

Supplemental information

**Setdb1 safeguards proper differentiation of adult
intestinal stem cells by controlling chromatin
accessibility and transcriptome variability**

Ioanna Peraki, Liwei Zhang, Dimitris Botskaris, Marianna Stagaki, Ioannis K. Deligiannis, Haroula Kontaki, Elena Deligianni, Ioannis Giannoulakis, Orsalia Hazapis, Matthieu D. Lavigne, Celia P. Martinez-Jimenez, and Iannis Talianidis



Mouse line Name:

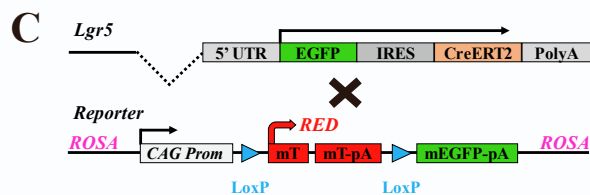
Lgr5-GFP-Cre^{ERT2}

Barker et. al. Nature (2007) 409:1003-1007



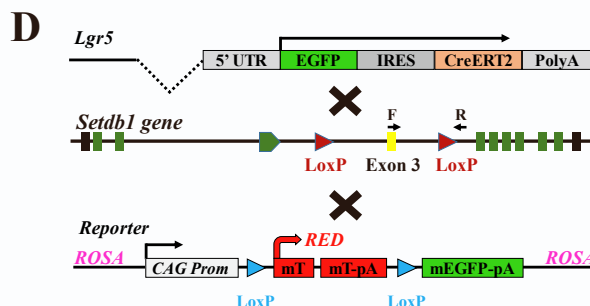
Mouse line Name:

Lgr5-GFP-Cre^{ERT2} / Setdb1^{lox/lox}



Mouse line Name:

Lgr5-GFP-Cre^{ERT2} / nTnG



Mouse line Name:

Lgr5-GFP-Cre^{ERT2} / Setdb1^{lox/lox} / nTnG

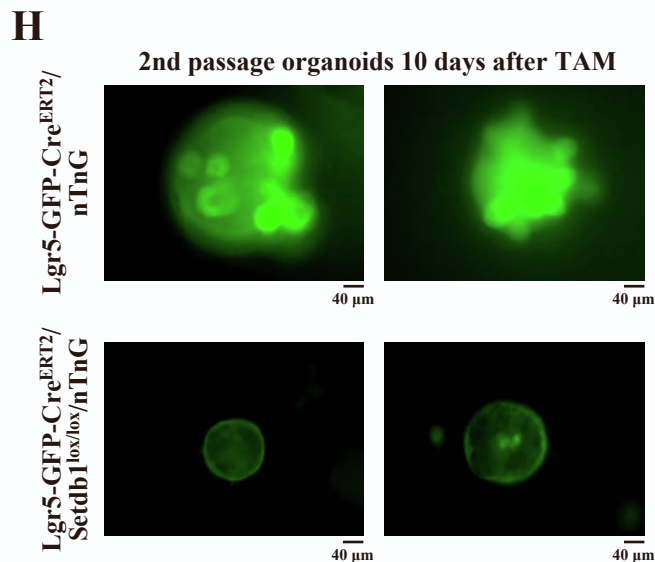
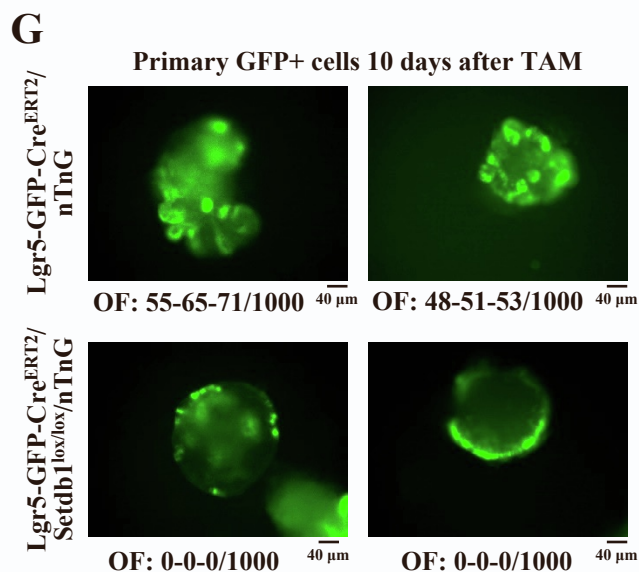
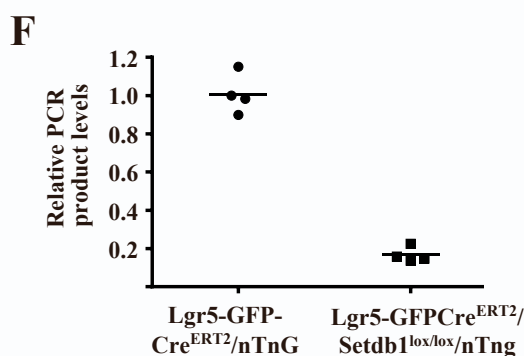
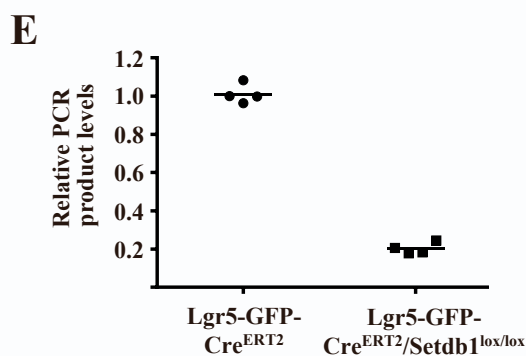


Figure S1

Schematic presentation of the genotypes of mouse models, analysis of Setdb1-exon 3 deletion and quantitation of immune-stained HNF4a⁺, Lysozyme⁺ and Mucin⁺ cells in wild-type and Setdb1-deficient ileal epithelium. Related to Figure 1 and Star Methods.

(A) Lgr5-GFP-Cre^{ERT2} mice described in Ref 22, contains one knock-in allele expressing GFP and Cre^{ERT2} specifically in Lgr5⁺ stem cells.

(B) Crossing Lgr5-GFP-Cre^{ERT2} mice with Setdb1^{lox/lox} mice generates Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox} mice, which upon tamoxifen treatment results in the deletion of exon 3 of Setdb1 and the generation of a premature stop codon in the area of exon 4, specifically in Lgr5⁺ stem cells.

(C) Crossing Lgr5-GFP-Cre^{ERT2} mice with ROSA(CAG-tdTomato*EGFP*)Ees mice generates mice named Lgr5-GFP-Cre^{ERT2}/nTnG, which express tomato-red in all cells and GFP and Cre^{ERT2} specifically in Lgr5⁺ stem cells. Upon tamoxifen treatment, all Lgr5⁺ stem cell progeny will express GFP from both Lgr5 and the ROSA locus. In parallel, these cells also loose tomato-red expression.

