,18]. In line with these observations, we detected a small fraction of *Lgr5* expressing cells that were also positive for known secretory markers, such as *Atoh1* and *Dll1*[1] (Extended Data Fig.5d). Analysis of a published dataset[43] confirmed expression of multiple secretory-associated genes in Lgr5^{eGFP+} sorted cells (Extended Data Fig.5e).

Surprisingly, we also observed high levels of *Lgr5* expression in differentiated Tuft cells, marked by *Dclk1* expression (Fig.2e, Extended Data Fig.5e)[1,44]. Dclk1 staining validated its overlap with Lgr5^{eGFP+} cells, both within the crypt as well as the villi region (39.2±9.5 % of Dclk1^{+} cells; Fig.2f**,** Extended Data Fig.6a). Moreover, sorted Dclk1$^{ZsGreen+}$ cells[45] confirmed expression of *Lgr5* and *Ascl2*[46], both proposed ISC markers, in differentiated Tuft cells (Extended Data Fig.6b-d). Taken together, these data show that *Lgr5* expression is not restricted to intestinal stem and progenitor cells but also spans differentiated subpopulations, including Tuft cells. Furthermore, they suggest that among *Lgr5* expressing cells, only a subset of Lgr5^{low} cells are characterized by high levels of inferred stemness.

Isthmal stem cells sustain intestinal tissue homeostasis: Lgr5$^{eGFP-CreERT}$ mice show clear labelling of a subset of crypt epithelial cells exhibiting long term lineage tracing (ISCs)[14]. Accordingly, this mouse line was initially reported to almost uniquely label CBC cells, marked by high levels of Lgr5 expression[14]. However, our analysis suggests that only a fraction of Lgr5^{low} -expressing cells is characterized by high cell potency (Fig.2a). We therefore investigated whether the Lgr5$^{eGFP-CreERT}$ mouse could label cells other than the Lgr5^{Hi} CBC population. Staining for eGFP and Cre recombinase revealed a similar distribution along the crypt epithelium, suggesting functional expression of the Tg allele above the crypt base in the isthmus region, where Lgr5^{low} cells are located (Fig.2d, Fig.3a). In line with this observation, following Tamoxifen (TAM) induction, scattered tdTomato^{+} cells could be consistently observed around the +4/+6 isthmus region (Fig.3b Extended Data Fig.7a), raising the possibility that long term lineage tracing may originate from Lgr5^{low} cells located at the crypt isthmus.

To investigate the contribution of CBC cells versus isthmus cells to long term lineage tracing, we studied the labelling of Troy^{CreERT} mice, as *Troy* has been described as an additional CBC specific marker[47]. Following TAM induction, tdTomato^{+} cells could be found near the crypt base, but again were not confined to the CBC population (Fig.3c, Extended Data Fig.7a). Comparison of position-based labelling revealed a greater number of non-CBC (+4 or higher) tdTomato^{+} cells in Troy^{CreERT} mice relative to Lgr5$^{eGFP\text{-}CreERT}$ (Fig.3d). Over time Troy^{CreERT} labelled cells traced significantly more than Lgr5$^{eGFP\text{-}CreERT}$ labelled cells (66.28%±14.75% vs 37.05%±8.37% at 6 months) (Fig.3e-f). Importantly, at day 1, both mouse models had a similar number of initially labelled cells (fraction of tdTomato^{+}: Lgr5$^{eGFP\text{-}CreERT}$ 7.94T%±1.58% vs Troy^{CreERT} 8.23%±2.82% at day 1) (Fig.3f), suggesting that cells higher up in the crypt, which were over-represented in Troy^{CreERT} mice, were largely responsible for long term lineage tracing.

Located immediately above the crypt base, in the crypt isthmus, there is a domain of highly proliferating cells, often referred to as transit amplifying (TA) progenitors[14]. Since our analyses suggest that ISCs are actively cycling (Extended Data Fig.7b) and staining for proliferative markers showed that a significant fraction of +4/+6 *Troy* expressing cells were indeed proliferating (Extended Data Fig.7c), we decided to test whether ISCs may localize to the crypt isthmus. For this purpose, we studied the contribution of proliferating cells, which are more abundant in this compartment, to long term lineage tracing using a novel mouse model (Lgr4^{CreERT}) that targets inducible Cre to *Lgr4* expressing cells (Extended Data Fig.7d), previously reported to mainly comprise intestinal progenitors[48]. Consistent with our expectation, following TAM induction, a majority of labelled cells were actively cycling (Extended Data Fig.7e) and localized to the +4/+10 position within the crypt (Fig.3g, Extended Data Fig.7f). Similarly, analysis of the Ki67^{CreERT} mouse model[49] confirmed efficient labelling of cycling cells in the isthmus region with little overlap (≤10% labelled cells at +1/+3 in both mouse models) with CBC cells (Fig.3g, Extended Data Fig.7f). Over time, both models gave rise to tracing ribbons, which persisted for up to six months, thus demonstrating labelling of some ISCs (Fig.3h-i).

Based on these results, we reasoned that the subset of *Lgr5* expressing cells that were also labelled in Lgr4^{CreERT} and Ki67^{CreERT} mice would be enriched for ISCs. Therefore, we analyzed their specific overlap by crossing the Lgr4^{CreERT} and Ki67^{CreERT} mice to the Lgr5$^{DTR\text{-}eGFP}$ reporter allele[2]. Many double positive cells could be observed above the CBC region, in the crypt isthmus, and qPCR analyses confirmed enrichment for Lgr5^{low} expressing cells in the double positive population (Extended Data Fig.7g-h). In line with these observations, analysis of histological sections revealed that some tracing ribbons originated at the level of the crypt isthmus, from +4/+6 positions, and persisted for a month or more in the absence of CBC labelling (Fig.3j, Extended Data Fig.7i). At 6 months post TAM induction, most units appeared either fully labelled or negative except for the occasional exclusive presence of Paneth cells (Fig.3h), in line with previous proposed dynamics of crypt clonality[50,51]. Taken together, these results support the hypothesis that ISCs are located within the isthmus compartment and tend to express low levels of *Lgr5* (Fig.3k). Taken together, such observations, which are fully consistent with our computational analyses, offer a new perspective on intestinal epithelial cells hierarchical organization.

***Lgr5^{neg}* isthmal cells compensate for loss of Lgr5 expressing cells:** Our data suggest that active cycling isthmus cells, some expressing low levels of *Lgr5*, comprise the ISCs responsible for maintaining long term tissue homeostasis. This is consistent with multiple studies showing that selective ablation of *Lgr5* expressing cells does not impact normal tissue turnover. In particular, two models of interpretation have been proposed to explain such observations, whereby Lgr5^{+} cells repopulation is driven by either (a) de-differentiation of terminally differentiated cells[8-11], or (b) activation of a reserve quiescent ISC[2,4].

To clarify the dynamics of this process we analyzed the Diphtheria Toxin (DT) ablation model using the Lgr5^{DTReGFP} mouse line[2]. Flow cytometry analysis confirmed consistent, complete elimination of Lgr5$^{DTReGFP+}$ cells after two doses of DT without compromising the intestinal mucosa; this was followed by the later reappearance of *Lgr5* expressing cells, returning to control values within 10 days (Fig.4a, Extended Data Fig.8a). Of note, the disappearance of Lgr5^{+} cells coincided with a marked reduction of Dclk1^{+} Tuft cells (Extended Data Fig.8a), corroborating their expression of *Lgr5* and aligning well with previous reports[52].

To gain additional insights into the response to Lgr5$^+$ cell ablation, we profiled the transcriptome of purified crypt epithelial cells at 24h after two doses of DT; this represents a time point when most Lgr5$^+$ cells are ablated (% of Lgr5-eGFP$^+$ cells: 0.44±0.85 vs 20.52±1.94 - Fig.4b) and regeneration has begun. ScRNA-seq profile analysis revealed that the overall cellular composition was intact, despite a marked reduction in *Lgr5* expressing cells (Fig.4c, Extended Data Fig.8c), consistent with the observation that Lgr5$^+$ cells are largely dispensable for crypt epithelium regeneration[2]. Importantly, these analyses did not show expansion of novel cell populations that may be ascribed to activation of a reserve ISC. Joint CytoTRACE analysis of CTRL and DT-treated samples revealed dramatic reduction of high potency cells, in the top decile of high potency scored cells, following DT treatment (Fig. 4d), corroborating the notion that *Lgr5* is expressed in some stem and progenitor cells. Nevertheless, a consistent fraction of high inferred stemness cells persisted at this time point, suggesting that some Lgr5^{neg} cells may possess the potential to serve as ISCs to support tissue homeostasis.

Similar to control, following DT treatment, high cell potency was confined to stem and progenitors, suggesting that ISC potential is restricted to these populations. To test this hypothesis, we first analyzed the behavior of Lgr4^{CreERT} labelled cells following DT treatment. TdTomato$^+$Lgr5$^{DTReGFP-}$ cells could be observed scattered across the crypt epithelium and expanded over time giving rise to new TdTomato$^+$Lgr5$^{DTReGFP+}$ cells at day 10 (Fig.4e, Extended Data Fig.8d). Next, we tested whether differentiated cells may also participate in Lgr5$^+$ cell repopulation alongside isthmal proliferating cells, thus reflecting the proposed broad plasticity of intestinal epithelial cells[11]. For this purpose, we generated a novel mouse model that targeted inducible Cre to *Dll1*-expressing cells (Dll1^{CreERT}, Extended Data Fig.9a). Dll1^{CreERT} mice efficiently label committed secretory cells that overlap only minimally with proliferative progenitors and show limited organoid forming capacity (Extended Data Fig.9b-d). Upon DT treatment, Dll1^{CreERT} labelled cells did not expand, rather labelled cells appeared as Paneth cells at day 10, suggesting that differentiated secretory cells do not contribute to Lgr5$^+$ cell repopulation (Fig.4f, Extended Data Fig.9e). A possible explanation for the discrepancy with previous reports[8,10] could be attributed to the higher specificity of our newly generated mouse line relative to the previously reported Dll1^{eGFPCreERT}, which shows lineage tracing events also in homeostatic conditions [8]. Indeed, Dll1^{low} cell can be detected broadly within the stem/progenitor compartment, including with some *Lgr5*^{low} cells (Extended Data Fig.5d). Taken together, these data show that, upon loss of *Lgr5* expressing cells, isthmal Lgr5^{neg} cells have the capacity to act as ISCs and to support tissue turnover. Furthermore, they show that regenerative potential is restricted to stem and progenitor cells and that de-differentiation or activation of a reserve quiescent-ISC are not the source of Lgr5$^+$ cell repopulation in this model (Fig.4g).

Surviving isthmal stem cells regenerate the intestinal epithelium following IR damage: As discussed, analysis of CTRL and DT-ablated crypt epithelial cells revealed neither the presence of a putative reserve ISC population nor any sign of active de-differentiation. Rather, it supports a model where some Lgr5^{neg} isthmal cells retain ISC potential and sustain intestinal tissue turnover. To determine whether this may be restricted to the ablation model in which only *Lgr5*-expressing cells are perturbed, we assessed whether intestinal regeneration after lethal irradiation (IR) follows similar cellular dynamics. Damage due to high dose IR exposure can induce gastrointestinal syndrome, with highly proliferative cells and *Lgr5*$^+$ CBCs representing subpopulations more likely to be susceptible to irreversible IR damage[14].

To pinpoint the identity of cells with regenerative potential, we studied crypt epithelial cell composition after exposure to 12 Gy whole body IR (WBI), by scRNA-seq analysis. First, to identify the earliest stage of intestinal regeneration, we characterized the dynamics of intestinal proliferation following IR damage (Extended Data Fig.10a). We postulated that at this time point, regenerating stem cells would be undertaking their first or second round of cell divisions, marked by an increase in Ki67 labeling. We identified 60 hours post IR as the earliest time point when proliferation increases, reflecting the first regenerative wave and the focal point for our analysis.

Following IR damage, surviving cells aligned well with previously identified clusters and no novel subpopulations emerged (Fig.5b). As expected[14], proliferative progenitors as well as *Lgr5* expressing cells were largely depleted at this time point (Extended Data Fig.10b). Similar to the ablation model, surviving differentiated cells retained low levels of inferred cell potency (Extended Data Fig.10c), and evaluation of Dll1$^{\mathrm{CreERT}}$ lineage tracing corroborated the absence of *Dll1*$^{+}$ secretory cell participation in intestinal regeneration (Extended Data Fig.10d).

