Intestinal stem cell differentiation is associated with dynamic changes in the chromatin landscape

Jesse R Raab[1], Kortney E Wager[1], Jeremy M Morowitz[1], Scott T Magness[2,3,4], Adam D Gracz[1,2,5]

Affiliations:
[1] Department of Genetics, University of North Carolina at Chapel Hill
[2] Center for Gastrointestinal Biology and Disease, University of North Carolina at Chapel Hill
[3] Department of Cell Biology & Physiology, University of North Carolina at Chapel Hill
[4] Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill/North Carolina State University
[5] Lead contact: agracz@med.unc.edu

The authors declare no conflicts of interest

Author contributions:
ADG conceived the study. JRR, STM, and ADG designed experiments. JRR, KW, JMM, and ADG performed experiments and analyzed data. JRR, STM, and ADG wrote and edited the manuscript.

SUMMARY

Chromatin is central to cellular identity and is remodeled as stem cells differentiate and acquire mature cell fates. In the adult intestinal epithelium, functional plasticity facilitates de-differentiation and reacquisition of stemness by progenitors and some post-mitotic cell types following damage to the intestinal stem cell (ISC) pool. Previous studies have suggested that this plasticity is regulated by largely homogenous or "static" chromatin landscapes shared by ISCs and their progeny. Here, we utilize a Sox9^{EGFP} reporter mouse to examine native chromatin across the continuum of ISC differentiation. Our data reveal a highly dynamic chromatin landscape in IECs, associated mainly with open chromatin regions (OCRs) found in intergenic or intragenic loci consistent with putative distal enhancers. By integrating gene expression with chromatin accessibility at transcription factor (TF) binding motifs in context of Sox9^{EGFP} populations, we classify "static" and "dynamic" chromatin relative to specific TF usage. These analyses identify known and potential regulators of ISC differentiation via their association with dynamic changes in chromatin and suggest that differential *cis*- and *trans*-regulatory significance can be defined in a locus-specific manner. Finally, we apply the Sox9^{EGFP} model to examine the relationship between gene expression and a specific chromatin modification, 5-hydroxymethylcytosine (5hmC), which reveals a role for 5hmC in absorptive lineage specification. Together, our data demonstrate a previously unappreciated degree of global chromatin changes in intestinal epithelial cells (IECs), identify novel potential regulators of ISC differentiation, and provide a chromatin roadmap for further dissecting the role of *cis* regulation of cell fate in the intestine.

INTRODUCTION

ISCs maintain the intestinal epithelium, which is replaced once a week throughout adult life. To preserve digestive and barrier function, ISCs balance proliferation with differentiation into post-mitotic IECs: absorptive enterocytes, goblet cells, Paneth cells, enteroendocrine cells (EECs), and tuft cells. Models for cellular hierarchy in the intestinal epithelium have been complicated by observations of extensive cellular plasticity, wherein secretory and absorptive progenitors as well as post-mitotic Paneth cells can de-differentiate following damage to or ablation of the ISC pool (Buczacki et al., 2013; Schmitt et al., 2018; Tetteh et al., 2016). The genetic mechanisms that allow IECs to adopt metastable differentiated fates while exhibiting facultative ISC function are not fully understood.

Chromatin landscapes consist of histone post-translational modifications, DNA modifications, and higher-order structural organization, and are known to exert regulatory control on cell fate. Chromatin is progressively remodeled during differentiation, and is highly dynamic in embryonic stem cells as well as tissue-specific stem cells in the hair follicle and hematopoietic system (Dixon et al., 2015; Lara-Astiaso et al., 2014; Lien et al., 2011). Recent studies of chromatin in ISCs and differentiated IECs suggest that intestinal chromatin is largely homogenous, with similar enhancer, open chromatin, and DNA methylation profiles observed in ISCs and differentiated progeny (Kaaij et al., 2013; Kim et al., 2014). The observed "static", or unchanging, nature of chromatin has been proposed as a mechanism for cell fate plasticity in the intestine (Kim et al., 2014). However, these studies relied on pharmacologic or genetic modulation of intestinal epithelial populations to force over-production of absorptive or secretory cells, or have compared *Lgr5*+ ISCs to bulk differentiated populations (Barker et al., 2007; Kaaij et al., 2013; Kim et al., 2014). Whether similarly homogenous chromatin profiles are present in native IECs, or specific IEC subpopulations, remains unknown.

