Supplemental information

Supplemental figures 1-5

Supplemental Methods for RNA sequencing, pathway analysis, ATAC sequencing and analyis and Chromatin immunoprecipitation (ChIP) and sequencing/analysis

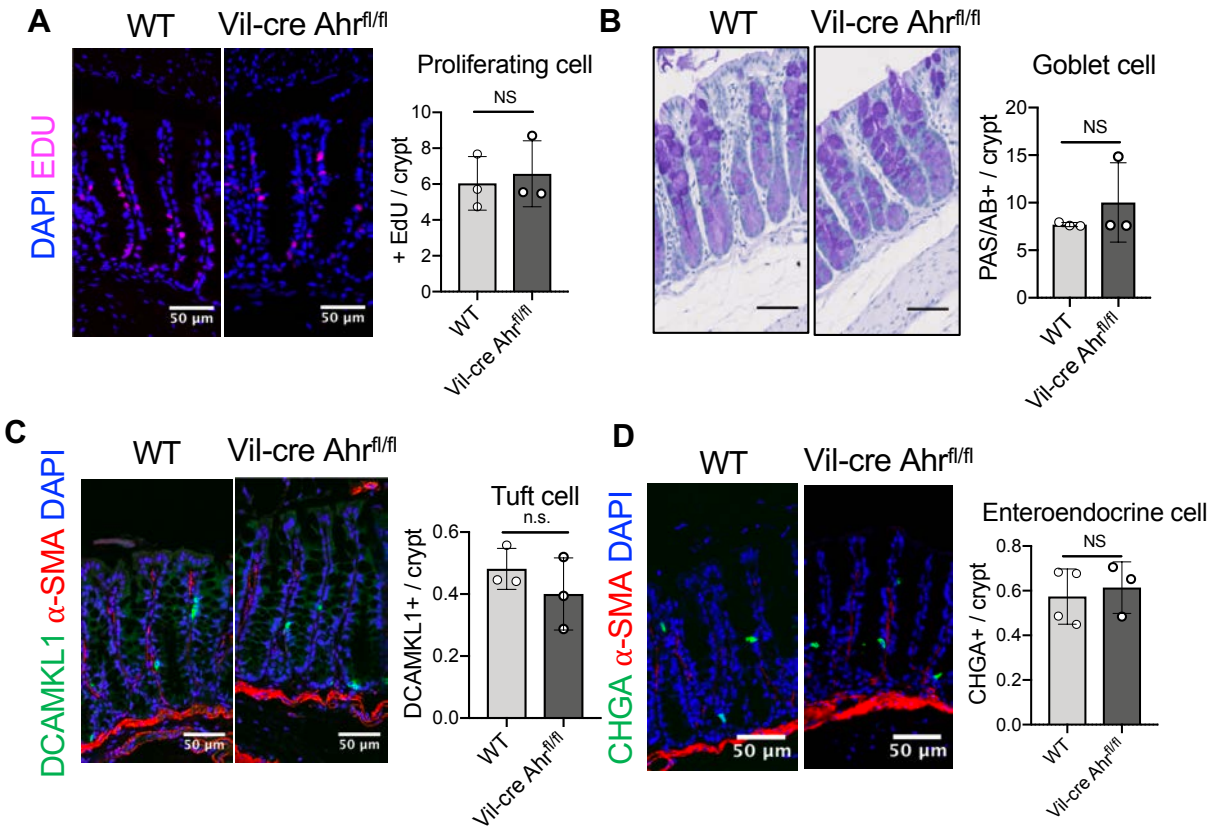
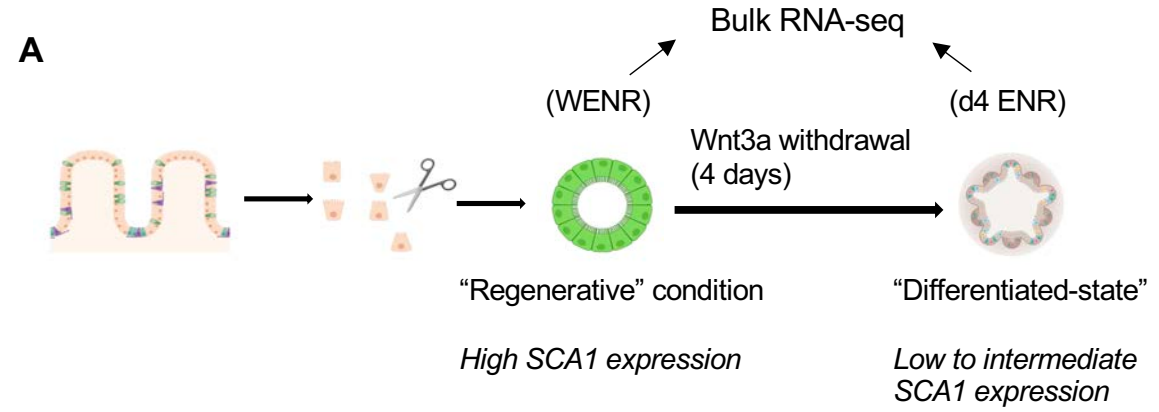
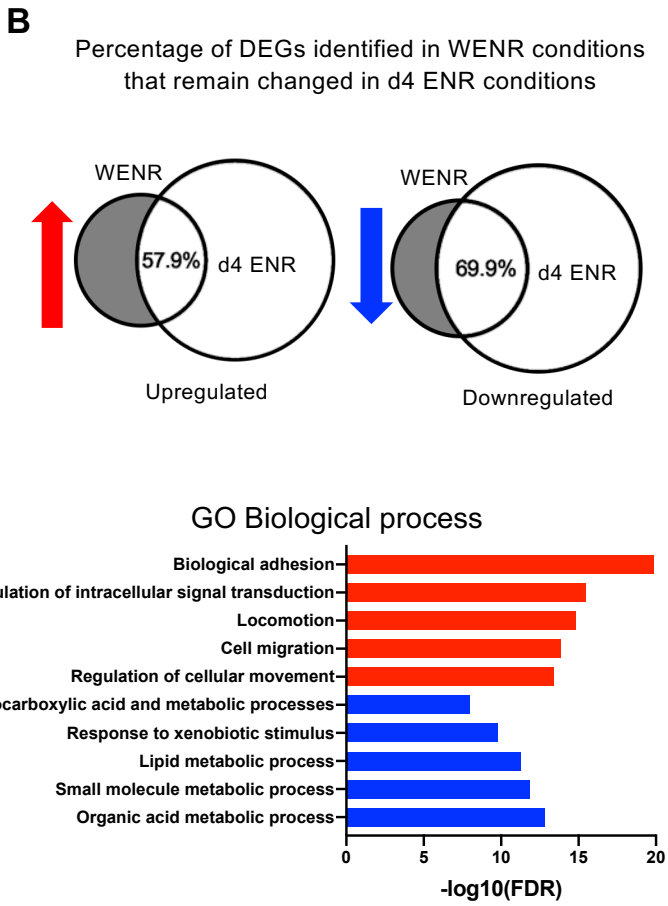