(D) Crossing Lgr5-GFP-Cre^{ERT2} mice with Setdb1^{lox/lox} and ROSA(CAG-tdTomato*EGFP*)Ees mice, generates mice named Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG. These mice express tomato-red in all cells and GFP and Cre^{ERT2} specifically in Lgr5⁺ stem cells. Upon tamoxifen treatment, exon 3 of Setdb1 is deleted specifically in Lgr5⁺ stem cells and all Setdb1-deficient Lgr5⁺ stem cell progeny will express GFP from both Lgr5 and the ROSA locus.

(E-F) Deletion of Setdb1 exon 3 was verified by PCR from FACS-sorted GFP⁺ cells of Tamoxifen-treated Lgr5-GFP-Cre^{ERT2} and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox} (E) or Lgr5-GFP-Cre^{ERT2}/nTnG and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG (F) mice, using a primer pair hybridizing to sequences inside Exon 3 (F) and outside the LoxP site (R). The data from 4 cell preparations and mean their values are presented as relative to the values obtained with cells from Lgr5-GFP-Cre^{ERT2} and Lgr5-GFP-Cre^{ERT2}/nTnG mice.

(G) Representative images of intestinal organoids from FACS-sorted GFP⁺ cells, isolated from two Lgr5-GFP-Cre^{ERT2}/nTnG mice (upper panel) and two Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG mice (bottom panel). After seeding, the cells were treated with 100nM tamoxifen (TAM) for 10 days, before image acquisition. Images from organoids developed from GFP⁺ cells isolated from two different mice (biological replicates) are shown. OF (Organoid Formation) values correspond to the number of fully developed, GFP-labeled, complex organoids with distinguishable crypt-like domains from 1000 seeded GFP⁺ cells at day10. Numbers from 3 parallel seedings (technical replicates) are shown.

(H) Representative images of 2nd passage organoids originating from tamoxifen-treated Lgr5-GFP-Cre^{ERT2}/nTnG organoids (upper panel) and from Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG organoids (bottom panel). Images were taken 10 days after reseeding the cells.

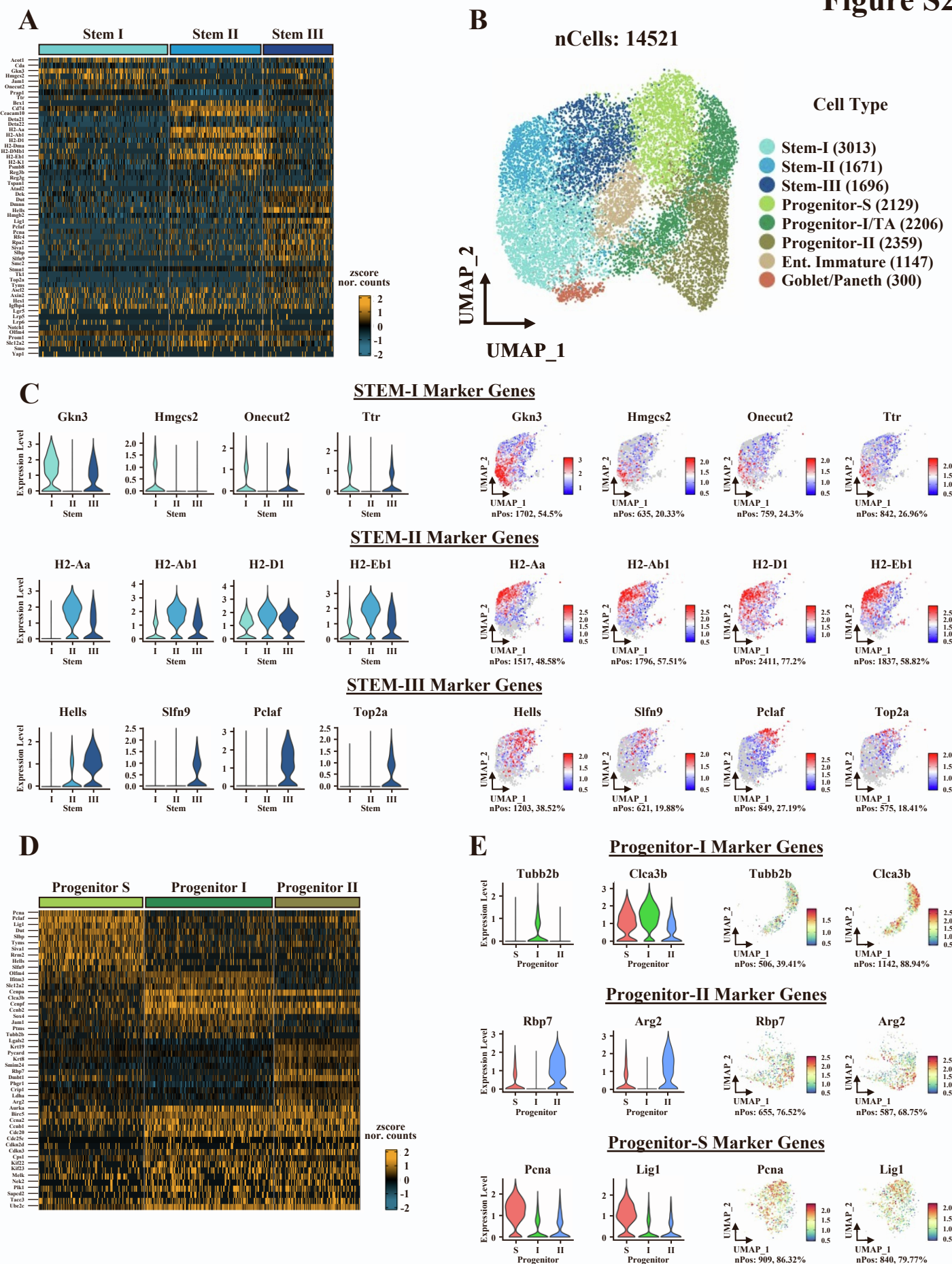


Figure S2

Gene signatures characterizing Stem cell and Progenitor cell subsets. Related to Figure 2.

(A) Heatmap of relative mRNA levels (zscore normalized counts) of genes in Stem-I, Stem-II and Stem-III cell clusters.