Differential expression levels of *Sox9*, observed via a Sox9^{EGFP} BAC transgene, are associated with phenotypically distinct IEC populations. Histological and gene expression studies have shown that *Sox9*^{low} cells are consistent with ISCs, *Sox9*^{sublow} cells with transit-amplifying (TA) progenitors, *Sox9*^{high} cells with label-retaining secretory progenitors/facultative ISCs in the crypts and mature EECs and tuft cells in the villus, while *Sox9*^{neg} cells consist of the remaining post-mitotic populations (goblet, Paneth, and absorptive enterocyte) (Formeister et al., 2009; Gracz et al., 2010; Gracz et al., 2015; Roche et al., 2015). We reasoned that this model could be applied to omics-scale assays to understand dynamics in native IEC chromatin. Here, we profile the transcriptome, open chromatin (Omni ATAC-seq), and 5hmC (Nano-hmC-Seal) to interrogate the relationship between gene expression changes and intestinal chromatin dynamics in the Sox9^{EGFP} model. These analyses demonstrate clear examples of both static and dynamic chromatin in IECs and suggest that specific transcription factors define the distinct chromatin landscapes found in subpopulations of IECs.

RESULTS

Sox9 expression levels define transcriptomically distinct populations

We isolated *Sox9*^{neg}, *Sox9*^{sub}, *Sox9*^{low}, and *Sox9*^{high} IEC populations from adult Sox9^{EGFP} mice by fluorescence-activated cell sorting (FACS) and subjected them to RNA-seq (Fig. 1A&B, Fig. S1A), which confirmed that Sox9^{EGFP} levels faithfully recapitulate endogenous *Sox9* expression (Fig. S1B) (Formeister et al., 2009). Differential expression analysis identified 2,852, 253, 1,499, and 3,521 genes that were upregulated in *Sox9*^{neg}, *Sox9*^{sub}, *Sox9*^{low}, and *Sox9*^{high} populations, respectively (Fig. 1C and Supplemental Table 1). Previous studies have defined Sox9^{EGFP} populations by RT-qPCR and immunofluorescence, and we next sought to confirm our model with transcriptomic data (Formeister et al., 2009; Gracz et al., 2010). Gene Ontology (GO) analysis identified expected relationships between Sox9^{EGFP} populations and cellular function, including mitosis in *Sox9*^{sub} progenitors, stem cell maintenance in *Sox9*^{low} ISCs, secretory and hormone processes in EEC-enriched *Sox9*^{high}, and absorption and secretion in *Sox9*^{neg}, which contains absorptive enterocytes as well as secretory goblet and Paneth cells (Fig. S1D).

We then compared gene expression levels from each Sox9^{EGFP} population against gene sets that define distinct IEC subtypes in previously published scRNA-seq data (Fig. 1D) (Haber et al., 2017). Consistent with our prior characterization of Sox9^{EGFP} populations, genes that define absorptive enterocytes and Paneth cells were enriched in *Sox9*^{neg}, TA and enterocyte progenitors (EP) in *Sox9*^{sub}, *Lgr5*-positive ISCs in *Sox9*^{low}, and EECs and tuft cells in *Sox9*^{high}. Similar trends of enrichment between Sox9^{EGFP} populations and IEC-specific gene sets were observed in comparisons to a second, independently-generated scRNA-seq data set (Fig. S1C) (Yan et al., 2017). We also validated correlation between Sox9^{EGFP} levels and protein biomarkers of IEC populations by semi-quantitative confocal microscopy, which demonstrated that cells identified by cell type-specific markers express EGFP levels predicted by RNA-seq (Fig. 1E&F). Collectively, our RNA-seq data support the distinct identities of Sox9^{EGFP} populations, define the transcriptional output of each population, and establish the Sox9^{EGFP} model as a viable, single-biomarker approach for isolating ISCs, progenitors, and post-mitotic cells for studies of chromatin dynamics.

OCRs are dynamic across *Sox9* populations

Next, we sought to determine the contribution of chromatin dynamics to the unique transcriptomic identities of Sox9^{EGFP} populations. We mapped OCRs in *Sox9*^{neg}, *Sox9*^{sub}, *Sox9*^{low}, and *Sox9*^{high} by Omni-ATAC-seq (Corces et al., 2017). 81,919 high confidence OCRs were identified across all four populations (Fig. S2A and Supplemental Table 2). We observed a strong correlation between biological replicates within each Sox9^{EGFP} population, and distinct clustering of *Sox9*^{neg}, *Sox9*^{sub}, *Sox9*^{low}, and *Sox9*^{high} by ATAC signal (Fig 2A). We classified peaks found in all four Sox9^{EGFP} populations as "shared" and compared shared peaks to peaks found in single populations relative to genomic features. A majority of peaks present in a single population were found in introns or intergenic regions, while promoters represented a greater

relative proportion of shared peaks (Fig 2B). Site-specific analysis of genome browser tracks confirmed dynamic changes in OCRs, with open chromatin found in loci encoding genes associated with function in Sox9^{EGFP} subpopulations. *Lct*, which is expressed in absorptive enterocytes was enriched for OCRs specific in *Sox9*^{neg}, with a similar correlation between expression and chromatin status observed for *Olfm4* in *Sox9*^{low} and *Neurod1* in *Sox9*^{high} (Fig. 2C).