In line with our hypothesis, we observed a small fraction of surviving cells characterized by high levels of inferred cell potency and expression of proliferative markers (Fig.5c), making them the most suitable candidates for the regenerating ISCs. Intriguingly, regenerating and homeostatic ISCs had a high degree of overlap (Extended Data Fig.10e), suggesting that those cells may correspond to surviving ISCs or early progenitors. To clarify the nature of these cells, we lineage traced intestinal isthmal cells following high doses of IR. Sixty hours after IR exposure, surviving tdTomato^{+} cells (Lgr4$^{\mathrm{CreERT}}$) could be observed within the damaged epithelium (Fig.5D, Extended Data Fig.10f), thereby demonstrating that not all proliferative cells are lost following IR damage. Moreover, Ki67 staining (Fig.5d) and lineage tracing analysis at 5 days post IR (Fig.5e) confirmed active participation of surviving tdTomato^{+} cells in intestinal regeneration.

We next sought to study the regulatory programs that characterize regenerating ISC. For this purpose we searched for features unique to cells with highest inferred cell potency pre- and post-IR (Fig.5f-g, Extended Data Fig.11a, Table 9-10). Pathway analysis showed that regenerating ISCs activate cell cycle damage checkpoints and upregulate multiple damage response factors[53-55], reflecting the extended damage generated by IR exposure. This was further corroborated by analysis of the regulatory proteins that define the regenerating ISC compartment (Fig.5g). Apex1 and Baz1a, both known to be involved in DNA damage response[56,57], were among the most statistically significant activated proteins. Moreover, regenerating ISCs expressed high levels of *Ly6a* (Sca1), as well as *Areg* and other proposed intestinal regeneration markers[58] (Fig.5f). Interestingly, we observed upregulation of *Clu* (Fig.5f, Extended Data Fig.11b), whose expression has been proposed as specific to a quiescent radio-resistant stem cell population (revival ISC[4]).

These results raise the possibility that regenerating ISCs may represent an expanding reserve stem cell subpopulation. To clarify the nature of these cells, we analyzed *Clu* expression in traced intestinal proliferating cells (Ki67^{CreERT}) pre- and post- IR damage. Results showed *Clu* upregulation in surviving tdTomato+ cells sixty hours post IR exposure (Extended Data Fig.11c), providing evidence that following IR damage, the 'revival' ISC correspond to surviving cycling cells rather than a separate quiescent stem cell population. Lastly, although we were unable to discriminate the relative contribution of pre-existing isthmal stem cells or early progenitors, our analyses strongly suggest that the potency to serve as regenerating ISCs is restricted to these populations (Fig.5h).

Discussion

Here we provide an unbiased characterization of intestinal crypt epithelial cells in homeostasis and regeneration. We find that ISCs are located in the isthmus, the crypt region previously thought to accommodate only TA progenitors. Furthermore, we provide evidence that intestinal regeneration arises from surviving stem and progenitor cells.

Our optimized isolation protocol—using the B6A6 Ab[24] to exclude villi cells—ensured high enrichment of the crypt epithelium and reproducibility across all high-throughput datasets, especially for the injured intestine where loss of tissue integrity greatly affect the isolation strategy. Furthermore, our computational pipeline integrated multiple algorithms to unbiasedly recover cell identities and study their hierarchical organization based on single cell transcriptomic and epigenetic profiles. Compared to previous studies, the results obtained show a high degree of similarity[17,18] yet provide improved molecular resolution on intestinal stem and progenitor cells. Moreover, the proposed regulatory network-based classification schema offers novel insights on the mechanistic determinants of intestinal epithelial cell subpopulations. Intriguingly, many of the identified factors align well with previous reports[1], thus further confirming their findings and suggesting that such signatures may be useful to study the role of novel factors in defining specific intestinal lineages.

Surprisingly, however, *Lgr5* expression did not correlate with inferred cell potency. *Lgr5*$^{\text{hi}}$ cells were largely found in the clusters associated with secretory progenitor cells as previously suggested for some intestinal Lgr5+ cells[18,59] and similar to the stomach, where high levels of *Lgr5* can be detected in differentiated chief[60] and antral basal secretory cells[61]. In addition, we find that *Lgr5* expression is not restricted to stem or progenitor cells but can also be detected in differentiated Tuft and other secretory cells[18]. In sharp contrast, some *Lgr5*$^{\text{low}}$ cells were characterized as high potency cells and lineage tracing studies in multiple mouse models narrowed ISC localization to the crypt isthmus, where *Lgr5*$^{\text{low}}$ cells are present. While the inducible Cre drivers (Lgr4, Ki67) may occasionally label cells at the crypt base, tracing ribbons, which persisted independent of crypt base labelling, could be detected starting at the level of the isthmus. Furthermore, flow-cytometry analyses revealed the highest tracing efficiency in Troy$^{\text{CreERT}}$ mice that best labeled the isthmus crypt compartment, reflecting a greater labelling of ISC. Importantly, these observations resonate well with our and others previous reports[3,5,7,12] and highlight similarities with the gastric mucosa where stem cells localize in the isthmus region[62].

Cells with the highest cell potency were detected at the boundary between the progenitor clusters (Sec-Progenitor and Abs-Progenitor). Notably, the difficulty in segregating ISC from early progenitors suggest their high-degree of similarity and may indicate they co-exist in closely comparable cellular states. Analysis of the factors that best associate with inferred cell potency revealed multiple chromatin remodelers, including Atad2, which has been recently shown to determine cell potency in the skin[35], as well as other factors previously suggested to regulate intestinal stemness[37,38]. Similarly, to Lgr5, other proposed markers for ISC, such as Ascl2[63], also failed to correlate with inferred cell potency and were rather broadly expressed within the secretory lineage.

In addition to characterizing homeostatic crypt cell composition at steady state, we also analyzed intestinal regeneration following Lgr5$^+$ cell ablation and IR damage. In line with previous observations[2,4], both models mostly eliminated *Lgr5* expressing cells, without evoking major composition changes within the intestine, at least for the DT ablation model. Notably, our results indicate that cells with high inferred cell potency persist within the intestine and align well with the proposed homeostatic ISC compartment. Furthermore, they highlight that differentiated cell types retain low cell potency states in both injury models and do not appear to acquire features of stemness. In fact, differentiated secretory cells, marked by high levels of *Dll1* expression, do not participate in intestinal regeneration, whereas surviving isthmal cells do. While the mouse models studied do not allow for the discrimination of the relative contribution between surviving ISC or early progenitors, they provide evidence that the potential to regenerate is restricted to these populations. In addition, when we analyzed the changes in the transcriptional profile of ISCs following IR damage, we detected a specific signature of regenerating cells[58]. This included upregulation of *Ly6a* (Sca-1) and *Areg* for example, and was accompanied by increased activation of several damage response factors, such as Baz1a and Apex1[56-58,64]. Intriguingly, we noted upregulation of *Clu*, an additional target of YAP signaling[65], indicating that previously proposed 'revival' stem cells[4] correspond to surviving ISC and early progenitors rather than a distinct quiescent stem cell population.

In summary, our unbiased analysis suggests a new regulatory model where a subpopulation of isthmal stem cells sustains the intestine under conditions of both homeostasis and regeneration.

Previous findings suggesting the presence of a reserve stem cell or extensive cellular plasticity likely reflected overlap of inducible Cre driver with this isthmal population. Our finding of broad expression by intestinal progenitors of multiple genes used for lineage tracing can help reconcile previous reports[8,9]. Lineage allocation to either an absorptive or secretory fate occurs early on, resulting in bidirectional migration of cells from the isthmal ISC. Future studies will need to focus on the regulation of these early events and the key niche signals which maintain stem cell renewal within the intestinal isthmus.

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

Acknowledgments: This research was funded in part through the NIH/NCI Cancer Center Support Grant P30CA013696 and used the resources of the Herbert Irving Comprehensive Cancer Center Flow Cytometry Shard Resources, Molecular Pathology/MPSR, Genomics and High Throughput Screening, as well as the Genetically Modified Mouse Model Shared Resource (GMMMSR). This work was supported by grants from the NIH/NCI including UO1DK103155, R35CA210088, RO1NK128195 to TCW; as well as a NCI Outstanding Investigator Award (R35 CA197745) and two NIH Shared Instrumentation Grants (S10 OD012351 and S1 0OD021764), all to AC. AV is supported by a U.S. Department of Defense Early Investigator Research Award (W81XWH19-1-0337) and an Early Career Development Pilot Award NIH/NCI Cancer Center, funded through the Cancer Center Support grant, P30CA013696. We thank Vladimir Korinek for the generous gift of the Troy^{CreERT} mice. We are grateful for the support we received by Columbia University shared resources and we would like to thank Sun Dajiang “Kevin” (Molecular

Pathology/MPSR), Kissner, Michael (Columbia Stem Cell Initiative Flow Cytometry core facility), Erin Bush (Genomics and High Throughput Screening), and Lin Chyuan-Sheng “Victor” (Genetically Modified Mouse Model), for their incredible expertise and help in this project.

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

References

1 Beumer, J. & Clevers, H. Cell fate specification and differentiation in the adult mammalian intestine. *Nat Rev Mol Cell Biol* **22**, 39-53, doi:10.1038/s41580-020-0278-0 (2021).

2 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).

3 Asfaha, S. *et al.* Krt19(+)/Lgr5(-) Cells Are Radioresistant Cancer-Initiating Stem Cells in the Colon and Intestine. *Cell Stem Cell* **16**, 627-638, doi:10.1016/j.stem.2015.04.013 (2015).

4 Ayyaz, A. *et al.* Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* **569**, 121-125, doi:10.1038/s41586-019-1154-y (2019).

5 Powell, A. E. *et al.* The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell* **149**, 146-158, doi:10.1016/j.cell.2012.02.042 (2012).

6 Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc Natl Acad Sci U S A* **108**, 179-184, doi:10.1073/pnas.1013004108 (2011).

7 Sheng, X. *et al.* Cycling Stem Cells Are Radioresistant and Regenerate the Intestine. *Cell Rep* **32**, 107952, doi:10.1016/j.celrep.2020.107952 (2020).

8 van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat Cell Biol* **14**, 1099-1104, doi:10.1038/ncb2581 (2012).

9 Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* **18**, 203-213, doi:10.1016/j.stem.2016.01.001 (2016).

10 Murata, K. *et al.* Ascl2-Dependent Cell Dedifferentiation Drives Regeneration of Ablated Intestinal Stem Cells. *Cell Stem Cell* **26**, 377-390 e376, doi:10.1016/j.stem.2019.12.011 (2020).

11 Shivdasani, R. A., Clevers, H. & de Sauvage, F. J. Tissue regeneration: Reserve or reverse? *Science* **371**, 784-786, doi:10.1126/science.abb6848 (2021).

12 Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed in vivo in intestinal stem cells. *Nat Genet* **40**, 915-920, doi:10.1038/ng.165 (2008).

13 Takeda, N. *et al.* Interconversion between intestinal stem cell populations in distinct niches. *Science* **334**, 1420-1424, doi:10.1126/science.1213214 (2011).

14 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-1007, doi:10.1038/nature06196 (2007).

15 Obradovic, A. *et al.* Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell* **184**, 2988-3005 e2916, doi:10.1016/j.cell.2021.04.038 (2021).

16 Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov* **9**, 1102-1123, doi:10.1158/2159-8290.CD-19-0094 (2019).

17 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339, doi:10.1038/nature24489 (2017).

18 Bottcher, A. *et al.* Non-canonical Wnt/PCP signalling regulates intestinal stem cell lineage priming towards enteroendocrine and Paneth cell fates. *Nat Cell Biol* **23**, 23-31, doi:10.1038/s41556-020-00617-2 (2021).

19 Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167 e1115, doi:10.1016/j.cell.2018.08.063 (2018).

20 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).

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 Bjerknes, M. & Cheng, H. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* **116**, 7-14, doi:10.1016/s0016-5085(99)70222-2 (1999).

33 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).

34 Ding, Y. *et al.* Smarca5-mediated epigenetic programming facilitates fetal HSPC development in vertebrates. *Blood* **137**, 190-202, doi:10.1182/blood.2020005219 (2021).

35 Baggiolini, A. *et al.* Developmental chromatin programs determine oncogenic competence in melanoma. *Science* **373**, eabc1048, doi:10.1126/science.abc1048 (2021).

36 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).

37 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).

38 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).

39 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).

40 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).

41 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).

42 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).

43 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).

44 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).

45 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).

46 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).

47 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).

48 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).

49 Basak, O. *et al.* Troy+ brain stem cells cycle through quiescence and regulate their number by sensing niche occupancy. *Proc Natl Acad Sci U S A* **115**, E610-E619, doi:10.1073/pnas.1715911114 (2018).

50 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).