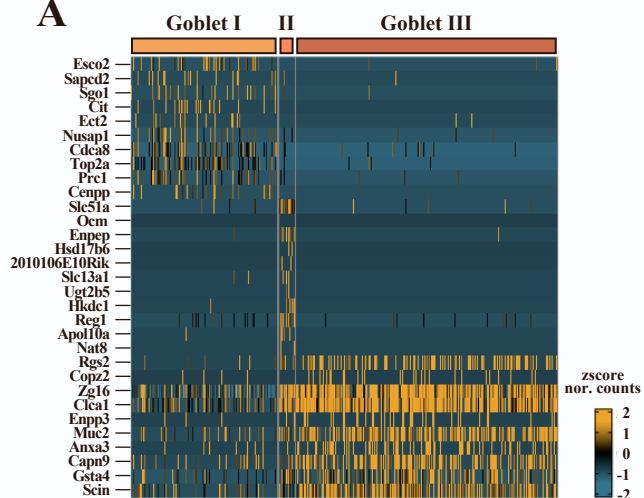
(B) UMAP projection of Stem cell and Progenitor cell types after sub-clustering of the indicated cell types. The color codes of the resulting re-clustered cell types are shown at the right.

(C) Violin plots and Feature plots of selected marker genes characterizing Stem cell subpopulations.

(D) Heatmap of relative mRNA levels (zscore normalized counts) of genes in Progenitor-I and Progenitor-II cell clusters. The gene expression profile of Progenitor S cells indicate that this cluster represents Progenitor cells in the S phase of the cell cycle.

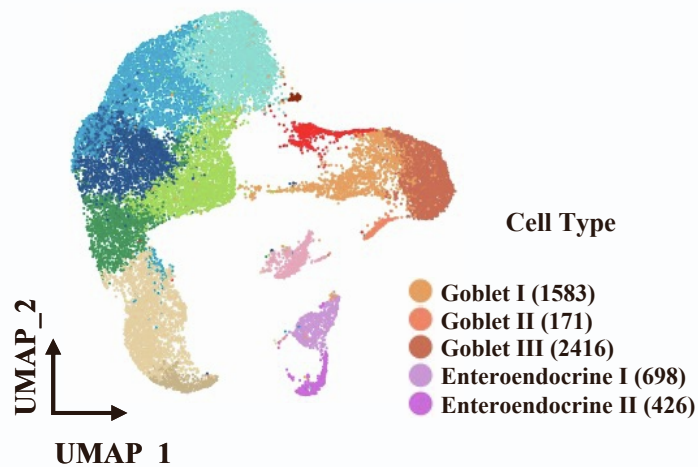
(E) Violin plots and Feature plots of selected marker genes characterizing Progenitor cell subpopulations.

A

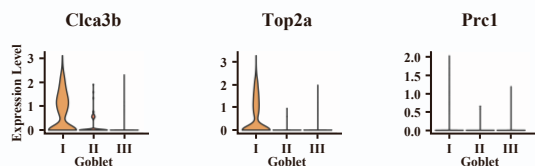
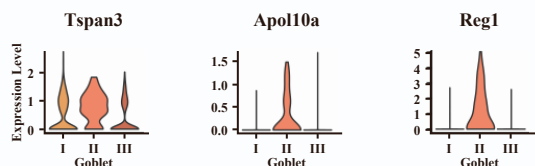
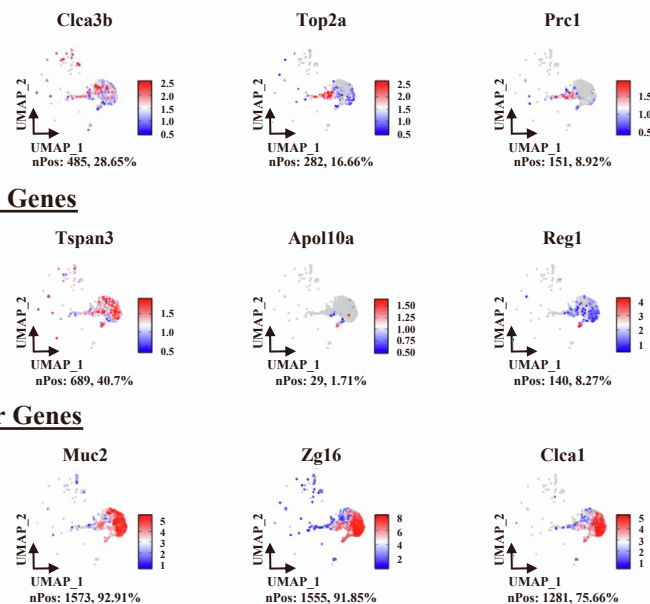
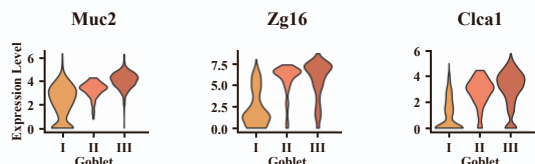


B

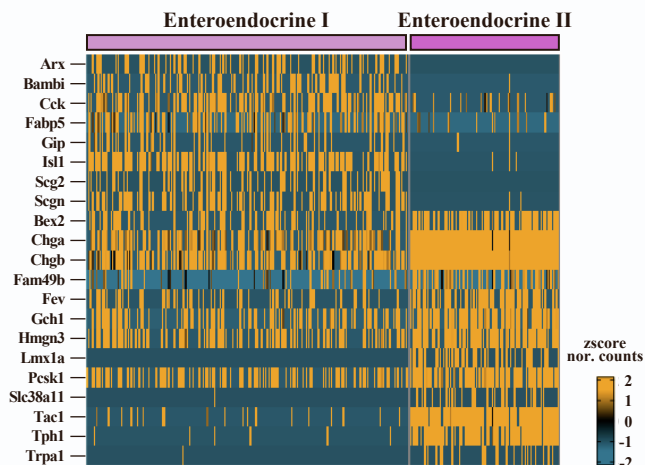
nCells: 24300



C

Goblet-I Marker GenesGoblet-II Marker GenesGoblet-III Marker Genes

D



E

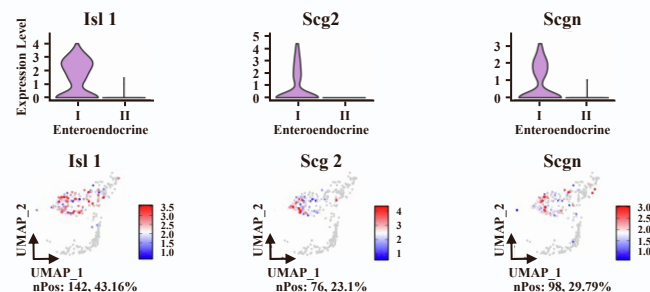
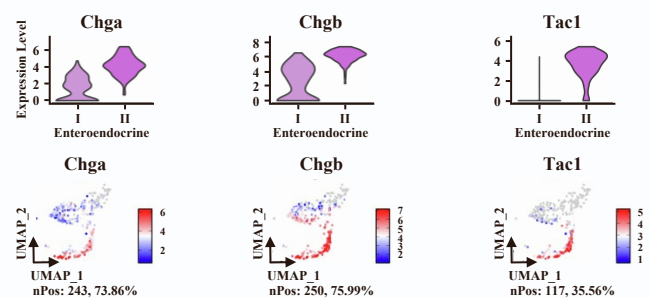
Enteroendocrine-I Marker GenesEnteroendocrine-II Marker Genes

Figure S3

Gene signatures characterizing Goblet cell and Enteroendocrine cell subsets. Related to Figure 2.