Fig.S1 related to Fig.1

Representative images and quantifications of (A) proliferating cells (2h-post EdU incorporation) and various mature epithelial subsets such as (B) goblet cells identified by PAS/AB staining (C) DCAMKL1+ Tuft cells and (D) CHGA+ Enteroendocrine cells in WT and Vil-cre AHRfl/fl mice under steady state conditions. Each data point represents the mean ± SD count per mouse, n=3-4 mice per group. Statistical test - unpaired-t tests, not significant (n.s.).





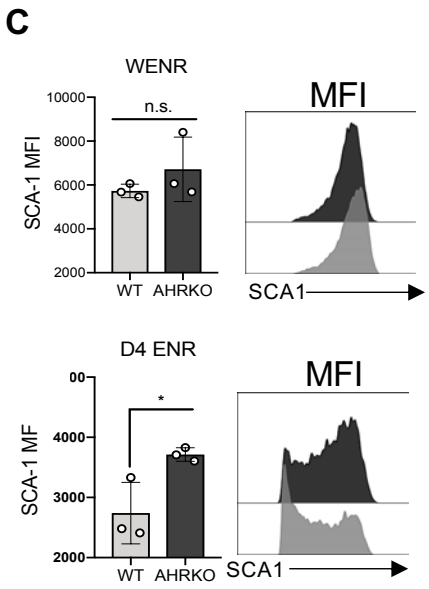


Fig.S2 related to Fig.2 and 3

(A) Schematic for RNA-seq of organoids in WENR or d4 ENR conditions. (B, left) Overlap between upregulated or downregulated genes (AHR KO v WT organoids) in WENR and d4 ENR conditions (B) gene ontology analysis (GO Biological process) of DEGs commonly up (in red) or downregulated (in blue) in AHR KO vs WT organoids grown in WENR and d4 ENR conditions. (C) % of cells expressing high surface SCA1 expression as determined by flowcytometry. Statistical test - unpaired-t tests, not significant (n.s.).