To quantify chromatin dynamics, we analyzed OCR variability across Sox9^{EGFP} populations using Gini index, a statistical measure of inequality that assigns higher values to more variable distributions (Yoshida et al., 2019). We visualized Gini index by UMAP dimensionality reduction, where each point on the plot represents a single OCR and the Gini index indicates the extent of variable accessibility across all four Sox9^{EGFP} populations (Fig. 2D) (McInnes et al., 2018). As predicted by genome feature analysis of shared and single OCRs, promoters were mostly associated with OCRs bearing low Gini index values, while introns were associated with a greater number of OCRs bearing high Gini index values (Fig. 2D&E). Additionally, promoters were clustered together in the UMAP representation of the data, whereas intronic OCRs were spread throughout the plot, highlighting similarities in accessibility between promoter-associated OCRs. Comparison of Gini values at different genomic features confirmed that the highest degree of dynamic chromatin is found in introns and intergenic regions (Fig. 2E). Despite the overall low diversity of promoters, we identified 738 promoter-associated OCRs with Gini $\geq$ 0.4 (Fig. 2D&E and Fig. S2). These dynamics promoters were found in all four Sox9^{EGFP} populations (n = 56 *Sox9*^{neg}, 97 *Sox9*^{sub}, 216 *Sox9*^{low}, 341 *Sox9*^{high}) and were associated with genes enriched in their respective population (Fig S2C&D).

Next, we visualized OCR intensity and distribution by Sox9^{EGFP} population. OCRs for *Sox9*^{low} ISCs and *Sox9*^{sub} progenitors were distributed across UMAP plots, with overlapping adjacent regions of density (Fig. S2E). *Sox9*^{neg} and *Sox9*^{high} populations demonstrated highest intensity OCRs at opposite ends of the plot, suggesting that the most significant differences in OCRs are found between populations containing post-mitotic IEC subtypes (Fig. S2E). To focus our analysis, we plotted dynamic OCRs by Sox9^{EGFP} population in which accessibility was highest. OCRs clustered by Sox9^{EGFP} population, suggesting that many OCRs shared between 2 or more populations are subject to regulatory activity that reaches a maximum in a single cell type (Fig. 2F). Globally, ATAC-seq in Sox9^{EGFP} populations demonstrates that chromatin in IEC differentiation is dynamic, associated with population-defining genes, and exhibits the most significant global changes in introns and intergenic regions in post-mitotic cells.

Chromatin dynamics are linked to transcription factor networks

We next sought to establish a framework to further examine the relationship between chromatin dynamics and gene expression in Sox9^{EGFP} populations. We reasoned that changes in TF motif usage could be leveraged to distinguish static from dynamic accessibility in IEC chromatin. chromVAR, which quantifies motif-dependent gain or loss of OCRs, identified altered motif accessibility across Sox9^{EGFP} populations (Schep et al., 2017). Ranking TF motifs

by variability in Sox9^{EGFP} populations revealed motifs associated with static chromatin (n = 602 motifs, chromVAR value < 5) as well as highly dynamic chromatin (n = 282 motifs, chromVAR value ≥ 5) (Fig 3A and Supplemental Table 3).

To further explore relationships between TF expression and chromatin dynamics, we analyzed correlation between these two readouts and generated a subset of TFs demonstrating: (1) significant correlation between gene expression and chromatin accessibility over motifs and (2) chromVAR values > 5 (Fig 3B). Increased chromatin accessibility over a given motif was mostly associated with increased expression of the corresponding TF. This category included known regulators of IEC differentiation, which validated the computational approach. *Foxa1*, which drives EEC differentiation, is most highly expressed with the greatest degree of motif accessibility in *Sox9*^{high} cells and *Hnf4g*, which drives absorptive differentiation, is most highly expressed with the greatest degree of motif accessibility in *Sox9*^{neg} cells (Fig. 3C) (Chen et al., 2019; Ye and Kaestner, 2009). We also observed several cases where TF expression and chromatin occupancy were inversely correlated. *Klf7*, *Klf8*, and *Klf12* were observed to by highly expressed but have the lowest level of chromatin occupancy in *Sox9*^{high} cells, consistent with known roles for Klf family TFs as repressors (Fig. 3D and Fig. S3A) (McConnell and Yang, 2010). A similar inverse correlation between TF expression and chromatin occupancy was observed for *Id3*, which is known to function as a transcriptional repressor but has no known role in ISC differentiation (Fig. S3B) (Benezra et al., 1990). In addition to *Id3*, we identified other TFs with no established role in ISC differentiation, including nuclear receptor superfamily genes in *Sox9*^{neg}, and Ets family genes in *Sox9*^{high} (Fig. 3B).