(A) Heatmap of relative mRNA levels (z-score normalized counts) of genes in Goblet-I, Goblet-II and Goblet-III cell clusters.

(B) UMAP projection of cell types from Figure 2A, used to characterize Goblet and Enteroendocrine cell subtypes.

(C) Violin plots and Feature plots of selected marker genes characterizing Goblet cell subpopulations.

(D) Heatmap of relative mRNA levels (z-score normalized counts) of genes in Enteroendocrine-I (EEC-I) and Enteroendocrine-II (EEC-II) cell clusters.

(E) Violin plots and Feature plots of selected marker genes characterizing Enteroendocrine cell subpopulations.

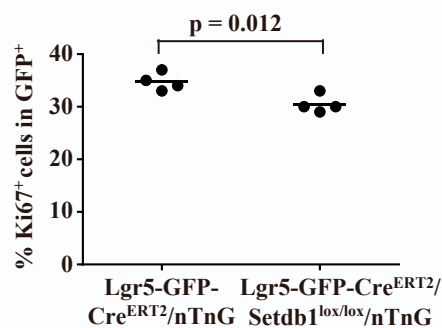
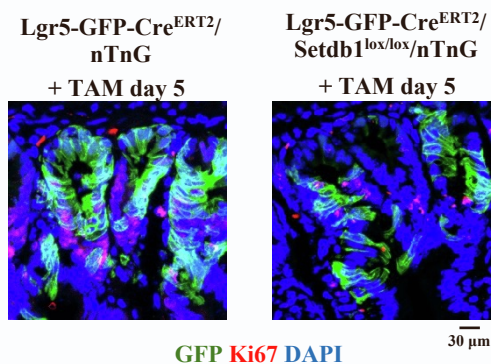
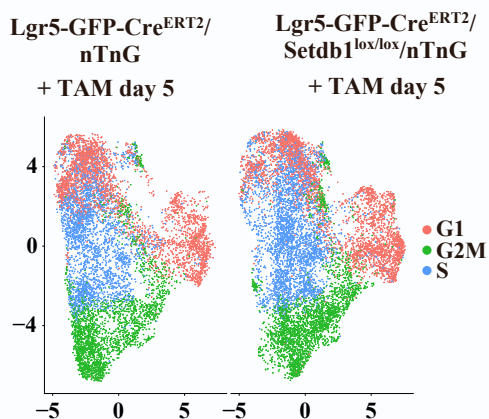
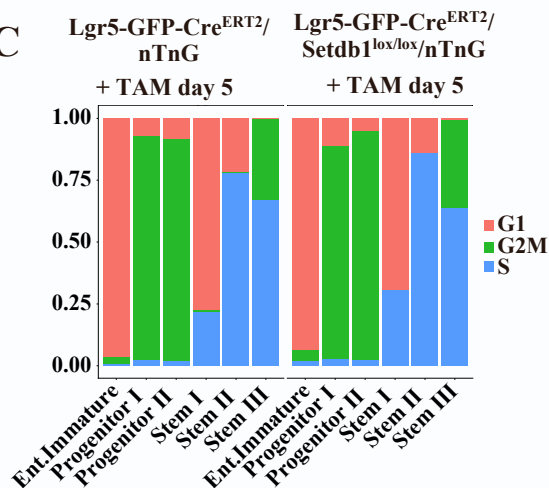
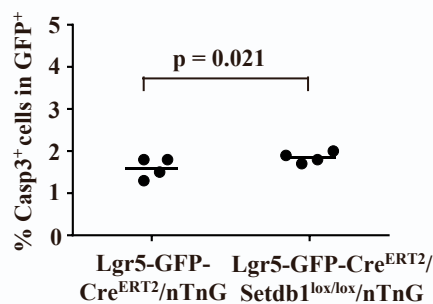
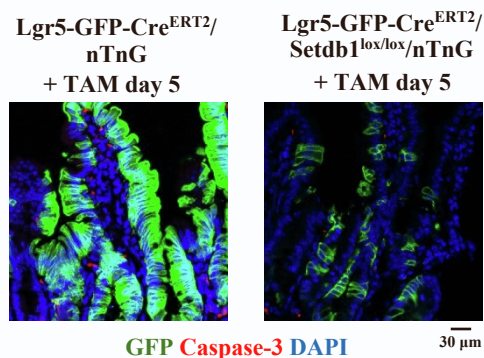
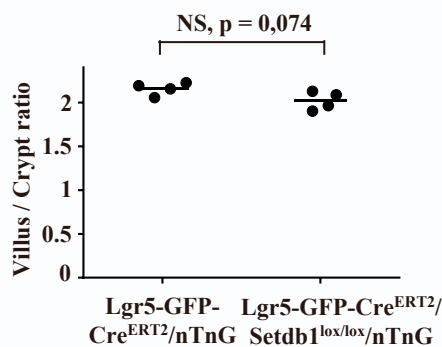
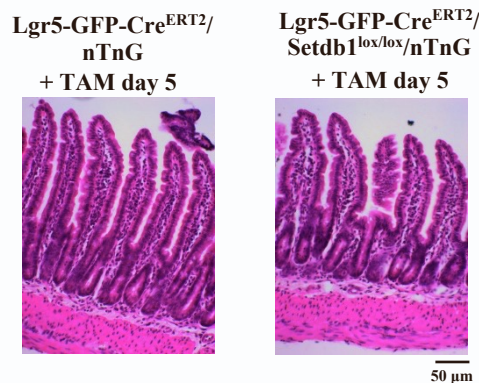
A**B****C****D****E**

Figure S4.

Analysis of cell proliferation and apoptosis in Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG cells 5 days after tamoxifen treatment. Related to Figure 4.

(A) Immunostaining with Ki67 antibody in cryosections from ileal tissue of Lgr5-GFP-Cre^{ERT2}/nTnG and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG mice 5 days after tamoxifen treatment. The graph at right shows percentages of Ki67/GFP double positive cells identified in GFP⁺ cells counted in 30 villus-crypt areas from individual mice (n=4). p-values were determined by Student's t-test. Note: Counting only GFP⁺ cells warrant the evaluation of cells originating from Lgr5⁺ cells where Setdb1 is inactivated.