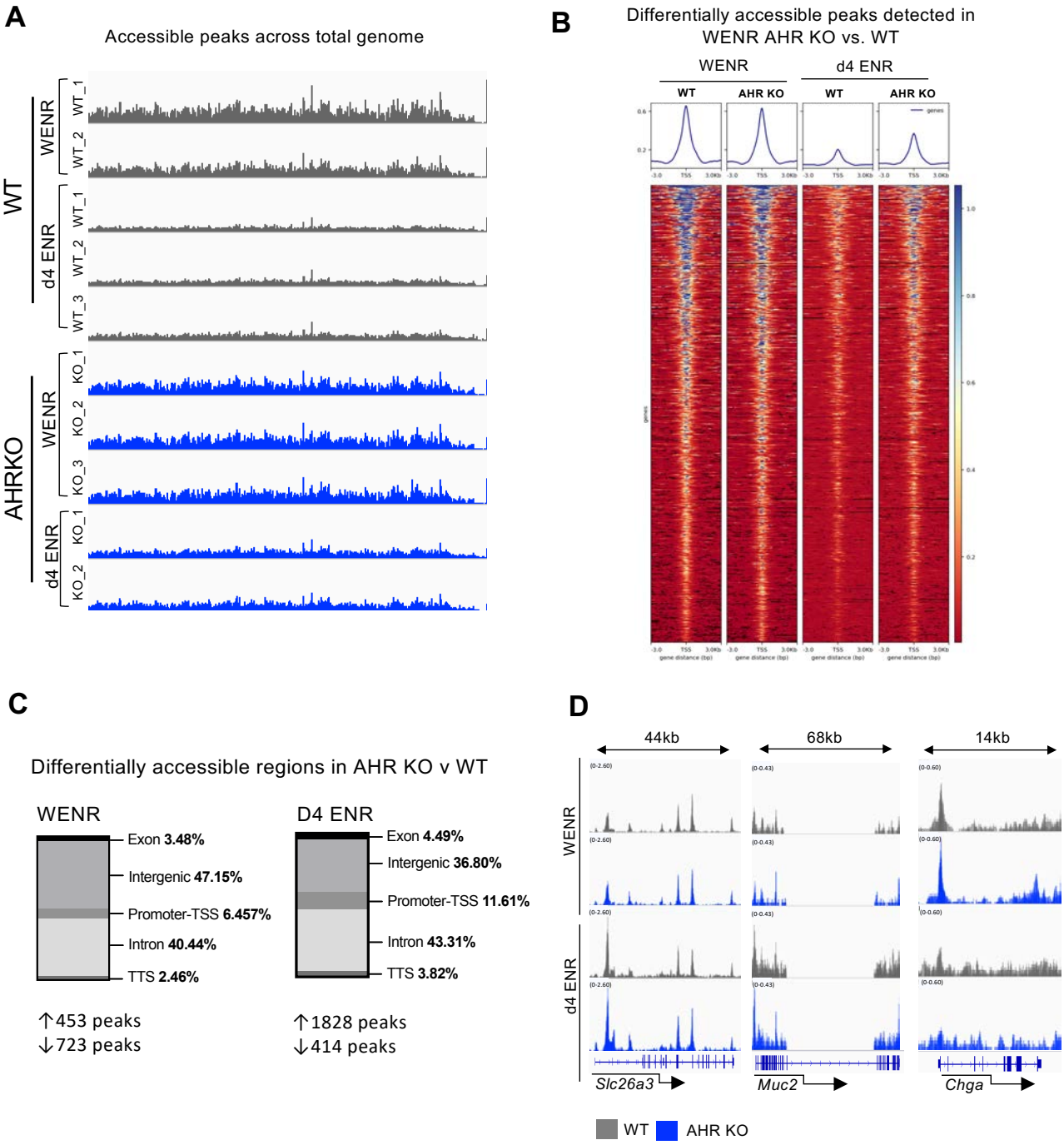


Fig.S3 related to figure 4

(A) IgV image depicting all accessible regions identified across the genome as determined by ATAC-seq (B) Deeptools heatmap showing accessibility data of differentially expressed peaks identified in AHR KO vs WT organoids across samples grown in WENR conditions (FDR< 0.05) (C) Horizontal slice graph shows regional distribution of peaks identified in AHR KO v WT organoids grown in either WENR or d4 ENR conditions (D) IgV images of accessibility peaks (open regions) of various maturation markers such as *Slc26a3*, *Muc2* and *ChgA* in d4 ENR AHR KO and WT organoids.