51 Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822-825, doi:10.1126/science.1196236 (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).

62 Hayakawa, Y., Nakagawa, H., Rustgi, A. K., Que, J. & Wang, T. C. Stem cells and origins of cancer in the upper gastrointestinal tract. *Cell Stem Cell* **28**, 1343-1361, doi:10.1016/j.stem.2021.05.012 (2021).

63 Schuijers, J. *et al.* Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell* **16**, 158-170, doi:10.1016/j.stem.2014.12.006 (2015).

64 Shao, J. & Sheng, H. Amphiregulin promotes intestinal epithelial regeneration: roles of intestinal subepithelial myofibroblasts. *Endocrinology* **151**, 3728-3737, doi:10.1210/en.2010-0319 (2010).

65 Gregorieff, A., Liu, Y., Inanlou, M. R., Khomchuk, Y. & Wrana, J. L. Yap-dependent reprogramming of Lgr5(+) stem cells drives intestinal regeneration and cancer. *Nature* **526**, 715-718, doi:10.1038/nature15382 (2015).

Isthmal stem cells sustain intestinal homeostasis and regeneration

Malagola E., Vasciaveo A., Ochiai Y., Kim W., Middelhoff M., Nienhüser H., Belin B., LaBella J.,

Zamechek LB., Wong M.H., Li L., Guha C., Yan K., Califano A., and Wang T.C.

Figure panels

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

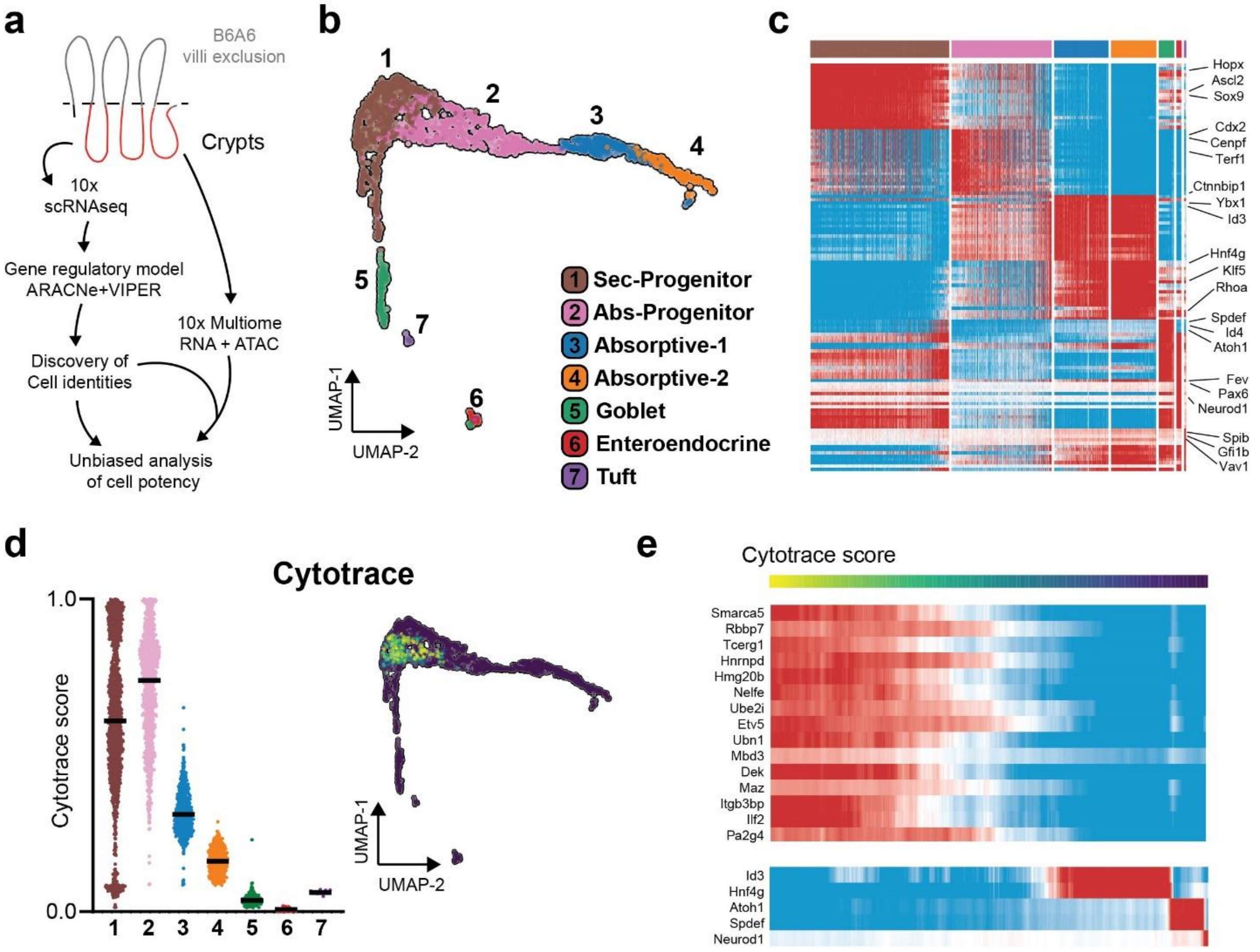
Figure 2: *Lgr5* expression is not restricted to intestinal stem and progenitor cells

Figure 3: Isthmal stem cells sustain intestinal tissue homeostasis

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

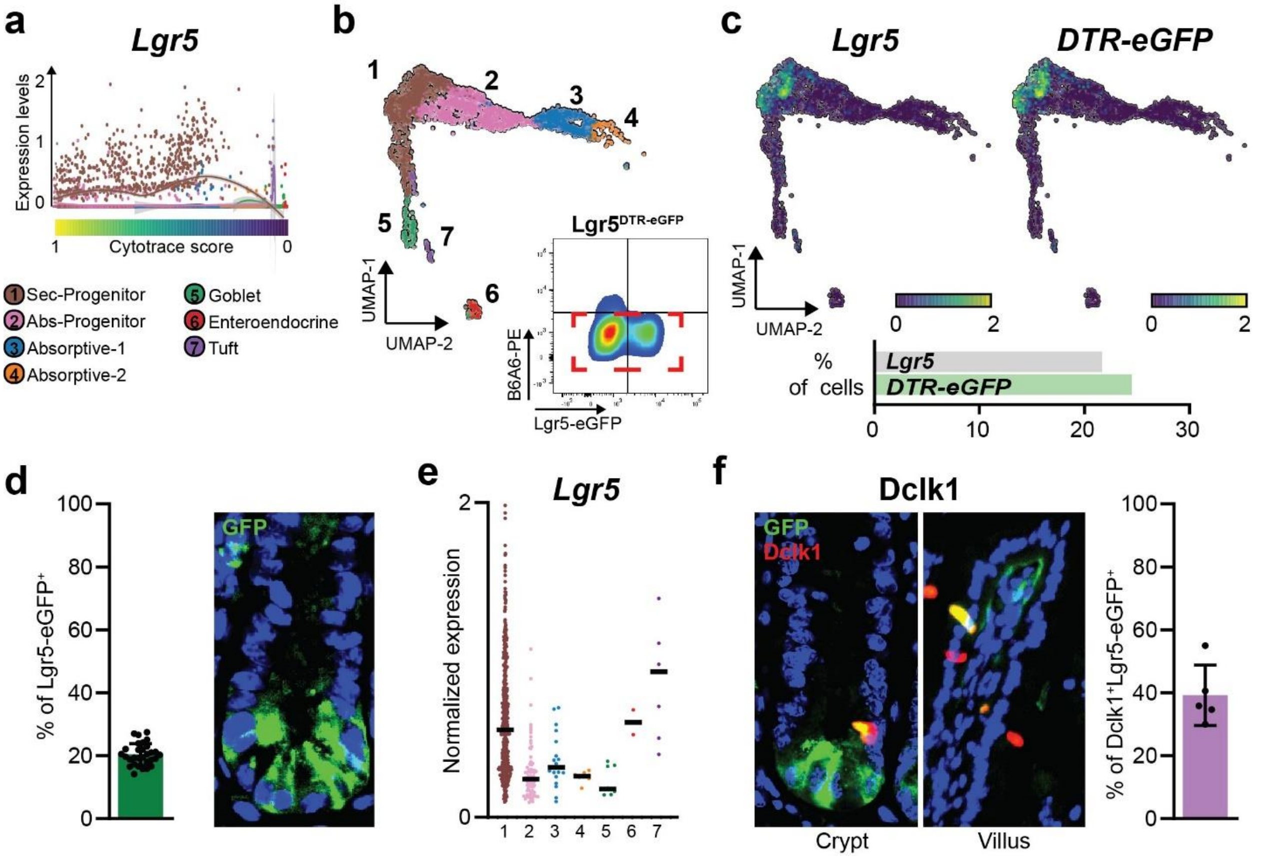
Figure 5: Surviving isthmal stem cells regenerate the intestinal epithelium following IR damage

Figure 1



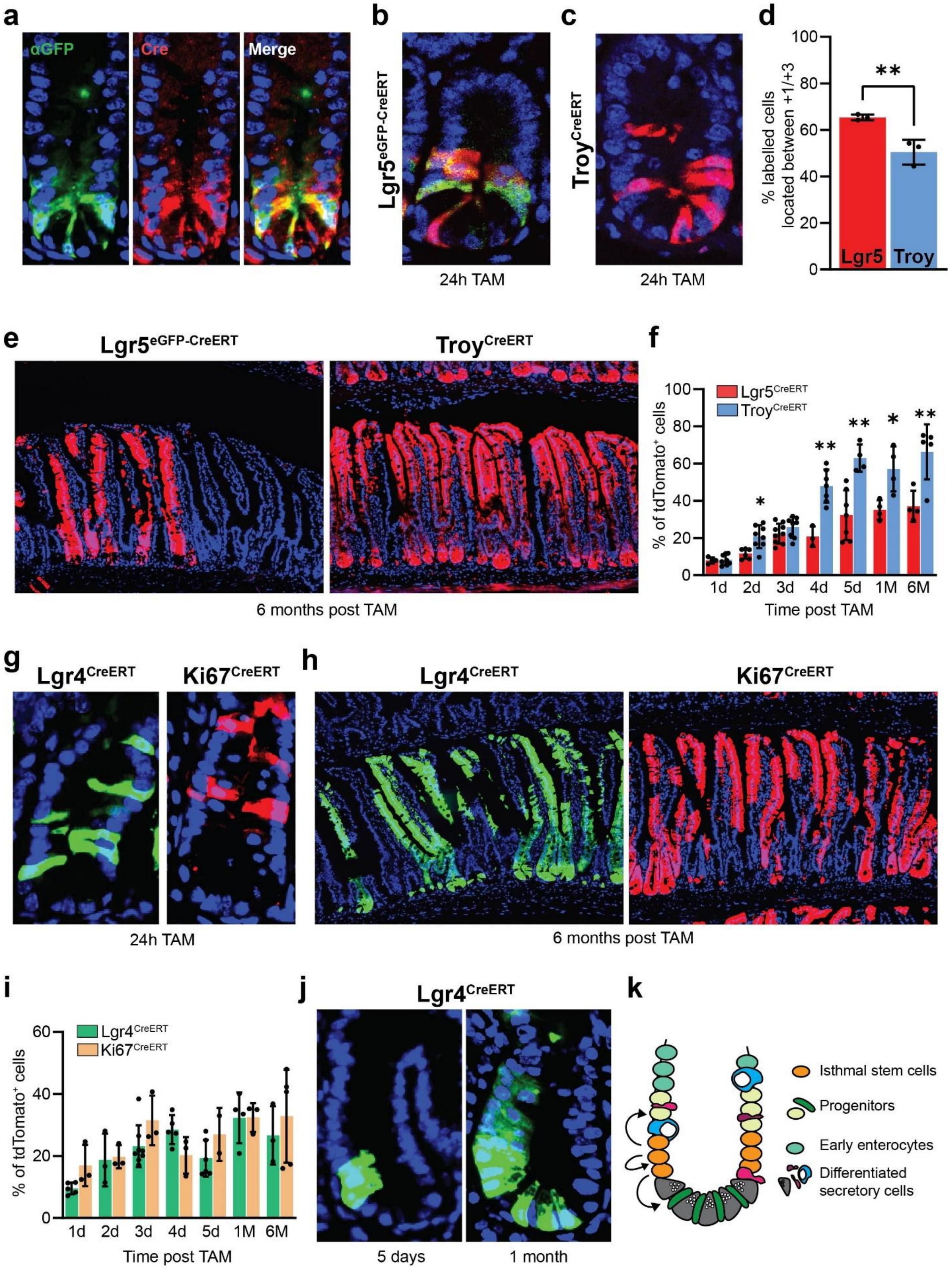
An unbiased approach to elucidate the organization of intestinal crypt epithelial cells: a. Schematic representation of the experimental workflow and computational pipeline for the analysis of crypt epithelial cells. (*See methods*) **b.** UMAP plot showing protein activity -based clustering solution based on gene regulatory network analysis. **c.** Heatmap showing top differentially activated regulatory proteins. **d.** Violin-plot and UMAP plot (top 10% cells are colored) showing CytoTRACE scores for individual cells; black line indicate median value per cluster. **e.** Pseudotime analysis using CytoTRACE scores to order cells based on inferred cell potency (high to low 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.