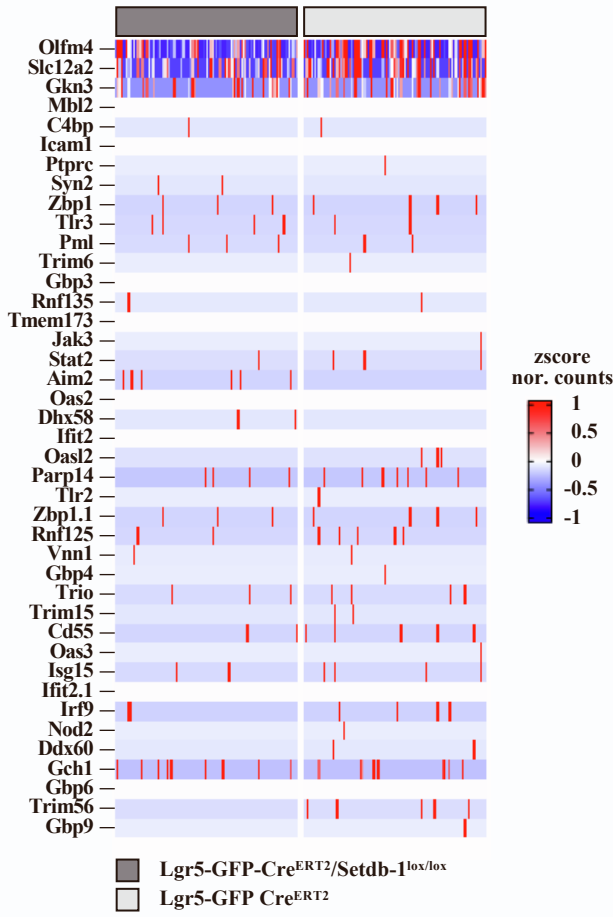
(B) UMAP projection of single cells from Tamoxifen-treated Lgr5-GFP-Cre^{ERT2}/nTnG and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG mice based on the expression of cell cycle marker genes.

(C) Relative distribution of G1, S and G2/M-phase cells in the individual cell clusters in Tamoxifen-treated Lgr5-GFP-Cre^{ERT2}/nTnG and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG mice.

(D) Immunostaining with Caspase-3 antibody in cryosections from ileal tissue of Lgr5-GFP-Cre^{ERT2}/nTnG and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG mice 5 days after tamoxifen treatment. The graph at right shows percentages of Caspase-3/GFP double positive cells identified in GFP⁺ cells counted in 30 villus-crypt areas from individual mice (n=4). p-values were determined by Student's t-test.

(E) Representative pictures of Hematoxylin and Eosin (H&E)-stained paraffin-embedded ileal tissues from Lgr5-GFP-Cre^{ERT2}/nTnG and Lgr5-GFP-Cre^{ERT2}/Setdb1^{lox/lox}/nTnG mice 5 days after tamoxifen treatment. Villus-crypt ratios were calculated from 50 ileal areas in individual mice (n=4). p-values were determined by Student's t-test.

A



B

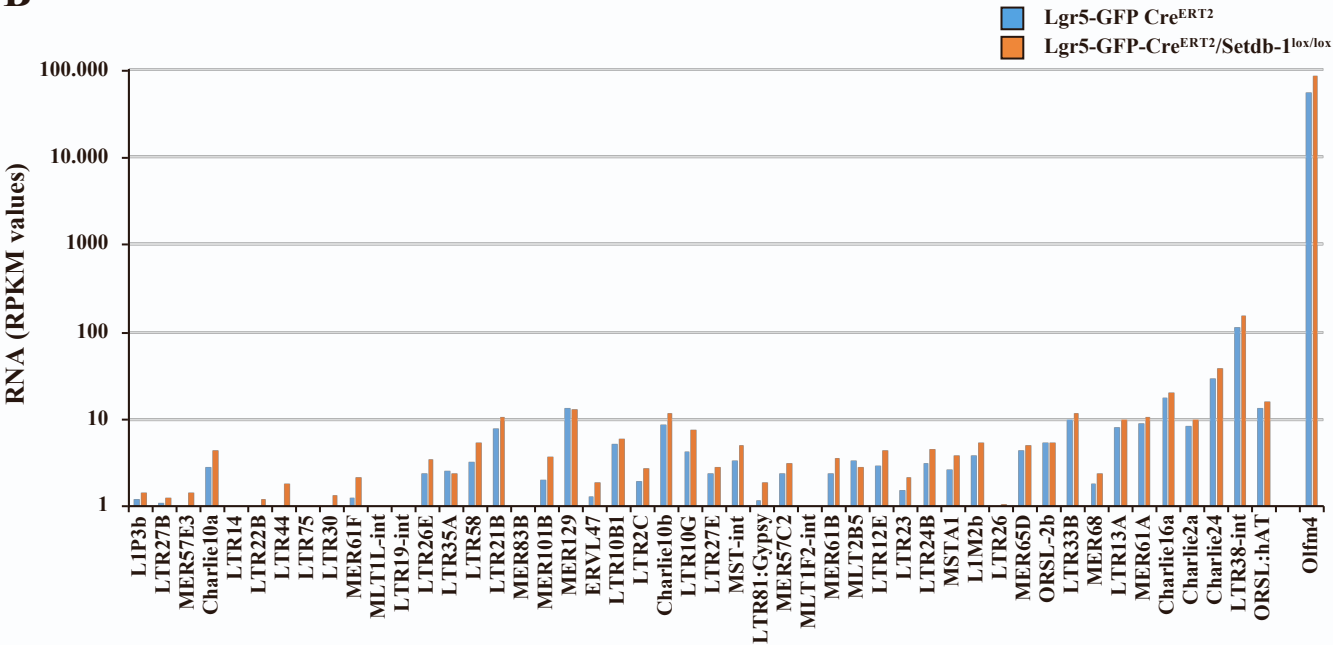


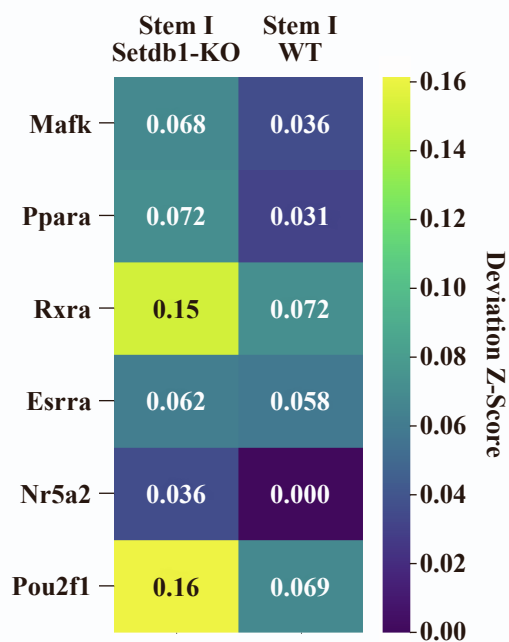
Figure S5

Lack of activation of innate immunity genes and endogenous retrovirus activation in Lgr5-GFP-Cre^{ERT2}/ Setdb1^{lox/lox} cells 5 days after tamoxifen induction. Related to Figure 2.