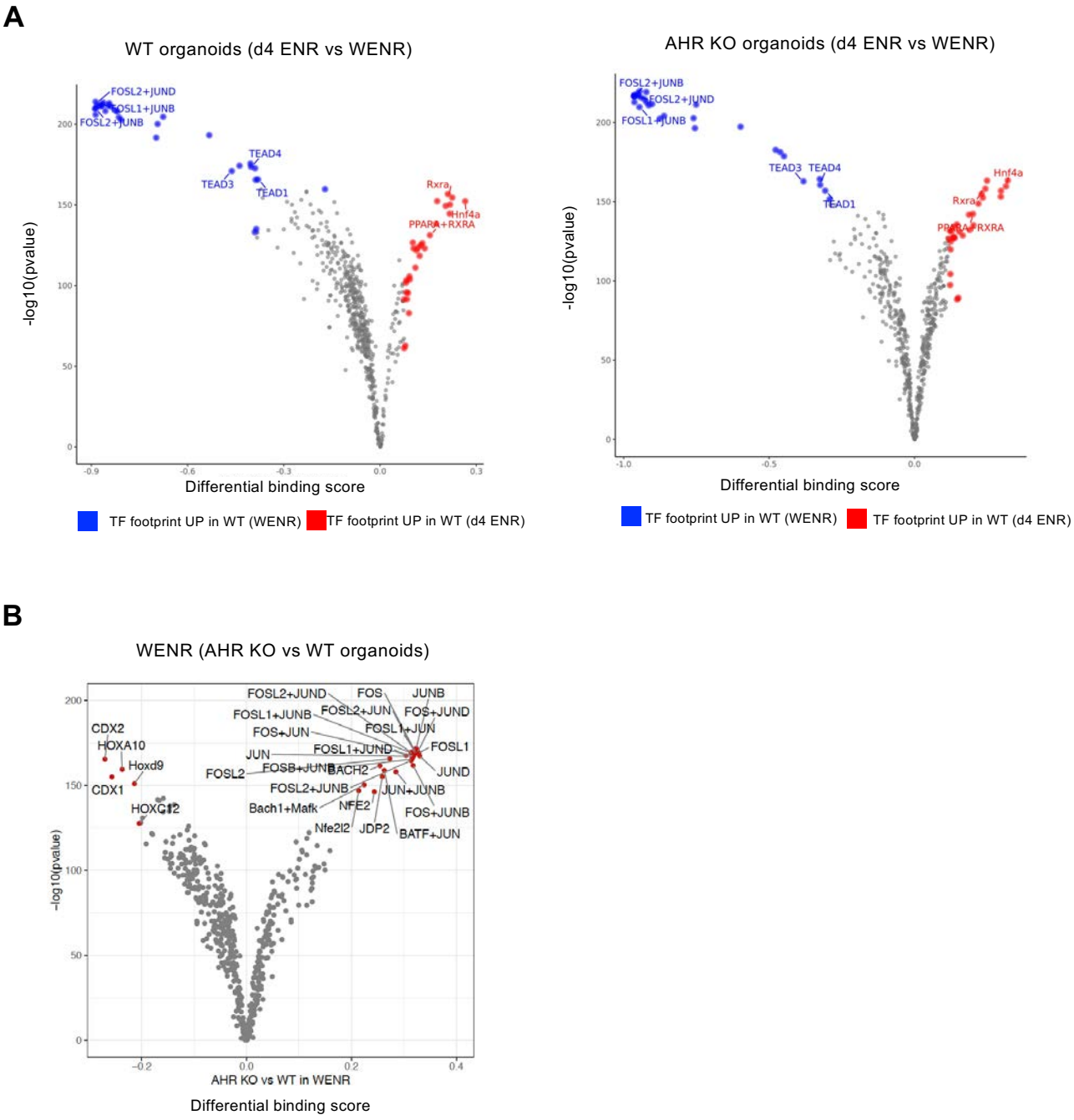
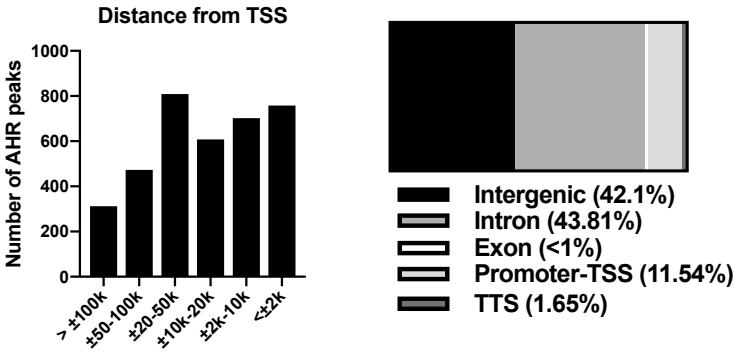