Finally we visualized TF motif usage in OCRs relative to Sox9^{EGFP} population by UMAP. Motif accessibility correlated with populations as predicted by accessibility and TF expression relationships and revealed differential distribution of TF motif usage within Sox9^{EGFP} population clusters (Fig. 3C-E & Fig. 3F, G). *Foxa1* and *Foxa2* demonstrated similar distribution within *Sox9*^{high} associated OCRs, while *Rfx3* and *Rfx6* motifs were also enriched within the *Sox9*^{high} cluster, but with unique distribution (Fig. 3G). These data suggest the existence of multiple *cis*-regulatory landscapes within each population and are supported by differential regulation of EEC subpopulations by *Foxa1/2* (L- and D-cells) versus *Rfx3/6* (K-cells) (Suzuki et al., 2013; Ye and Kaestner, 2009). By analyzing OCRs in Sox9^{EGFP} populations relative to associated TF motifs, our data infer *trans*-regulatory interactions with dynamic chromatin that are confirmed by identification of TFs with known roles in IEC differentiation.

5-hydroxymethylation is dynamic and facilitates absorptive gene expression

We next sought to examine a specific chromatin modification in Sox9^{EGFP} populations to further assay *cis*-regulatory dynamics. 5hmC is associated with activation of gene expression and can be mapped in small numbers of primary cells using Nano-hmC-Seal (Han et al., 2016). *Tet1-3*, which catalyze hydroxymethylation of 5mC, are differentially expressed in Sox9^{EGFP} populations, suggesting the potential for dynamic regulation of 5hmC (Fig. S4A). Nano-hmC-Seal revealed broad, dynamic domains of 5hmC across gene bodies and extending into the

surrounding genome (Fig. 4A). Since genic 5hmC has been linked to increased transcription, we compared fold change in gene expression to fold change of gene body 5hmC in each Sox9^{EGFP} population (Han et al., 2016). We noted a moderate positive correlation between expression and genic 5hmC in all populations, with *Sox9*^{high} and *Sox9*^{neg} cells displaying the strongest relationship (Fig. 4B). When comparing the significantly altered genes for RNA-seq and Nano-hmC-Seal, we found the vast majority were concordantly regulated (gene expression increasing with genic 5hmC) (Fig. 4C).

To determine if 5hmC dynamics correlate with predictable patterns in ISC differentiation, we assessed the enrichment of 5hmC signal over defined marker genes for cell types identified by scRNA-seq (Haber et al., 2017). Consistent with previous reports, this analysis showed that 5hmC is depleted near the TSS, with high levels of the mark found throughout the gene body (Fig. 4D and S4B) (Han et al., 2016). We observed predictable relationships between enrichment of 5hmC in gene bodies associated with specific IEC subpopulations and their corresponding Sox9^{EGFP} population. For example, *Sox9*^{neg} cells were enriched for high 5hmC signal over genes associated with absorptive enterocytes, while *Sox9*^{high} cells demonstrated modest enrichment of 5hmC over genes associated with tuft cells (Fig. 4D and S4B). Most cell type specific marker gene sets were devoid of 5hmC enrichment associated with any one Sox9^{EGFP} population. This suggests that 5hmC may play specific regulatory roles in different IEC populations.

To functionally test the role of 5hmC, we sought to modulate enzymatic activity of TETs in intestinal organoid cultures. Site-specific and non-enzymatic roles for TET1-3 complicate analyses of knockout/knockdown approaches. Since all TET enzymatic activity is sensitive to intracellular succinate/alpha-ketoglutarate ratios, we treated organoids with cell-permeable succinate to inhibit 5hmC, as previously applied in ES cells (Carey et al., 2015). Dot blot analysis revealed a significant reduction in global 5hmC levels in organoids treated with 10mM succinate for 5 days *in vitro* (Fig. S4C-E). To assess lineage allocation, we conducted RT-qPCR for IEC subtype-specific markers in control and succinate-treated organoids. While marker expression was highly variable between biological replicates in both control and succinate conditions, reflecting heterogeneity of organoid cultures, we observed a significant and reproducible downregulation of genes associated with absorptive enterocytes (Fig. 4E). No other IEC subtypes demonstrated significant changes in gene expression, consistent with Nano-hmC-Seal assays demonstrating the greatest degree of 5hmC enrichment over absorptive enterocyte genes (Fig. 4D).