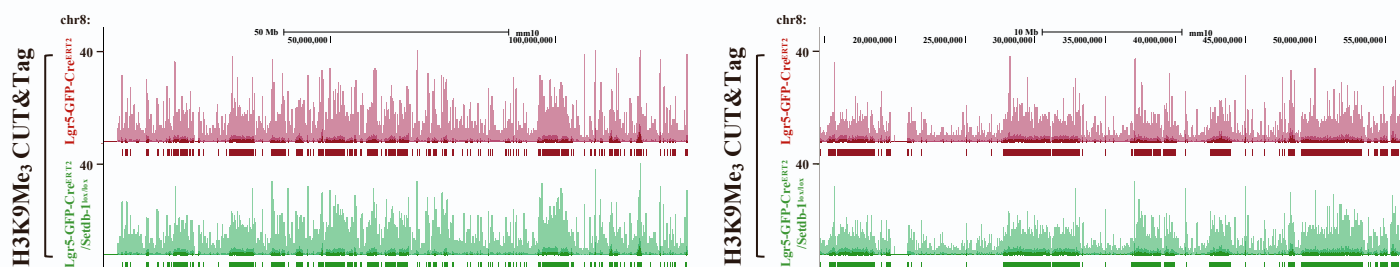
(A) Heatmap showing RNA levels of innate immunity genes in individual cells from tamoxifen-treated Lgr5-GFP-Cre^{ERT2} (WT, light gray) and Lgr5-GFP-Cre^{ERT2}/ Setdb1^{lox/lox} (Setdb1-KO, dark gray) mice. Heatmap of the highly expressed Olfm4 is included in the top panel.

(B) Bar-graph showing relative mRNA levels of the indicated endogenous retroviral transcripts in FACS-sorted GFP⁺ cells from tamoxifen-treated Lgr5-GFP-Cre^{ERT2} mice and Lgr5-GFP-Cre^{ERT2}/ Setdb1^{lox/lox} mice, as determined Tettranscript analysis of total RNA-seq data. Bars represent average RPKM values from two biological replicates. For comparison Olfm4 mRNA levels are included in the last bars. Note the 3 to 4 orders of magnitude difference between the retroviral RNAs and Olfm4 RNA levels and the small or no difference between the values obtained with wild type and Setdb1-deficient cells.

A



B



C

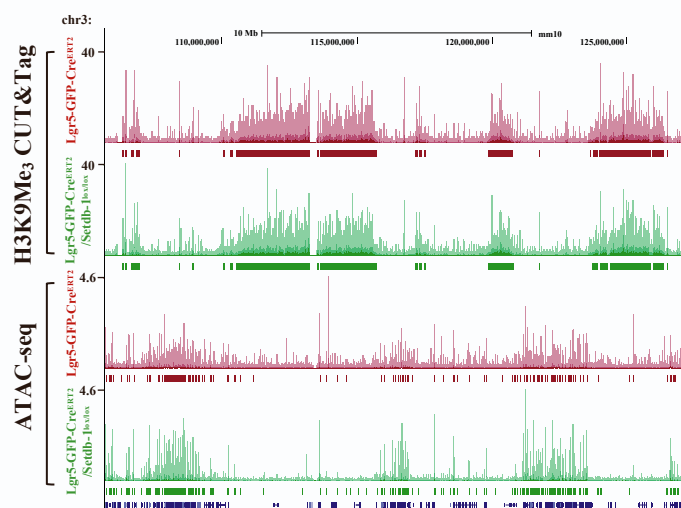


Figure S6

Comparisons of specific Transcription Factor Regulons and large H3K9Me₃-modified heterochromatin domains in wild-type and Setdb1-deficient cells. Related to Figure 5.

(A) Transcription Factor Regulons for Mafk, Ppara, Rxra, Esrra and Pou1f1 were compared using the PyScenic computational pipeline (Ref). Numbers inside the heatmaps correspond to mean AUC (Area Under Curve) scores of the indicated regulons, after ten runs of the Pyscenic algorithm on the scRNA-seq dataset of Stem-I cluster for either wild type (TAM-treated Lgr5-GFP-Cre^{ERT2}) or Setdb1-KO (TAM-treated Lgr5-GFP-Cre^{ERT2}/ Setdb1^{lox/lox}) conditions separately.

p-adjusted values for the different factors were as follows:

Pou2f1(+) 0.0; Rxra(+) $3.291e^{-179}$; Ppara(+) $2.255e^{-119}$; Mafk(+) $4.802e^{-107}$; Esrra(+) 0.999. Note that with the exception of Esrra, Setdb1 inactivation, significantly affected the downstream regulons of the interrogated Transcription Factors. For the Nr5a2 (+) we couldn't provide a statistical test, since it never appeared as an active regulon on the wildtype Stem-I cluster.

(B) Limited changes in large H3K9Me₃-modified heterochromatin domains in Setdb1-deficient cells. Genome Browser tracks showing normalized H3K9Me₃ reads and SICER-called peaks (bars shown below the tracks) along the entire chromosome 8 (left panel) and a 40 Mb region of chromosome 8, from tamoxifen-treated Lgr5-GFP-Cre^{ERT2} mice (WT) and Lgr5-GFP-Cre^{ERT2}/ Setdb1^{lox/lox} mice (Setdb1-KO). Note the highly similar overall distribution of the peak locations especially the broad peaks.

(C) Combined Genome Browser tracks of H3K9Me₃ CUT&Tag and ATAC-seq reads spanning a 20Mb region of chromosome 3. Note the tandemly organized H3K9Me₃-containing heterochromatin domains and the ATAC-seq signal containing euchromatin domains. Note the rare co-occurrence of ATAC-seq peaks and broad H3K9Me₃ CUT&Tag peaks in cells from Lgr5-GFP-Cre^{ERT2} mice, and their more frequent co-occurrence in Setdb1-KO cells.