Fig.S4 related to figure 4

(A) Pairwise comparison of TF activity as WT (left panel) or AHR KO (right panel) organoids transition from WENR to d4 ENR conditions and of AHR KO vs WT organoids grown in WENR conditions. Top 5% of TFs motifs are highlighted (upregulated in red, downregulated in blue) (B) The volcano plots show the differential binding activity against the −log10(p value) (both provided by TOBIAS) of all investigated TF motifs; each dot represents one motif. Top 5% of TF motifs enriched in AHR KO vs WT organoids grown in WENR are labelled in red.

5A



5B

Rank	Motif	Transcription factor	P-value	% of targets
1		Arnt:Ahr(bHLH)	1e-32	9.45%
2		Jun-AP1(bZIP)	1e-20	5.31%
3		E2F3	1e-14	0.33%
4		ETS	1e-14	0.37%
5		BAPX1	1e-13	0.41%

Fig.S5 related to Figure 7

(A) Bar graph shows the distribution of AHR CHIP-peaks in accessible regions relative to the nearest identified TSS across the genome of FICZ-treated WT organoids grown under WENR conditions and the horizontal slice graph shows a breakdown of the genomic regions where they are found (B) HOMER Motif enrichment in AHR binding regions identified in (A).

Supplemental methods

RNA sequencing

WT and AHR-deficient generated from n=3 biological replicates were cultured in WENR or d4 ENR (differentiation experiments) or stimulated with 5nm FICZ for 4 hours at 37C. Total RNA was isolated from organoids using TRIreagent®, and cDNA was synthesized using the high-capacity cDNA reverse transcription kit according to manufacturer's instructions. Total RNA was prepared into Illumina compatible libraries using the TruSeq Stranded RNA-Seq with RiboZero Gold according to the manufacturer's instructions. Sequencing was carried out on the Illumina platform. Bam files were aligned TxDb.Mmusculus.UCSC.mm10.knownGene using GenomicAlignments summarizeOverlaps to produce raw gene counts. Further analysis was performed using DESeq2. PCA of the variance stabilized transform showed 90% of variance in the first PC corresponding to stem cells vs. differentiated organoids, 6% in the second PC corresponding to genotype and very low variance between replicates. Differentially expressed genes were called using a cutoff for adjusted p-value of 0.05. Heatmaps using RNA-seq data were generated using Morpheus (https://software.broadinstitute.org/morpheus).

Identification of activated transcriptional regulator and pathway analysis Predicted upstream transcription factors of DEGs were identified by using the Core Analysis-Upstream Regulatory analysis tool (QIAGEN's Ingenuity® Pathway analysis software) where the analysis was restricted to identified upstream "transcriptional regulators". All identified predicted upstream regulators are reported in (supplemental table 2) but only regulators with p-value of <0.0001 were considered as significant. Geneset enrichment analysis (GSEA) was conducted using the Broad institute GSEA tool (software.broadinstitute.org/gsea/index.jsp) using standard settings. We assessed overlaps with the H (hallmark gene sets), C5 (GO Biological processes geneset: Digestion, Intestinal Absorption) and C6 (Oncogenic signature genesets: Cordenonsi_Yap_conserved signature). For correlation with published data, expression dataset (.gct) files containing gene lists of DEGs identified in fetal spheroids or adult organoids were acquired from Mustata et al., 2013 (Mustata et al., 2013) the list of genes used to identity intestinal v gastric specific genes (selected from GTex repository) was acquired from Luknonin et al., 2020 (Lukonin et al., 2020). The CDX2KO organoid dataset (GSE62784) and the intestinal stem

cell specific signature was taken from a single cell RNA-sequencing dataset, were originally from Simmini et al., 2014 (Simmini et al., 2014) and Haber et al, 2017 (Haber et al., 2017) respectively.