We further explored 5hmC dynamics in IECs by correlating changes in gene expression with changes in 5hmC signal over gene bodies, focusing on genes identified as upregulated in any of the four Sox9^{EGFP} populations (Fig. 1C). By analyzing 5hmC signal relative to gene expression changes, we identified three classes of genes. While a majority of gene expression correlates positively with 5hmC (class I – 69.4%), some genes demonstrate no correlation with 5hmC (class II – 22.3%), and others demonstrate decreasing expression as 5hmC increases (class III – 8.4%) (Fig. S4E). To specifically examine 5hmC dynamics in differentiation, we restricted our analysis of expression:5hmC correlations to *Sox9*^{neg} specific genes identified by our bulk

RNA-seq. *Sox9*^{neg}-enriched genes were further classified by cell-type specific labels from scRNA-seq (Fig. 1C & 4F). Relative to goblet, Paneth, and non-classified genes, absorptive enterocyte genes exhibited the strongest overall correlation with gene body 5hmC (Fig. S4F). We hypothesized that genes directly involved in ISC differentiation should follow a logical progression of increasing gene expression and 5hmC from ISCs (*Sox9*^{low}) to TAs (*Sox9*^{sub}) to mature cells (*Sox9*^{neg}). When we plotted expression:5hmC correlations for *Sox9*^{neg} specific genes, we observed this progression clearly in genes associated with absorptive enterocytes, while goblet and Paneth cell genes had less well-defined relationships between Sox9^{EGFP} population progression (Fig. 4F). Our data demonstrate that 5hmC is dynamic across gene sets associated with all post-mitotic IEC populations. However, the correlation between gene expression and 5hmC is strongest in absorptive enterocytes, which organoid experiments demonstrate to be predictive of the sensitivity of absorptive genes to loss of 5hmC.

DISCUSSION

We mapped chromatin dynamics across IECs using a Sox9^{EGFP} model and identified widespread differences in OCRs and 5hmC signal between populations. Our data provide a multi-omic chromatin "roadmap" of the intestinal epithelium, and a longitudinal analysis of cellular differentiation in an continuously-renewing adult tissue. By combining transcriptome and chromatin analyses, we were able to identify potential *cis*-regulatory modules associated with differentiation of ISCs into mature cell types. Collectively, our data demonstrate that chromatin changes underlie transcriptome dynamics associated with IEC fate specification and suggest that active chromatin regulation may be important for facultative stemness in the setting of epithelial damage and disease.

To address whether or not intestinal chromatin undergoes significant remodeling during ISC differentiation, we faced the conceptual challenge of defining the relative extent of inter-population chromatin differences as "static" or "dynamic". The Gini index was developed for economic dispersion analysis, but was recently applied to quantify chromatin dynamics in the hematopoietic system (Yoshida et al., 2019). Our data affirm the Gini index as a valuable approach for quantifying chromatin dynamics in studies involving 3 or more groups, where pairwise analyses may be insufficient. As a complementary approach, combined analysis of OCR activity, TF motifs found within OCRs, and corresponding TF expression allowed us to identify potential *cis*-regulatory features associated with specific *trans*-regulatory modules. Importantly, it allowed us to classify intestinal chromatin as "static" or "dynamic" in a TF-associated manner and predict co-regulatory relationships for TF networks across stages of ISC differentiation. A majority of dynamic OCRs coinciding with ISC differentiation were found in intra- or intergenic regions, consistent with positioning of enhancer elements and reports of chromatin dynamics in other tissue systems (Yoshida et al., 2019). Functional association between specific enhancers and their associated genes would help further define the significance of specific *cis*-regulatory elements.

Since OCRs are subject to regulation by multiple factors, we examined 5hmC dynamics to understand a specific chromatin modification in ISC differentiation. While previous studies have identified changes in 5hmC between ISCs and villus epithelium, we were able to apply the $Sox9^{EGFP}$ model to include intermediate progenitors in our analysis of 5hmC dynamics (Kim et al., 2016). Doing so allowed us to plot the trajectory of 5hmC in subpopulations of epithelial cells, by combining $Sox9^{EGFP}$ population identities established here with gene sets identified by published scRNA-seq. This revealed a uniquely strong association between 5hmC and absorptive enterocyte differentiation that we confirmed in organoid assays, establishing a framework for functional validation of observed chromatin dynamics.