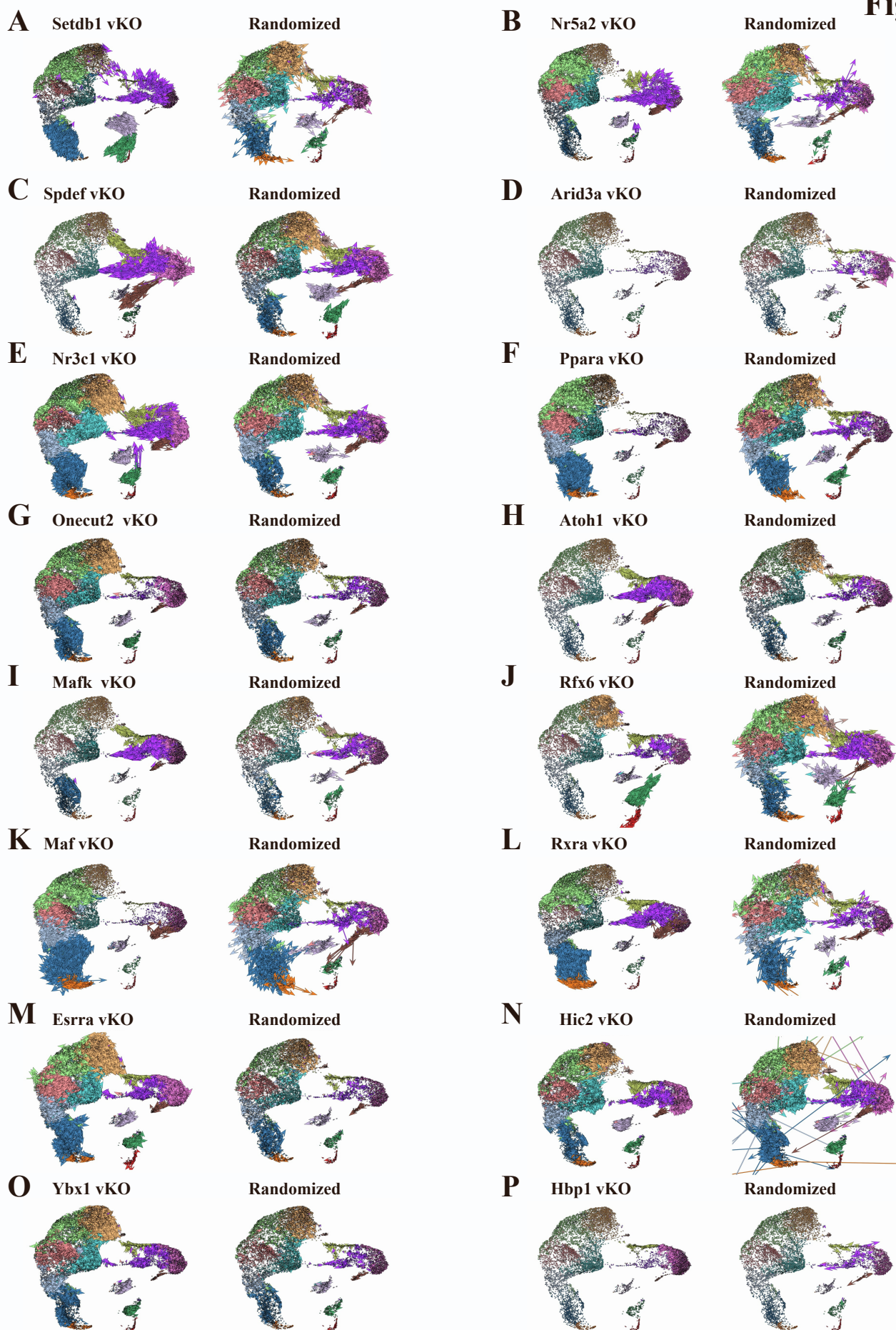


Figure S7

Quality control of in-silico perturbation experiments virtual Setdb1 and fifteen transcription factors. Related to Figure 7.

(A-P) CellOracle machine-learning algorithm⁴⁰ was employed to simulate the virtual knockout (vKO) of the transcription factors (TFs) across the single-cell manifold. Each transcription factor has two panels. Left panel shows UMAP plot displaying the predicted vector field of cell state shifts following specific TF deletion. The direction and length of the arrows represent the predicted trajectory and magnitude of transcriptomic change, respectively, for individual cells. Right panels show randomized control simulations where the regulatory network edges were shuffled, demonstrating that the directional flow observed in the specific vKO is specific to the inferred gene regulatory network. Colors represent distinct cell type clusters.