Figure 2



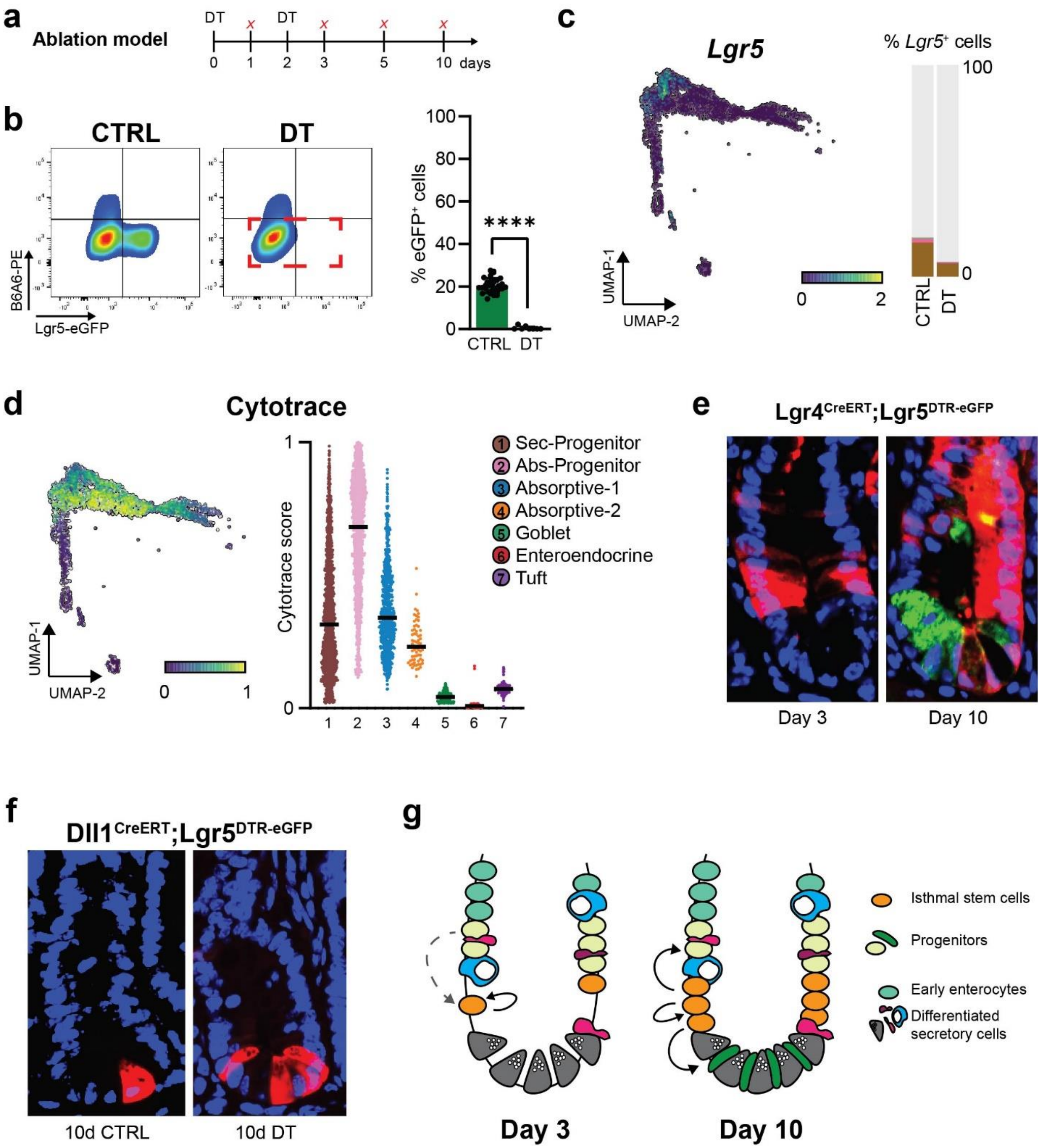
***Lgr5* expression is not restricted to intestinal stem and progenitor cells: a.** Pseudotime analysis using CytoTRACE scores to order cells based on inferred cell potency (high to low potency from left to right), showing expression levels of *Lgr5*. **b.** Sorting strategy and UMAP plot of Lgr5$^{DTR\text{-}eGFP}$ 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.** Flow cytometry quantification of Lgr5-eGFP^{+} cells in purified crypts (n=30); together with representative image of immunofluorescence staining for Lgr5-eGFP in mouse jejunum. **e.** Violin-plot showing expression levels of *Lgr5* in the identified clusters (only Lgr5^{+} cells are shown). **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
αGFP
Cre
Merge
b
Lgr5$^{eGFP-CreERT}$
24h TAM
c
Troy^{CreERT}
24h TAM
d
% labelled cells located between +1/+3
100
80
60
40
20
0
**
Lgr5
Troy
e
Lgr5$^{eGFP-CreERT}$
Troy^{CreERT}
6 months post TAM
f
Lgr5^{CreERT}
Troy^{CreERT}
% of tdTomato$^+$ cells
100
80
60
40
20
0
*
**
**
*
**
1d
2d
3d
4d
5d
1M
6M
Time post TAM
g
Lgr4^{CreERT}
Ki67^{CreERT}
24h TAM
h
Lgr4^{CreERT}
Ki67^{CreERT}
6 months post TAM
i
Lgr4^{CreERT}
Ki67^{CreERT}
% of tdTomato$^+$ cells
60
40
20
0
1d
2d
3d
4d
5d
1M
6M
Time post TAM
j
Lgr4^{CreERT}
5 days
1 month
k
Isthmal stem cells
Progenitors
Early enterocytes
Differentiated secretory cells

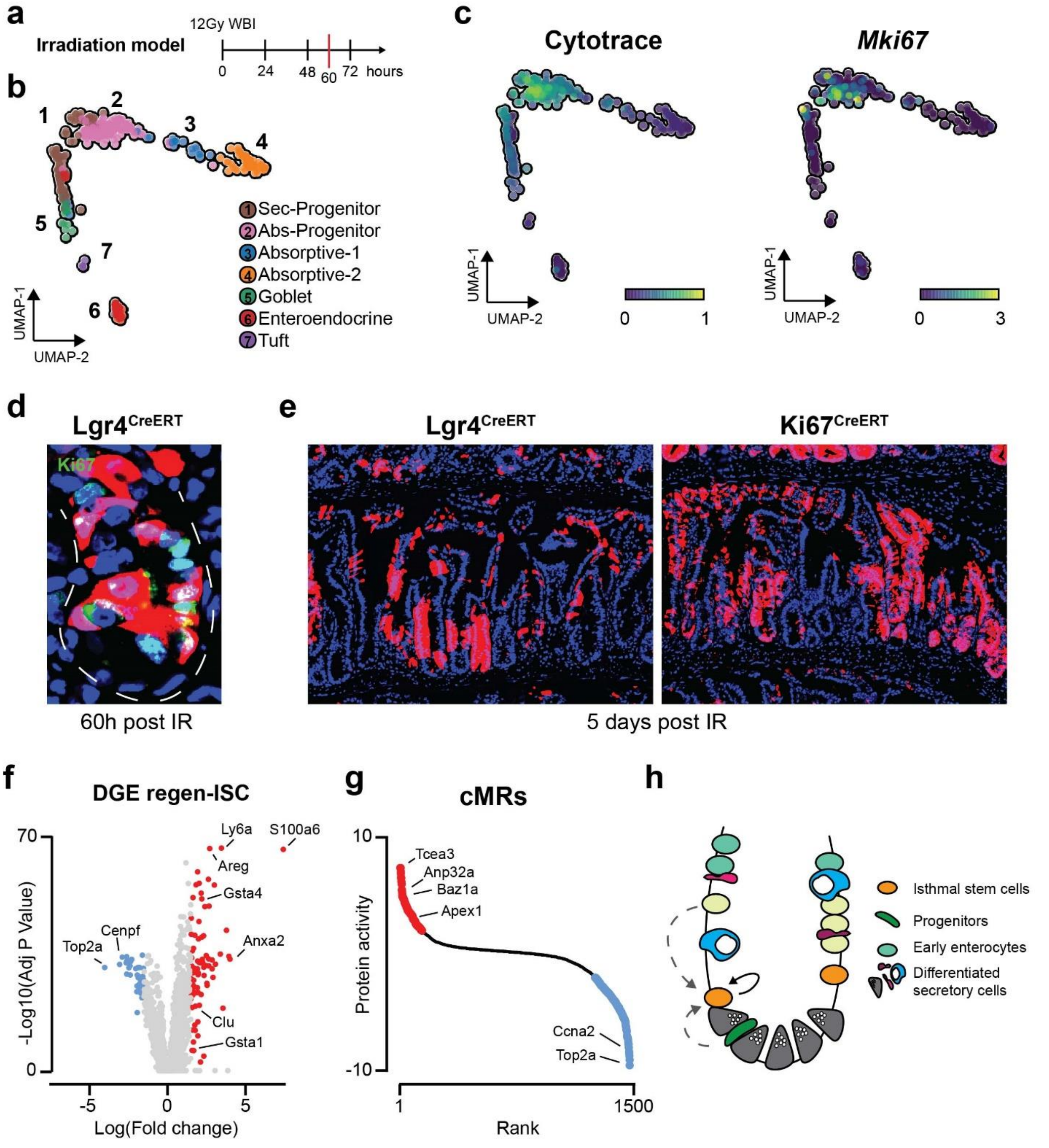
Isthmal stem cells sustain intestinal tissue homeostasis: a. Representative images of immunofluorescence staining for eGFP and Cre in Lgr5^{eGFPCreERT} mouse jejunum. **b.** Representative image of Lgr5$^{eGFP\text{-}CreERT}$ mice 24 hours post TAM induction. **c.** Representative image of Troy^{CreERT} mice 24 hours post TAM induction. **d.** % of tdTomato^{+} cells from Lgr5$^{eGFP\text{-}CreERT}$ and Troy^{CreERT} mice located within the +1/+3 position in intestinal crypts at day 1. (n=3) **e.** Representative images of Lgr5$^{eGFP\text{-}CreERT}$ and Troy^{CreERT} mice 6 months post TAM induction. **f.** Flow cytometry based quantification of tdTomato^{+} cells in crypt epithelial cells at indicated time points (n≥3). **g.** Representative images of Lgr4^{CreERT} and Ki67^{CreERT} mice 24h post TAM induction. **h.** Representative images of Lgr4^{CreERT} and Ki67^{CreERT} mice 6 months post TAM induction. **i.** Flow cytometry based quantification of tdTomato^{+} cells at indicated time points (n≥3). **j.** Representative images showing evidences of tracing ribbons at 5 days or 1 month post TAM without CBC labelling. **k.** Schematic representation of the model proposed. 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
x
DT
x
x
x
0
1
2
3
5
10
days
b
CTRL
DT
B6A6-PE
Lgr5-eGFP
% eGFP⁺ cells
100
80
60
40
20
0
****
CTRL
DT
c
Lgr5
% Lgr5⁺ cells
100
0
UMAP-1
UMAP-2
0
2
CTRL
DT
d
Cytotrace
UMAP-1
UMAP-2
0
1
Cytotrace score
1
0
1
2
3
4
5
6
7
① Sec-Progenitor
② Abs-Progenitor
③ Absorptive-1
④ Absorptive-2
⑤ Goblet
⑥ Enteroendocrine
⑦ Tuft
e
Lgr4CreERT;Lgr5DTR-eGFP
Day 3
Day 10
f
Dll1CreERT;Lgr5DTR-eGFP
10d CTRL
10d DT
g
Day 3
Day 10
Isthmal stem cells
Progenitors
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.** Sorting strategy (left) analysis of crypt epithelial cells in CTRL or DT ablated mice (sorted gate in red), (right) flow cytometry quantification of Lgr5$^{DTReGFP+}$ cells in control and DT treated crypts (n≥10). **c.** 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. **d.** UMAP and violin plots showing CytoTRACE score in DT treated epithelial cells. **e.** Representative images for Lgr4^{CreERT} lineage tracing upon concurrent DT treatment. **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. Gray dotted line indicate the possibility for early progenitors to serve as ISC. 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
b
1
2
3
4
5
6
7
①Sec-Progenitor
②Abs-Progenitor
③Absorptive-1
④Absorptive-2
⑤Goblet
⑥Enteroendocrine
⑦Tuft
UMAP-1
UMAP-2
c
Cytotrace
Mki67
UMAP-1
UMAP-2
0
1
UMAP-1
UMAP-2
0
3
d
Lgr4^CreERT
Ki67
60h post IR
e
Lgr4^CreERT
Ki67^CreERT
5 days post IR
f
DGE regen-ISC
70
0
-Log10(Adj P Value)
Ly6a
S100a6
Areg
Gsta4
Anxa2
Cenpf
Top2a
Clu
Gsta1
-5
0
5
Log(Fold change)
g
cMRs
10
-10
Protein activity
Tcea3
Anp32a
Baz1a
Apex1
Ccna2
Top2a
1
1500
Rank
h
Isthmal stem cells
Progenitors
Early enterocytes
Differentiated secretory cells