Large-scale chromatin mapping projects, including the Epigenome Roadmap, are critical for expanding our understanding of the functional genome, but often focus on whole tissues at the expense of cell-level resolution (Roadmap Epigenomics et al., 2015). Next generation studies require higher resolution approaches to avoid population averaging effects and fully understand the regulatory significance of chromatin landscapes. Our study provides a valuable cell type-specific chromatin "roadmap" in a continuously renewing adult tissue that will facilitate further mechanistic dissection of the chromatin modifying enzymes and TF networks involved in establishing functional genomes in IECs. Further refinement of cell-type specific isolation protocols and single cell assays are likely to reveal an even more dynamic chromatin landscape in ISC differentiation.

FIGURE LEGENDS

Figure 1: *Sox9^{EGFP} populations capture the spectrum of IEC differentiation*. (A) Sox9^{EGFP} is expressed at distinct levels in the crypt and as rare "high"-expressing cells in the villi. (B) Sox9^{EGFP} levels facilitate quantification and isolation by FACS and define (C) distinct cell populations bearing unique transcriptomic signatures. (D) Comparison to cell type specific gene expression signatures from scRNA-seq experiments demonstrates relative enrichment of differentiated cell types in *Sox9*^{neg}, progenitors in *Sox9*^{sub}, ISCs in *Sox9*^{low}, and EECs and tuft cells in *Sox9*^{high}. (E) Co-localization between cell type specific protein markers and Sox9^{EGFP} by immunofluorescence and (F) quantification of EGFP signal confirms these relationships (* indicates $p < 0.03$; ** indicates $p < 0.0001$).

Figure 2: *IEC populations have unique and dynamic OCR signatures.* (A) Sox9^{EGFP} populations cluster reproducibly by Pearson correlation of OCRs (n = 3 biological replicates per population). (B) Intergenic regions and introns represent the majority of OCRs present in single populations, while promoters represent a larger portion of OCRs shared across all four populations. (C) Representative browser tracks demonstrate relationships between OCR activity and gene expression. (D) UMAP dimensionality reduction and Gini index analysis of OCRs reveals static and dynamic chromatin and predictable clustering by genome feature, including static promoters and dynamic introns and intergenic loci (n = 81,919 OCRs). (E) Quantitative ranking of genome features by Gini index confirms that the highest degree of chromatin dynamics are associated with introns and intergenic regions, but demonstrates dynamic loci in all features, including promoters. (F) UMAP visualization of OCRs by highest activity demonstrates distinct clustering by Sox9^{EGFP} population.

Figure 3: *Identification of regulatory networks by OCR dynamics*. (A) Analysis of motif usage in Sox9^{EGFP} populations identifies TFs associated with dynamic (n = 282, chromVAR $\geq$ 5) and static (602 motifs, chromVAR value < 5) OCRs. (B) A subset of 64 TFs demonstrate significant correlation with dynamic motif usage. (C) A majority of gene expression and motif relationships are positively correlated, including *Foxa1* and *Hnf4g*. (D) Negative correlations are observed for transcriptional repressors including *Klf12*. (E) TFs lacking correlation between motif accessibility and expression, like *Stat3*, are associated with static chromatin. (F) Mapping motifs onto UMAP plots of Sox9^{EGFP} population specific OCRs reveals distributions predicted by gene expression/motif accessibility relationships (n loci = 12,117 *Hnf4g*; 7,877 *Klf12*). (G) Motifs for *Foxa1/2* and *Rfx3/6* are differentially distributed within *Sox9*^{high} OCRs, suggesting regulatory roles in subpopulations of the same parent Sox9^{EGFP} population (n loci = 7,472 *Foxa1*; 3,302 *Foxa2*; 3,680 *Rfx3*; 4,862 *Rfx6*).