Figure S8

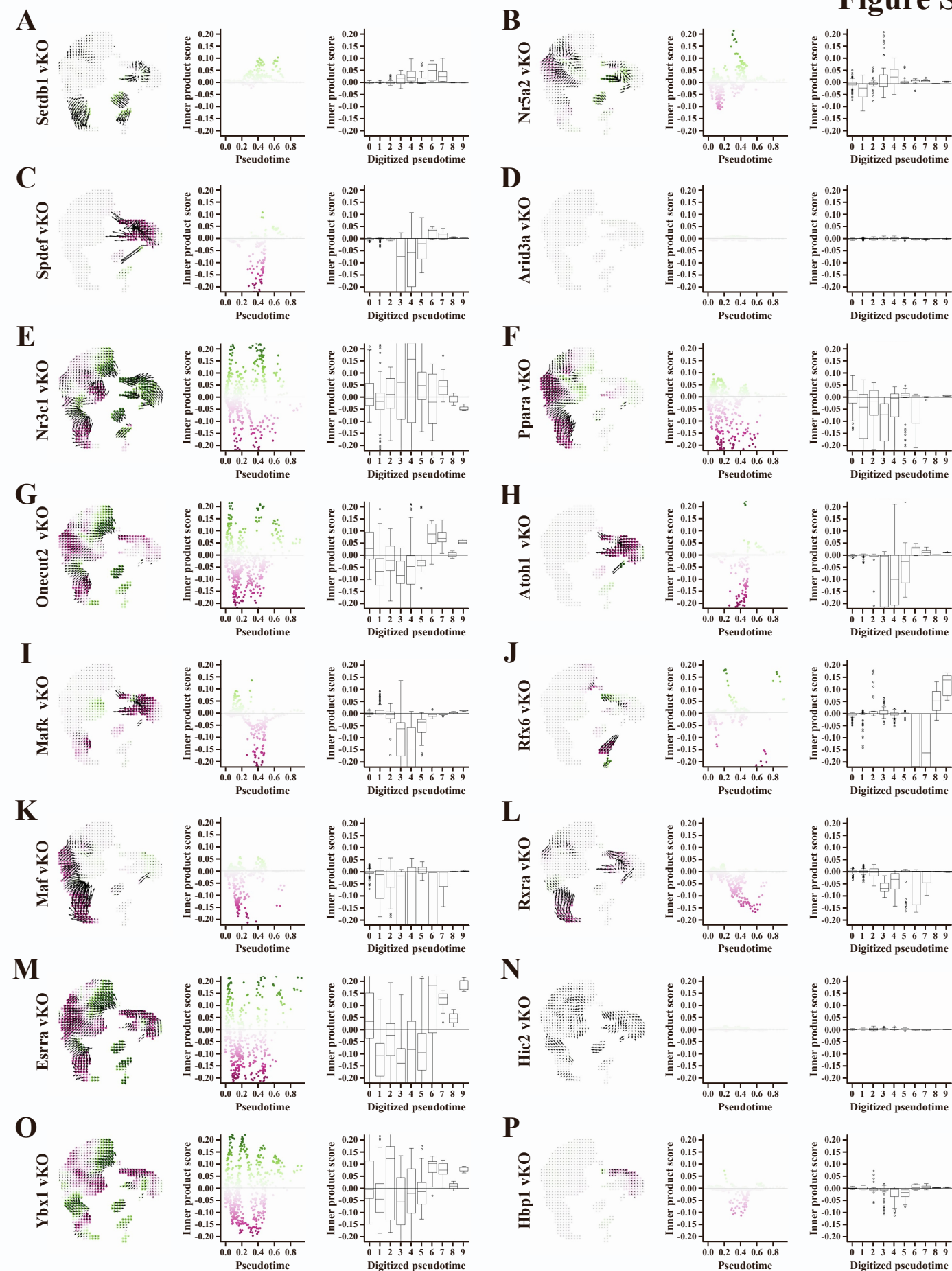


Figure S8. Quantitative comparison of CellOracle perturbation flows with the endogenous developmental program. Related to Figure 7.

Inner products between CellOracle-predicted perturbation vectors and the baseline developmental flow were calculated across the UMAP manifold for the indicated virtual knockouts (vKOs). Left panels, perturbation vector field overlaid with inner-product values (green, positive; magenta, negative). Middle panels, inner-product scores plotted against continuous pseudotime. Panels at right, boxplots of inner-product scores across digitized pseudotime bins.

Figure S9

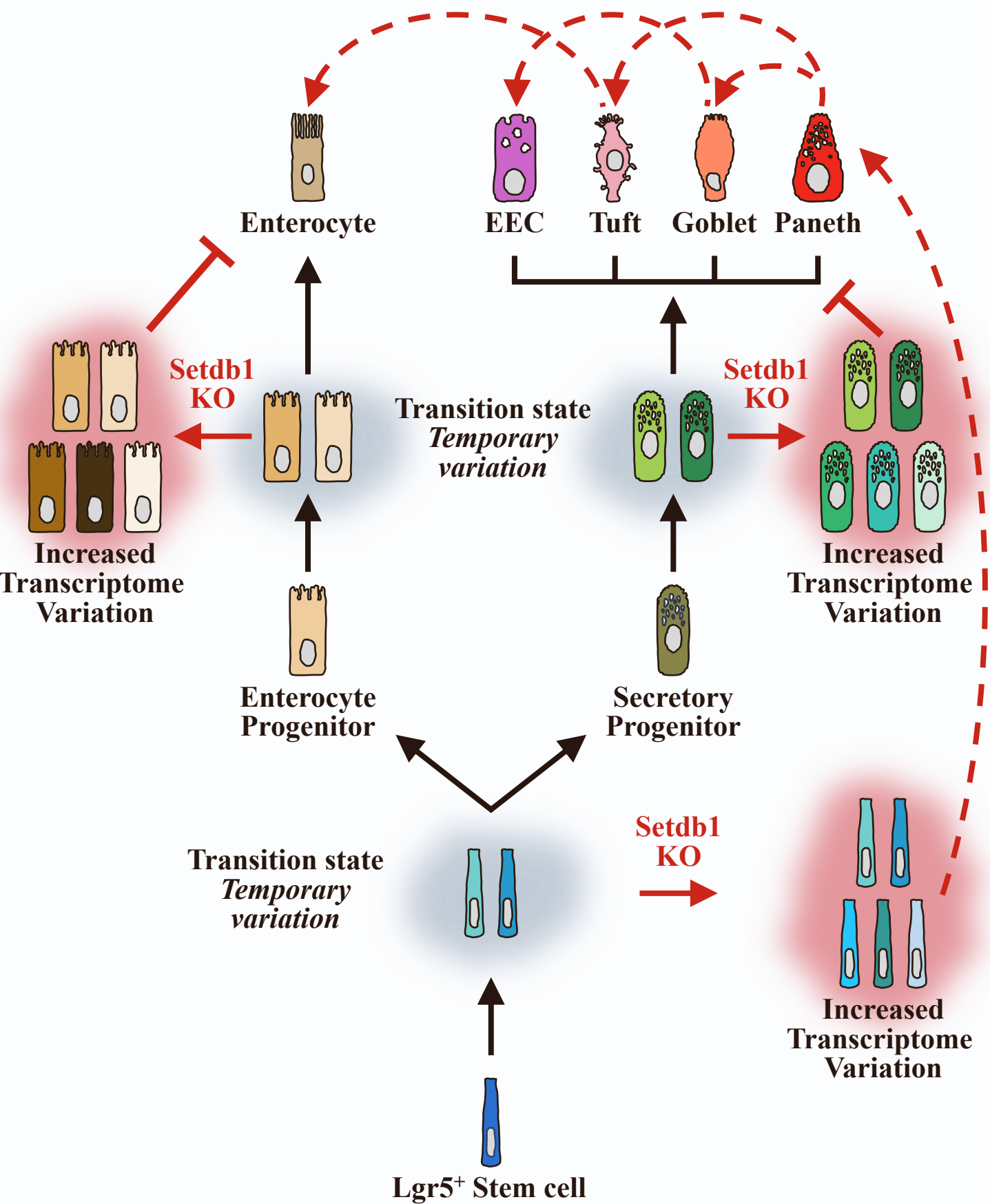


Figure S9**Model of chromatin defect-mediated increased cell-to-cell transcriptome variation during intestinal stem cell differentiation.**

The model proposes a multilevel control mechanism for the temporally distinct transcriptomes, at each consecutive differentiation state. Tight regulation promotes relatively uniform cellular transcriptomes, which is characteristic to Stem cells, Enterocyte Progenitor and Secretory Progenitors, as well as the differentiated epithelial cell types (Enterocytes, Paneth, Goblet, Tuft and Enteroendocrine cells). During the transition between the consecutive cellular differentiation states, partial transcriptome diversification occurs, leads to transition state-specific variable transcriptomes (cells in gray clouds). In *Setdb1*-KO mice, defects in chromatin structure-dependent regulation block cells in the transition states further increasing cell-to-cell transcriptional variations (cells in pink clouds), which cannot progress to fully differentiated cell types (flat-ended bar). Alternative pathways for the generation of secretory cell types directly from Stem cell population are triggered (dashed red arrow).