Surviving isthmal stem cells regenerate the intestinal epithelium following IR damage: a. Schematic representation of irradiation damage model. **b.** UMAP plot showing protein activity based clustering solution for irradiated crypt epithelial cells. **c.** UMAP plots showing both inferred cell potency through CytoTRACE and *Mki67* expression in irradiated crypt epithelial cells. **d.** Representative image for Ki67 staining 60h post IR. **e.** Representative images of Lgr4^{CreERT} and Ki67^{CreERT} linage tracing 5 days post IR. **f.** Volcano plot of differentially expressed genes in regenerating (positive LogFC) vs homeostatic ISC (negative LogFC). **g.** 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. **h.** Schematic representation of the model proposed of surviving intestinal cells after IR exposure. Dotted gray lines indicate the putative involvement of early progenitors to serve as ISC.

Isthmal stem cells sustain intestinal homeostasis and regeneration

Malagola E., Vasciaveo A., Ochiai Y., Kim W., Middelhoff M., Nienhüser H., Belin B., LaBella J.,

Zamechek LB., Wong M.H., Li L., Guha C., Yan K., Califano A., and Wang T.C.

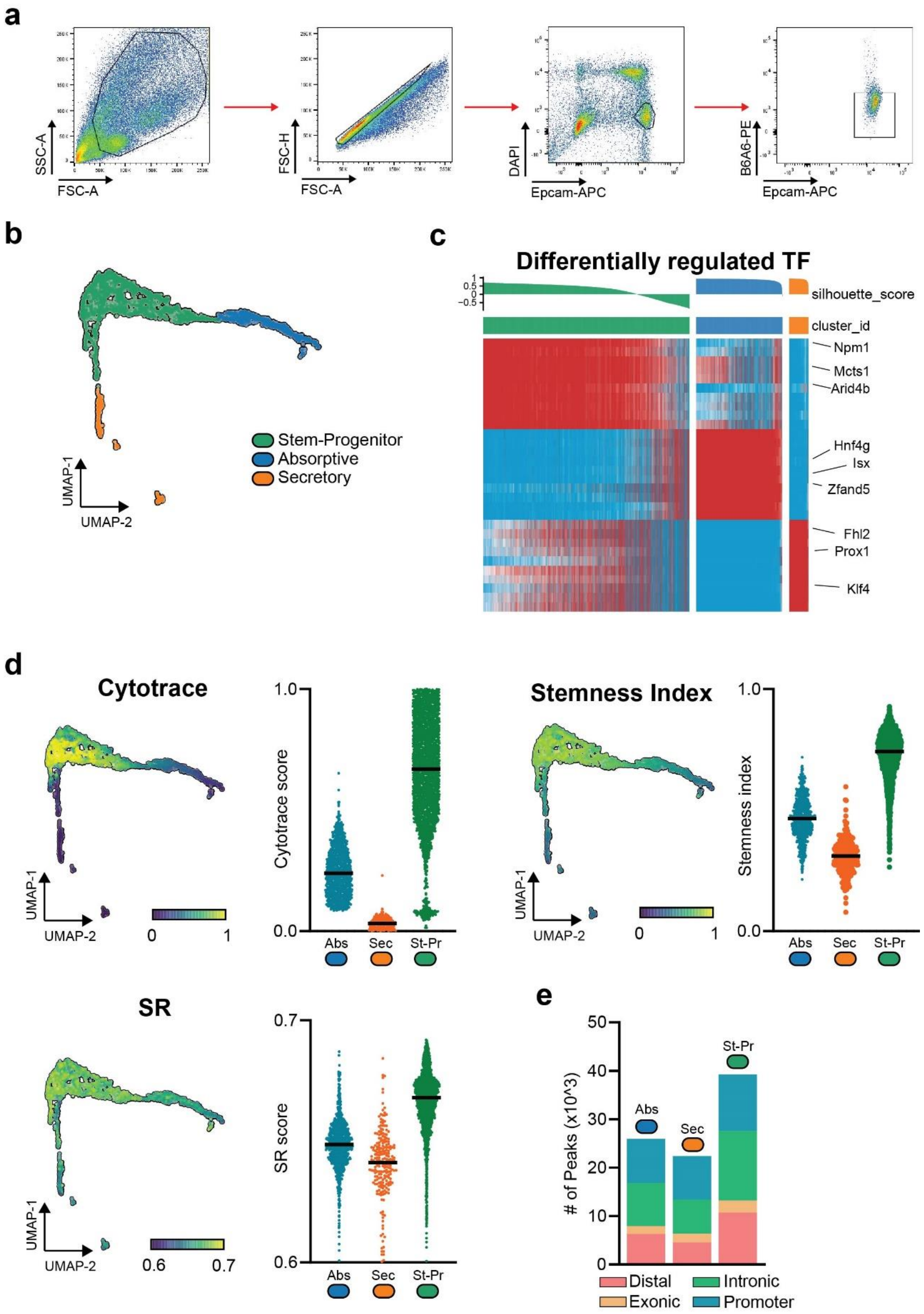
Extended Data and Methods

Extended Data Figures 1-11

List of Extended Data Table

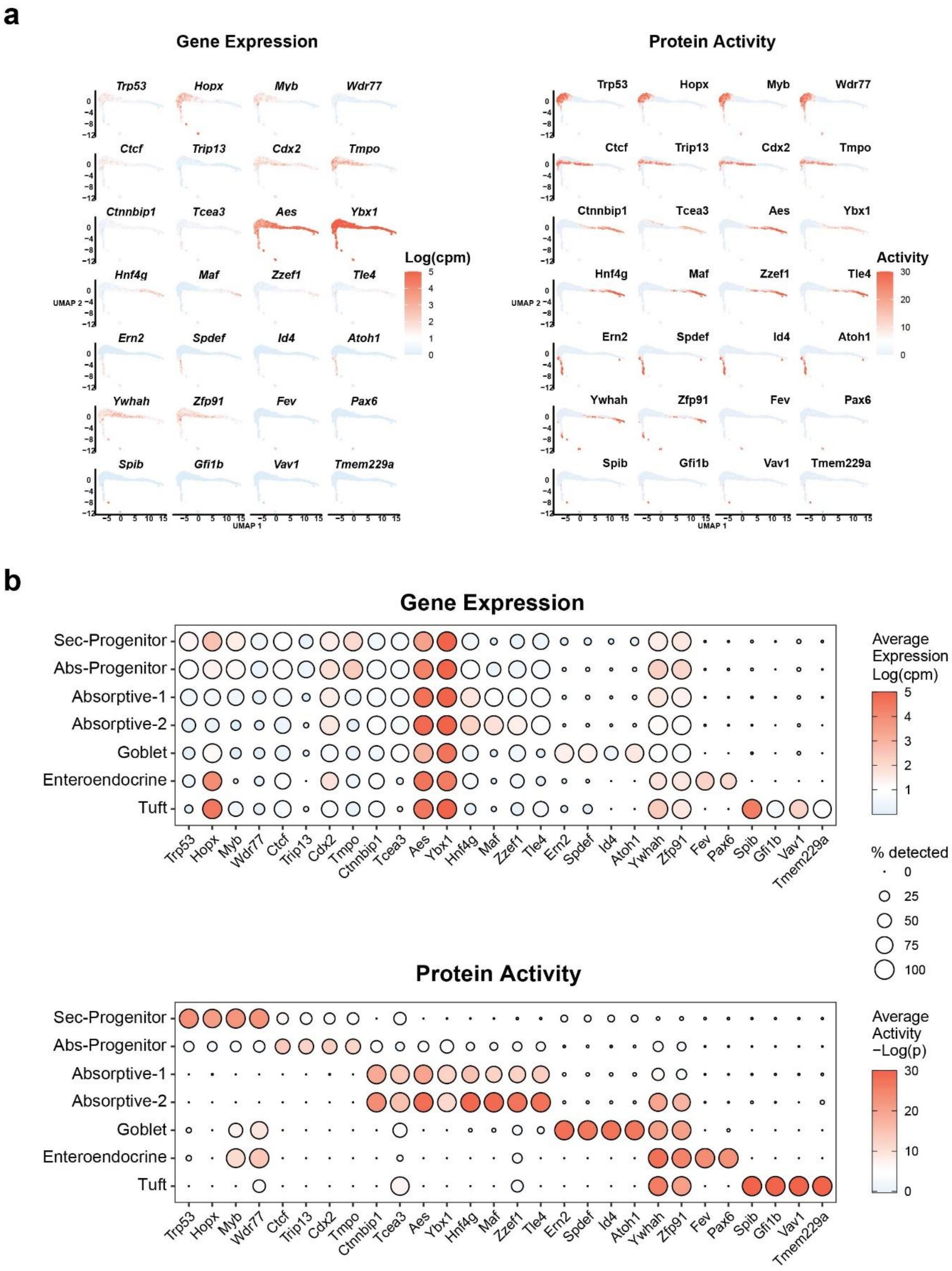
Methods description

Extended Data Figure1



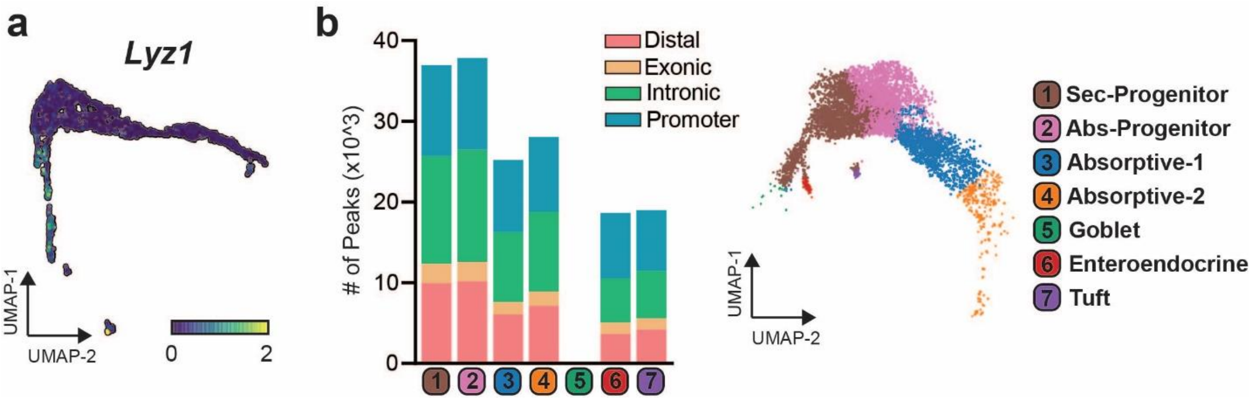
Extended Data Figure1: a. Gating strategy used for flow-cytometry based sorting of crypt epithelial cells. **b.** UMAP plot showing protein activity-based clustering solution. **c.** Heatmap showing top differentially activated regulatory proteins. On top, computed silhouette scores for individual cells. **d.** UMAP and Violin plots showing CytoTRACE, Stemness index, and Signaling Entropy (SR) scores computed for individual cells. **e.** Bar plot showing number of identified accessible peaks per cluster based on scATACseq.