Omni-ATAC sequencing

Wild type and AHR knockout colon organoids cultured in WENR or d4 ENR conditions were washed and digested with TrypLE buffer (12604-021, Life technologies) to make single cell suspensions and Omni-ATAC was performed as described in Buenrostro JD et al. (Buenrostro et al., 2013). In brief, 50,000 live cells were sort purified and immediately resuspended in 50 µl of ATAC-Resuspension Buffer (0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin) and incubated on ice for 3 minutes, and nuclei was spin pelleted and transposition was carried out for 30 minutes at 37C in a thermomixer. DNA was purified using the DNA Clean and Concentrator (D4014, Zymo Research) and transposed fragments were amplified for 10 cycles using NEBNext 2x MasterMix and adaptors to prepare libraries. Libraries were cleaned and quantified using Bioanalyzer followed by sequencing.

Data analysis of ATAC-seq The nf-core/atacseq pipeline (version 1.2.1) (Ewels et al., 2020) written in the Nextflow domain specific language (version 19.10.0) Di Tommaso et al., 2017) (Di Tommaso et al., 2017) was used to perform the primary analysis of the fastq samples in conjunction with Singularity (version 2.6.0) (Kurtzer et al., 2017). The command used was "nextflow run nf-core/atacseq -profile crick --input /Path_to_desing/design.csv --genome GRCm38". To summarise, the pipeline performs adapter trimming (Trim Galore! - https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), read alignment (BWA) and filtering (SAMtools (Li et al., 2009), BEDTools (Quinlan and Hall, 2010); BamTools (Quinlan and Hall, 2010); pysam - https://github.com/pysam-developers/pysam; picard-tools - http://broadinstitute.github.io/picard), normalised coverage track generation (BEDTools (Quinlan and Hall, 2010); bedGraphToBigWig (Kent et al., 2010)), peak calling (MACS) and annotation relative to gene features (HOMER (Heinz et al., 2010)), consensus peak set creation (BEDTools), differential binding analysis (featureCounts (Liao et al., 2014) R Core Team; DESeq2 (Love et al., 2014)) and extensive QC and version reporting (MultiQC (Ewels et al., 2020); FastQC (Daley and Smith, 2013); deepTools (Ramírez et al., 2016); ataqv (Orchard et al., 2020)). All data was processed relative to the mouse UCSC mm10 genome. A set of consensus peaks was created by selecting peaks that appear in at least one sample. Counts per peak per sample were then imported on DESeq2 within R

environment for differential expression analysis. Pairwise comparisons between genotypes in each condition, and between conditions per genotype were carried out and differential accessible peaks were selected with an FDR < 0.05. Heatmaps for differentially accessible peaks were generated using Deeptools.

For footprinting analysis TOBIAS (v 0.12.10) (Bentsen et al., 2020) was used by running the following pipeline (https://github.com/luslab/briscoe-nf-tobias). The pipeline runs TOBIAS' ATACorrect, ScoreBigwig, BINDetect and generates PlotAggregate metaplots on merged replicate bam files. TOBIAS was run on set of consensus peaks used for the differential analysis (see above), with the flag "--output-peaks" within TOBIAS BINDetect to set a different peak set for the output analysis. Differential binding scores and p-values are visualized as a volcano plot per pairwise comparison. JASPAR 2018 was used as the motif database for the foot printing analysis. As described before (Bentsen et al., 2020), all TFs with -log10(pvalue) above the 95% quantile or differential binding scores smaller/larger than the 5% and 95% quantile are colored. Selected TFs are also shown with labels. The full set of differential binding scores is included as supplemental table 6.