Figure 4: *5hmC is dynamic across IECs and is required for absorptive gene expression*. (A) 5hmC exhibits broad signal over gene bodies that is dynamic across Sox9^{EGFP} populations and

associated with changes in gene expression. (B) Increased 5hmC in gene bodies demonstrates a moderate positive correlation with gene expression in all Sox9^{EGFP} populations, with stronger relative relationships found in *Sox9*^{neg} and *Sox9*^{high}. (C) A majority of genes demonstrate concordant relationships between 5hmC and expression. (D) Enrichment of 5hmC over scRNA-seq defined IEC gene sets reveals increased signal in absorptive enterocyte genes in the *Sox9*^{neg} population and tuft cell genes in the *Sox9*^{high} population. (E) Organoids treated with dimethyl succinate to suppress global 5hmC levels demonstrate downregulation of absorptive enterocyte specific genes, with no change to canonical biomarkers associated with secretory lineages (n = 3 0mM and 10mM dimethyl succinate; * indicates $p < 0.05$, ** indicates $p < 0.005$). (F) 5hmC:RNA correlations demonstrate a positive relationship in gene sets associated with cell types found in *Sox9*^{neg} population, with absorptive enterocytes exhibiting a large dynamic range in 5hmC and gene expression increases that proceed in an ISC (*Sox9*^{low}) to TA (*Sox9*^{sub}) to differentiated (*Sox9*^{neg}) manner.

Supplemental Figure 1: *Transcriptomic characterization of Sox9^{EGFP} populations*. (A) Representative gating for FACS isolation of Sox9^{EGFP} populations. (B) RNA-seq demonstrates accurate enrichment of endogenous *Sox9* mRNA across EGFP populations. (C) Bulk RNA-seq of Sox9^{EGFP} populations correlates with cell-type specific gene signatures identified by (Yan et al., 2017). (D) GO analysis of Sox9^{EGFP} population transcriptomic signatures reveals enrichment of cellular functions consistent with predicted cell type.

Supplemental Figure 2: *OCRs and promoter dynamics in IEC differentiation.* (A) Summary of OCRs by Sox9^{EGFP} population demonstrates differential numbers of shared and unique peaks. (B) Omni-ATAC-seq signal is highly reproducible between biological replicates and can be quantified by the Gini index to identify relative similarities and differences in OCRs between populations. (C) UMAP visualization of promoters reveals that a majority have low Gini index values indicative of high similarity between Sox9^{EGFP} populations, but also identifies dynamic promoter OCRs (n = 56 *Sox9*^{neg}, 97 *Sox9*^{sub}, 216 *Sox9*^{low}, 341 *Sox9*^{high}). (D) Correlation between accessibility and corresponding gene expression for dynamic promoters. (E) Sox9^{EGFP} population-specific OCR intensity reveals clustering by population on UMAP dimensionality reduction.

Supplemental Figure 3: *TF expression and motif accessibility identifies repressive relationships*. (A) *Klf8* demonstrates an inverse relationship between expression and motif accessibility predictive of *Sox9*^{high} specific repressive function (n = 11,203 loci). (B) *Id3* is expressed and associated with decreased motif accessibility in all populations but *Sox9*^{low} ISCs, suggesting a role in IEC differentiation (n = 5,664 loci).

Supplemental Figure 4: *Regulation of 5-hydroxymethylcytosine in IECs.* (A) *Tet1-3* are differential expressed in Sox9^{EGFP} populations. (B) 5hmC signal over scRNA-seq gene

signatures from (Yan et al., 2017) independently confirms enrichment over absorptive enterocyte genes. (C) Representative 5hmC dot blot and loading control images from dimethyl succinate-treated intestinal organoids, two independent biological replicates. (D) Quantification of dot blot signal demonstrates significant decrease in global 5hmC with 10mM succinate treatment (n = 4 biological replicates per dose, samples normalized to their matched 0mM control, * indicates significance, p = 0.0093). (E) Ratios of gene expression vs genic 5hmC allow classification of genes based on 5hmC dynamics across $Sox9^{EGFP}$ populations and reveal positive and negative relationships between 5hmC and gene expression. (F) Among cell types associated with $Sox9^{neg}$ population, absorptive enterocytes genes demonstrate the highest expression:5hmC ratios relative to goblet and Paneth cell genes and genes not classified by scRNA-seq (n genes = 147 absorptive enterocyte, 57 goblet, 24 Paneth, 2,464 not classified).

Supplemental Table 1: RNA-seq differential expression analysis

Supplemental Table 2: Omni-ATAC peaks

Supplemental Table 3: chromVAR analysis of Omni-ATAC

Supplemental Table 4: Primary antibodies

METHODS

Mice

Sox9^{EGFP} mice were maintained as heterozygotes on the C57Bl/6 background. All experiments were carried out on adult males between 8-16 weeks of age. Mice were co-housed, littermates were used when possible, and age-matched non-littermates when not. All mice received Tekland Global Soy Protein-Free Extruded Rodent Diet (Envigo, 2020SX) and water *ad libitum*. Phenotyping for *Sox9*^{EGFP} was carried out by observation of EGFP expression in hair follicles of tail clippings, as previously described (Formeister et al., 2009). The Institutional Animal Care and Use Committee (IACUC) of the University of North Carolina at Chapel Hill reviewed and approved all protocols.