Extended Data Figure2



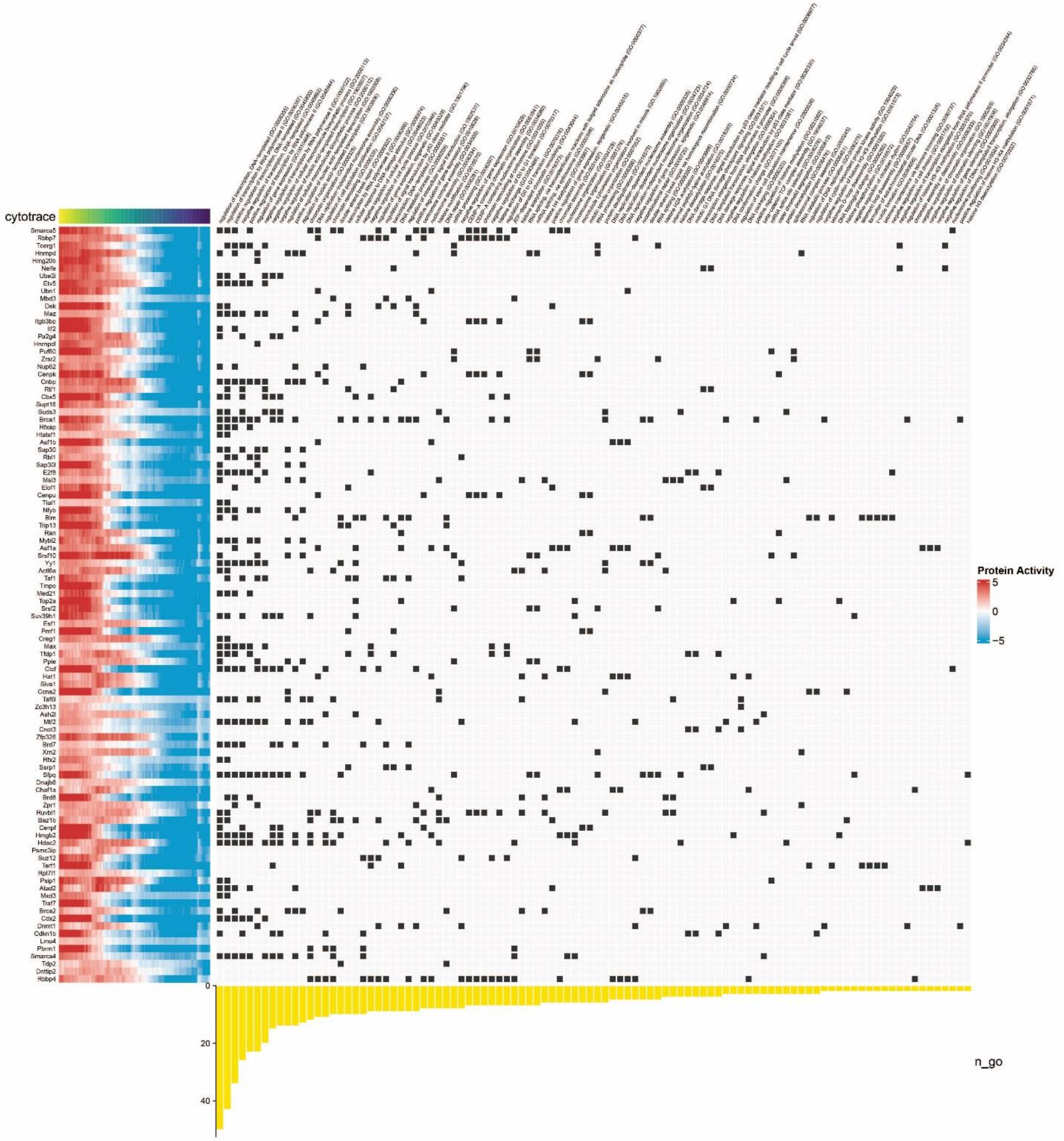
Extended Data Figure2: a. Showing UMAP plots of top 4 markers for each cluster (rows) as identified from protein activity-based clustering. Panel on the left shows gene expression as log(cpm), while panel on the right shows for the same markers the corresponding protein activity as -log10(p). **b.** Same cluster-specific markers are shown as dot plots. Protein activity shows higher sensitivity and specificity than gene expression.

Extended Data Figure 3



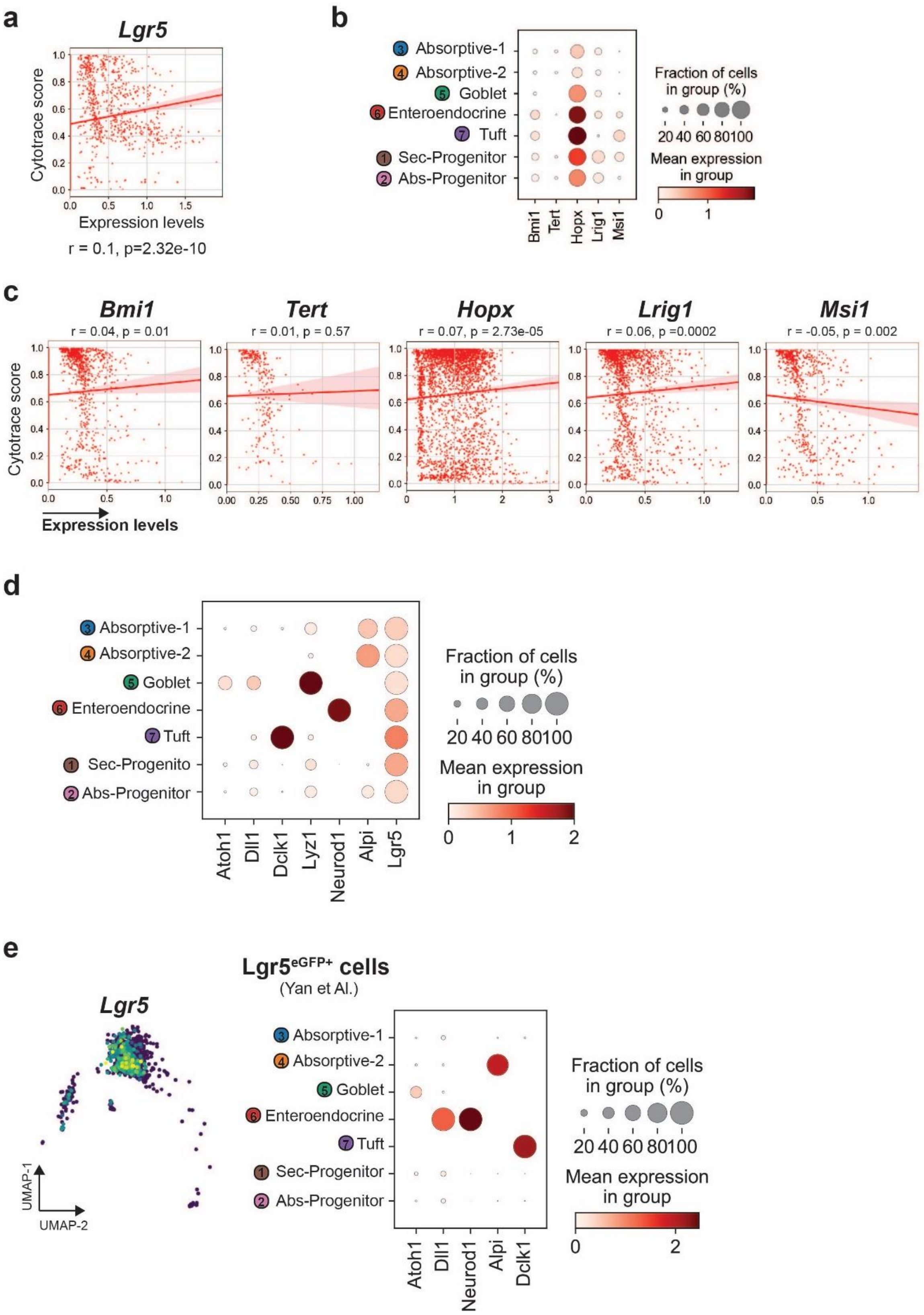
Extended Data Figure3: a. UMAP plot showing expression domain for Lyz1. **b.** Bar plot showing number of accessible peaks per cluster together with UMAP plot showing clustering results for multiome dataset; note cluster 5 consist of very few cells and was therefore excluded from further analyses.

Extended Data Figure 4:



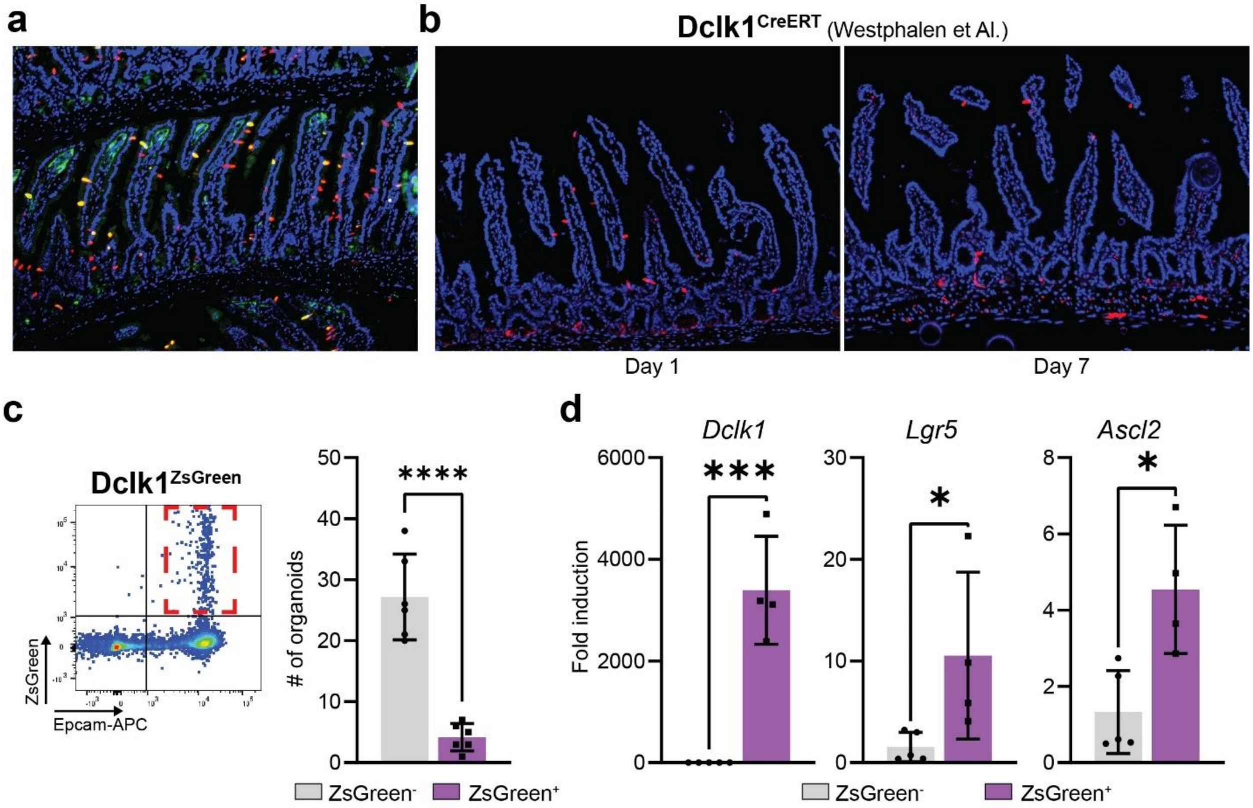
Extended Data Figure 4: On the left, strip plots showing inferred activity of the 100 CytoTRACE score-based top-correlating transcriptional regulators. On the right, top GO-associated terms for each regulatory protein, using 'Molecular Function" as query.