Chromatin immunoprecipitation (ChIP) and sequencing

WT and AHR knockout colon organoids generated from n=4 biological replicates grown in WENR condition were stimulated with 5nM FICZ for 1 hour in a glass vial at 37^{o}C. Samples were crosslinked with 1% formaldehyde (catalogue number: 28906; Thermofisher) for 8 minutes and quenched with 0.125M glycine at room temperature. Nuclei was extracted by Dounce homogenization and chromatin was sheared by sonication in a lysis buffer (0.25% SDS, 1mM EDTA, 1% Trition-X, 10mM Tris-HCl, pH 8). Debris was spin pelleted at 13,000 rpm for 30min at 4^{o}C. Supernatant was diluted in RIPA buffer (0.2% deoxycholate, 280mM NaCl, 1.1% Triton-X in H_2O) and cleaned with protein A/G Dynabeads (10003D & 10001D; Invitrogen) for 1 hour at 4^{o}C. 10% of supernatant was used as input and the remaining sample was immunoprecipitated overnight at 4^{o}C with anti-AhR antibody (BML-SA210; Enzo). Chromatin was then incubated with protein A/G Dynabeads for 3 hours at 4^{o}C. Chromatin bound beads were washed twice in low-salt buffer (0.1% SDS, 1% Triton X-100, 1mM EDTA, , 140 mM NaCl, 0.1% deoxycholate, 20 mM Tris–HCl, pH 8), high-salt buffer (same as low salt except for 500 mM NaCl), LiCl wash buffer (250mM LiCl, 0.5% NP40, 0.5% deoxycholate, 1 mM EDTA, 10 mM Tris–HCl, pH 8)

and with Tris-EDTA buffer (10 mM Tris–HCl 1mM EDTA). Protein-chromatin complexes were then eluted in elution buffer (1% SDS, 100 mM NaHCO3) and reverse cross-linked overnight with RNase and proteinase containing buffer (62.5 µg/ml RNaseA, 5 mg/ml proteinase K, 1.25M NaCl, 62.5mM EDTA pH 8, 250mM Tris-HCl pH 6.5). ChIP DNA was purified using the ChIP DNA Clean and Concentrator (D5205, Zymo Research) and quantified using Bioanalyzer followed by sequencing.

Data analysis for CHIP-seq Samples were sequenced on an Illumina HiSeq4000 generating 101bp single ended reads averaging 30 million reads per sample. ChIP-seq reads were aligned to the mouse mm10 genome assembly using BWA version 0.7.15 (Li et al., 2009)with a maximum mismatch of 2 bases. Picard tools version 2.1.1 (http://broadinstitute.github.io/picard) was used to sort, mark duplicates and index the resulting alignment BAM files. Normalised tdf files for visualisation purposes were created using the resulting BAM files using IGVtools (http://www.broadinstitute.org/igv) version 2.3.75 software by extending reads by 50 bp and normalising to 10 million mapped reads per sample. Peaks were called by comparing IP samples to their respective input using MACS version 2.1.1, (Zhang et al., 2008) using the standard parameters. Peaks called by MACS were annotated using the 'annotatePeaks' function in the Homer (version 4.8) (Heinz et al., 2010) software package. Common and unique peaks across experiments were determined using the 'intersect' function in Bedtools (version 2. 26.0) (Quinlan and Hall, 2010) and custom scripts.

Reference list for supplemental methods

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T.*, et al.* (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun *11*, 4267.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Daley, T., and Smith, A.D. (2013). Predicting the molecular complexity of sequencing libraries. Nat Methods *10*, 325-327.

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nat Biotechnol *35*, 316-319.

Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol *38*, 276-278.

Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y.*, et al.* (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333-339.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204-2207.

Kurtzer, G.M., Sochat, V., and Bauer, M.W. (2017). Singularity: Scientific containers for mobility of compute. PLoS One *12*, e0177459.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lukonin, I., Serra, D., Challet Meylan, L., Volkmann, K., Baaten, J., Zhao, R., Meeusen, S., Colman, K., Maurer, F., Stadler, M.B.*, et al.* (2020). Phenotypic landscape of intestinal organoid regeneration. Nature *586*, 275-280.

Mustata, R.C., Vasile, G., Fernandez-Vallone, V., Strollo, S., Lefort, A., Libert, F., Monteyne, D., Pérez-Morga, D., Vassart, G., and Garcia, M.I. (2013). Identification of Lgr5-independent spheroid-generating progenitors of the mouse fetal intestinal epithelium. Cell Rep *5*, 421-432.

Orchard, P., Kyono, Y., Hensley, J., Kitzman, J.O., and Parker, S.C.J. (2020). Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with ataqv. Cell Syst *10*, 298-306.e294.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

Simmini, S., Bialecka, M., Huch, M., Kester, L., van de Wetering, M., Sato, T., Beck, F., van Oudenaarden, A., Clevers, H., and Deschamps, J. (2014). Transformation of intestinal stem cells into gastric stem cells on loss of transcription factor Cdx2. Nat Commun *5*, 5728.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.