Extended Data Figure 5:



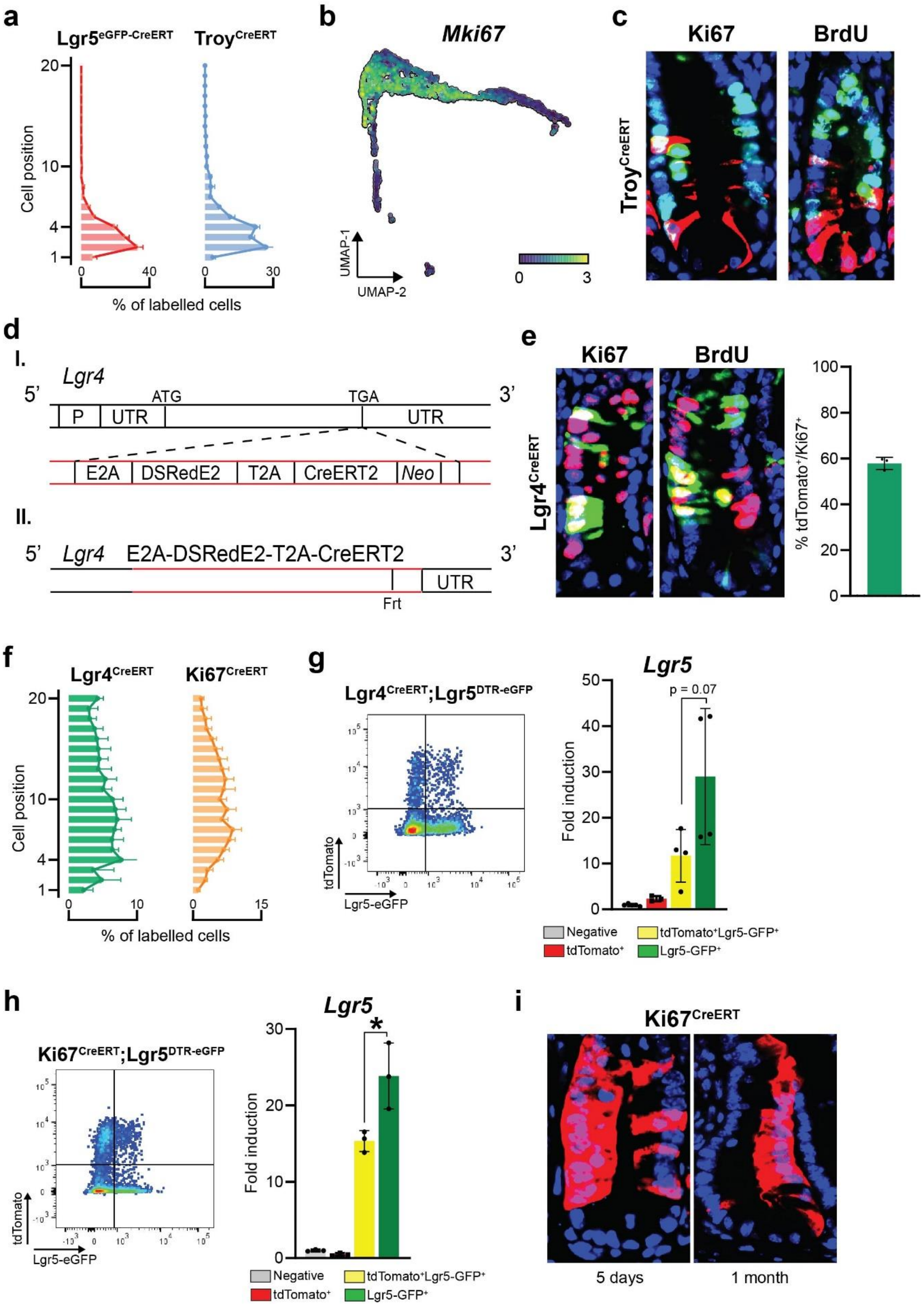
Extended Data Figure 5: a. Correlation plot of Lgr5 and CytoTRACE score. **b.** Dot-plot showing broad pattern of expression for multiple proposed markers of ISC. **c.** Correlation plots between CytoTRACE and expression levels of proposed ISC markers. **d.** Dot-plot showing expression of known markers of differentiation in Lgr5 expressing cells. **e.** Left: UMAP distribution plot showing expression of *Lgr5* in Lgr5^{eGFP+} sorted cells; right: Dotplot showing expression of markers of differentiated cells.

Extended Data Figure 6



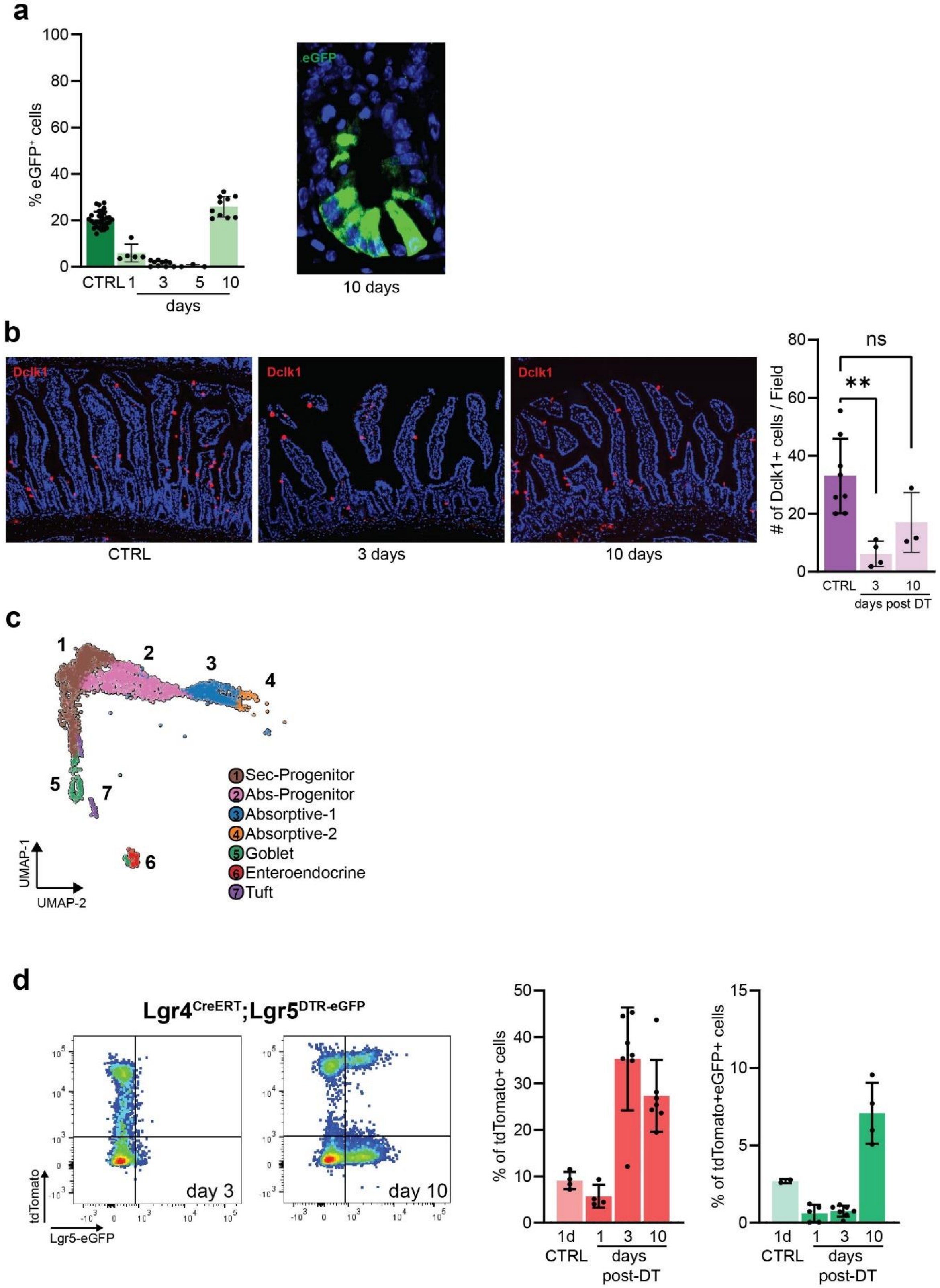
Extended data Figure 6: a. Representative image of low power magnification view of Dclk1 (Red) and eGFP (green) in intestinal jejunum of a Lgr5^{DTReGFP} mouse. **b.** Lineage tracing analysis of Dclk1^{CreERT} mouse model. At day 1 scattered tdTomato^{+} Tuft cells can be observed, at day 7 post TAM most tdTomato^{+} cells are washed out. **c.** On the left, sorting strategy for the isolation of Dclk1$^{ZsGreen+}$ cells (Red gate); on the right barplot showing quantification of the number of organoids *in vitro* (n=2). **d.** 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 7:



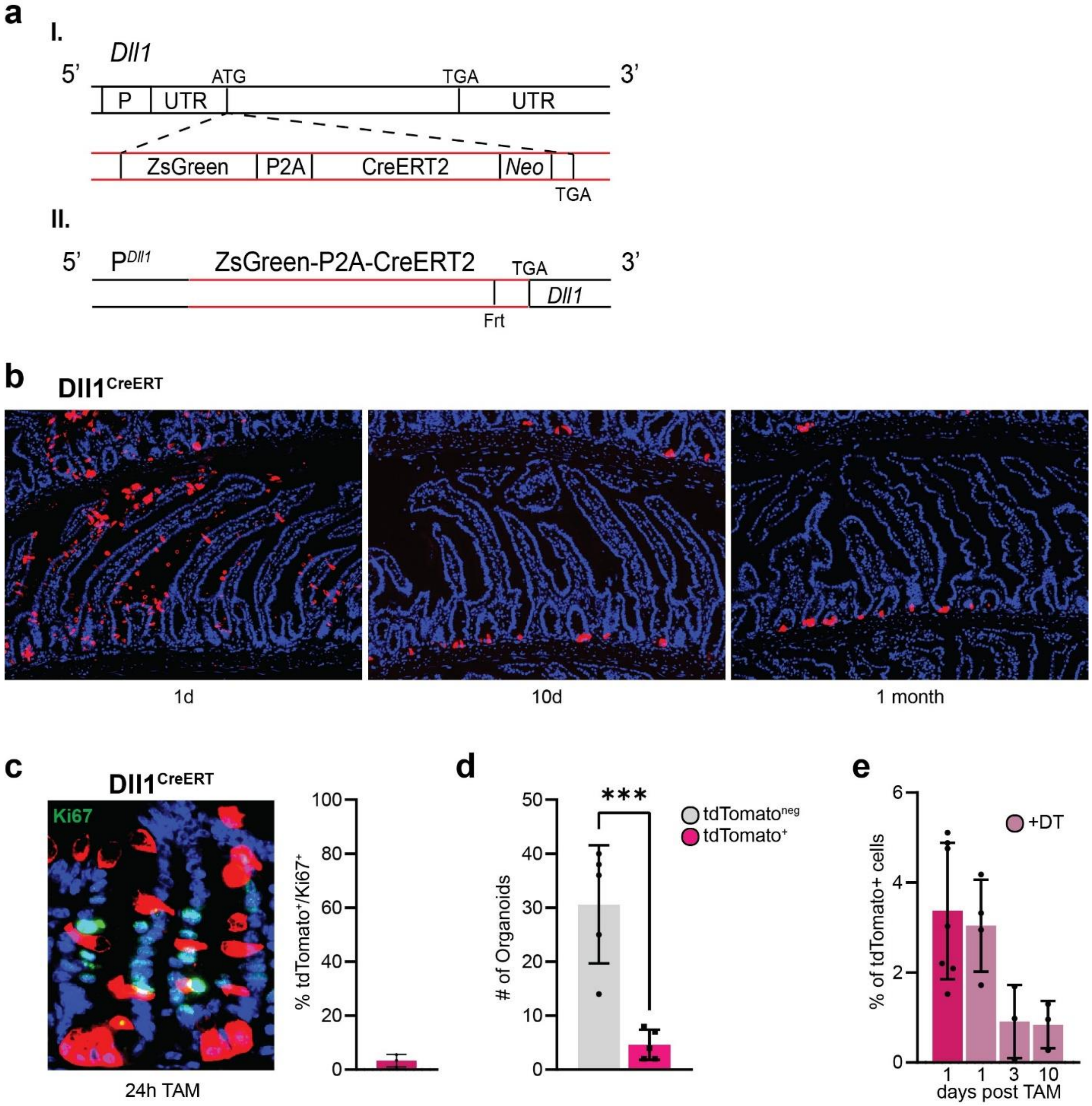
Extended Data Figure 7: a. Position based counting of tdTomato^{+} cells 24h post TAM administration in Lgr5^{CreERT} and Troy^{CreERT} mice. (n=3) **b.** UMAP plot showing expression of Mki67 in crypt epithelial cells. **c.** Representative images of IF staining for Ki67 and BrdU (in green) in Troy^{CreERT} mice 24 hours post TAM induction. **d.** Designing strategy for the Lgr4^{CreERT} mouse model. (See methods). **e.** Representative images of IF staining for Ki67 and BrdU (in red) in Lgr4^{CreERT} mice 24 hours post TAM induction (R26-ZsGreen); on the right quantification of labelled cells also positive for Ki67 (n=3) **f.** Position based counting of labelled cells 24h post TAM administration in Lgr4^{CreERT} and Ki67^{CreERT} mice. **g-h.** Flow cytometry analysis of overlap between Lgr4 and Ki67 labelled cells (24h) and Lgr5^{eGFP+} cells; together with bar plot showing expression levels of *Lgr5* in single sorted cells, groups labelled in the panel. (Lgr4: n=4, Ki67: n=3). **i.** Representative images of lineage traced glands (Ki67^{CreERT}) at the indicated time points post TAM induction. Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Extended Data Figure 8:



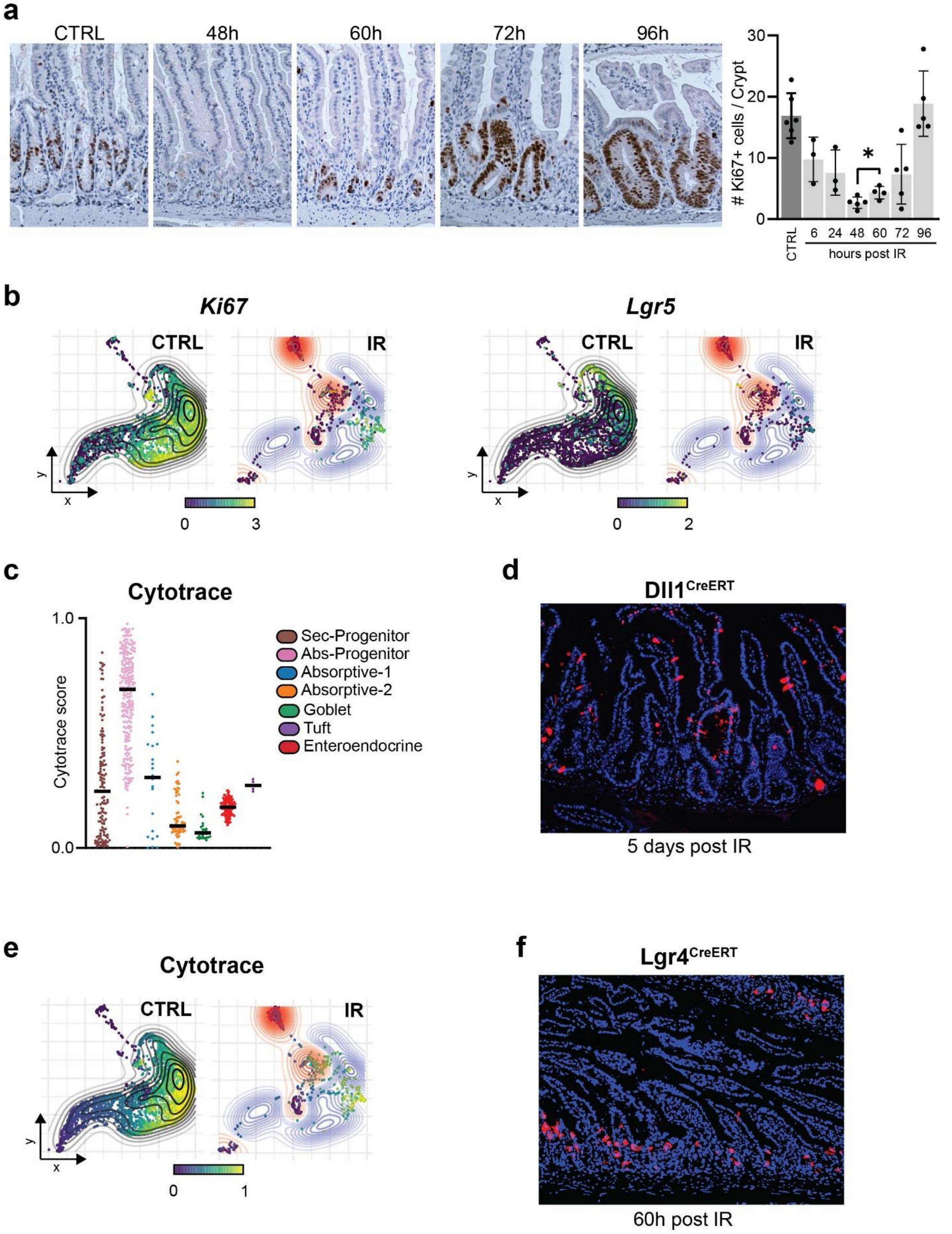
Extended Data Figure 8: a. Flow cytometry quantification of Lgr5^{eGFP+} cells in purified crypts (n≥3). On the right, representative images of Lgr5^{eGFP} 10 days after the first dose of DT. **b.** Representative images for Dclk1 staining of intestine following DT treatment together with bar plot showing 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.** Flow cytometry analysis and quantification of tdTomato+ cells (Lgr4^{CreERT};Lgr5^{DTReGFP}) at indicated time points, on the right barplot showing percentage of tdTomato+GFP+ double positive cells at indicated timepoints. Statistical method: Unpaired t-test, two-tailed; *: p<0.05, **: p<0.01, ***: p<0.01, ****: p<0.01.

Extended Data Figure 9



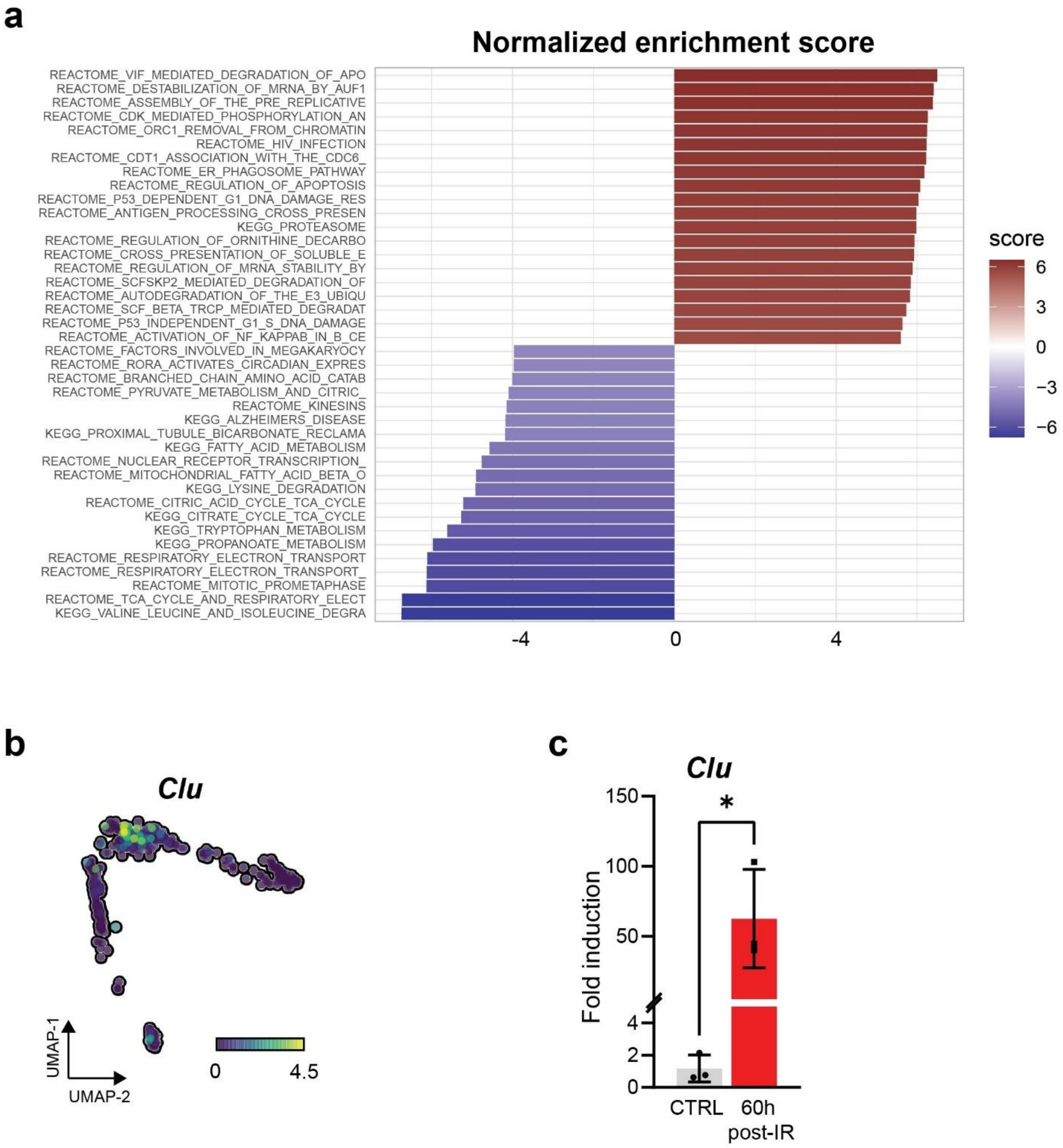
Extended Data Figure9: a. Designing strategy for the Dll1^{CreERT} mouse model. **b.** Representative images showing lineage tracing of Dll1^{CreERT} mice at indicated time points (n≥3). **c.** Representative image for Ki67 staining (in green) and bar plot showing % of Ki67^{+} /(Dll1^{CreERT}) tdTomato^{+} cells. (n=3) **d.** Bar plot showing quantification of the number of organoids generated from single sorted tdTomato^{-} and tdTomato^{+} (Dll1^{CreERT} , 24h post-TAM) cells. **e.** Bar plot showing flow cytometry based quantification of (Dll1^{CreERT}) tdTomato^{+} cells at indicated time points. (n≥3). Statistical method: Unpaired t-test, two-tailed; *: p<0.05, **: p<0.01, ***: p<0.01, ****: p<0.01.

Extended Data Figure 10



Extended Data Figure 10: 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. ($n \geq 3$) **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 (Right). **c.** Bar plot showing computed CytoTRACE score in irradiated crypt epithelial cells relative to control. **d.** Representative images showing tracing of tdTomato+ cells ($Dll1^{CreERT}$) five days post IR damage. **e.** KDE analysis showing computed CytoTRACE scores. **f.** Representative image of $Lgr4^{CreERT}$ mice 60 hours after IR damage. Statistical method: Unpaired t-test, two-tailed; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

Extended Data Figure 11



Extended Data Figure 11: a. Bar-plot showing gene expression-based pathway enrichment analysis**. b.** UMAP plot showing *Clu* expression in irradiated crypt epithelial cells. **c.** 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; *: $p<0.05$, **: $p<0.01$, ***: $p<0.01$, ****: $p<0.01$.

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 7-cluster Solution

Table 6: Protein activity 7-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) or R26-ZsGreen (Strain #:007906) reporter allele purchased from Jax (The Jackson Laboratory). TAM was dissolved in corn oil and administered 200uL per oral gavage at indicated time points. BrdU (Biolegend, 423401) was injected i.p. at 50mg/Kg 6 hours before harvesting. Diphtheria toxin (DT - EMD Millipore, 322326) was injected i.p. at 50ug/Kg as previously described[1]. 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 [2] 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[3] 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.

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)[4]. 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[4]. 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[5].

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[6], the most recent implementation of the ARACNe algorithm[7], 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[8,9]. The resulting network includes 1,797 regulators, 14,935 targets and 548,442 interactions. The 3,656 scRNA-Seq profiles were transformed to protein activity profiles using the metacell-derived regulatory network and the VIPER algorithm [10]. To avoid bias due to different regulon sizes, regulons were pruned to include only the 50 highest likelihood targets, as recommended in[10], 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. A gene expression signature (GES) was computed across all 3,656 cells by scaling the data after having normalized and variance-stabilized the UMI counts matrix using *sctransform* as implemented in the Seurat R package. Cells were computationally isolated based on their inferred lineage at the previous step. VIPER analysis was performed using cluster-specific regulatory networks and the 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. The absorptive lineage was divided in two clusters, the secretory lineage in three clusters and the stem/progenitor one in two clusters. The cells from the whole datasets were then labels based on these seven clusters. Next, the three lineage-specific VIPER analyses were merged together for all RPs that were represented across all the lineages. For the RPs that were lineage specific, we used the VIPER score from the lineage-specific regulatory network analysis. Protein activity of RPs that were not present in one of the networks were computed using metaVIPER across all the networks [11].

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[12], 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[13]. 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[14]. To improve the readout of the activation status for each protein in PPIs, we utilized VIPER-inferred protein activity scores rather than gene expression. PPIs were modeled using PrePPI[15], 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.

Comparison of pre and post-radiation cells: 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) [4]. 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 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).

scATAC-Seq analysis: scATAC-Seq data was processed using ArchR[16]. Peak calling was performed using MACS2 with default parameters[17]. Cluster analysis was performed using the RNA data modality by projecting the clustering solution identified on the CTRL sample using *ingest* over scanpy[18]. *De novo* 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 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).

2 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).

3 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).

4 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).

5 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).

6 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).

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

8 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).

9 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).

10 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).

11 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).

12 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).

13 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).

14 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).

15 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).

16 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).

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

18 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).