Intestinal epithelial dissociation and FACS

Full length intestine (duodenum, jejunum, and ileum) was used for all sequencing-based experiments, and jejunum was used for organoid experiments. Intestinal epithelial dissociations were carried out as previously described (Zwarycz et al., 2018). For full length preps, the entire small intestine was dissected and opened longitudinally before being rinsed in 1X DPBS (Gibco) to remove fecal matter. For jejunal preps, the proximal 5cm of intestine was removed and discarded, and the proximal half of the remaining intestine was opened longitudinally and rinsed in 1X DPBS to remove fecal matter. Rinsed intestine was transferred to a 50mL conical containing 10mL of 3mM EDTA (Corning) and incubated for 15min at 4C with gentle nutation (RPM = 40). Intestinal tissue was retrieved and transferred to a glass plate prior to being "brushed" with a pipette tip to remove ~50% of villi. The tissue was then minced to ~3-5mM pieces and transferred to a new 50mL conical containing 10mL of 3mM EDTA and incubated for and additional 35min at 4C with gentle nutation (RPM = 40). Tissue fragments were transferred to 10mL DPBS in a third 50mL conical and shaken by hand for 4min to remove crypt-enriched epithelium. To obtain single cells, crypt epithelium was rinsed once with 25mL DPBS then resuspended in 10mL HBSS with 0.3U/mL dispase (Corning) and 10uM Y27632 (Selleck Chem), and incubated at 37C for 12-16min with gentle shaking and observation by light microscopy every 2min until ~80% of prep consisted of single cells. Cells were rinsed twice with DPBS containing 10% FetalPlex (Gemini Biosciences) to quench dispase activity.

Dissociated single cells were resuspended in 1X Sort Media [Advanced DMEM/F12 (Gibco), N2 (Gibco), B27 (Gibco), 10mM HEPES (Corning), Glutamax (Gibco), Pen/Strep (Gibco), and 10uM Y27632 (Selleck Chem)] and stained with CD31-APC (Biolegend) and CD45-APC (Biolegend) for 45min on ice. Stained cells were rinsed once with Advanced DMEM/F12 and resuspended in 1X Sort Media. 7-AAD (Biolegend) and Annexin V-APC (Biolegend) were added immediately prior to sorting. Cells were sorted on a Sony SH800 FACS instrument with a 100um disposable nozzle "chip". Gating was carried out as previously described in (Gracz et al., 2018) and shown in Supplemental Figure 1A. Doublet discrimination

was visually validated by sorting ~500 cells directly on to a glass slide and observing by light microscopy.

Organoid culture and dimethyl succinate treatment

Organoids were cultured from jejunal crypts in Cultrex Type II Growth Factor-Reduced matrix (R&D Biosystems). For 5hmC dot blot experiments, ~300 crypts were seeded in 30uL Cultrex droplets in 48 well plates and overlaid with 200uL organoid media. For gene expression experiments, ~100 crypts were seeded in 10uL Cultrex droplets in 96 well plates and overlaid with 100uL organoid media. Media consisted of: Advanced DMEM/F12 (Gibco), N2 (Gibco), B27 (Gibco), 10mM HEPES (Corning), Glutamax (Gibco), Pen/Strep (Gibco), 50ng/mL recombinant murine EGF (Gibco), and 100ug/mL recombinant human Noggin (Peprotech). 10uM Y27632 (Selleck Chem) was added at initial plating. Organoids were allowed to establish for 48hr prior to treatment with dimethyl succinate (Sigma). Media and succinate were replaced every 48hr.

RNA isolation

12,000 cells from each Sox9^{EGFP} population were sorted directly into 500uL of Lysis Buffer (Ambion) and stored at -80C until RNA isolation. Total RNA was prepared using the RNAqueous Micro Kit (Ambion) according to manufacturer instructions and eluted in 15uL molecular-grade H2O (Corning). RNA was treated with Ambion RNAqueous Micro Kit DNase I for 30min at 37C to eliminate gDNA and DNase was inactivated using DNase Inactivation Beads. gDNA-free RNA was transferred to low-binding tubes and stored at -80C until QC and library preparation.

gDNA isolation for Nano-hmC-Seal-seq and dot blot

For FACS isolated Sox9^{EGFP} populations, cells were collected directly into 250uL 1X sort media, pelleted by centrifugation at 6,000g for 5min, then lysed in 200uL gDNA Lysis Buffer (10mM Tris HCl pH 8.0, 100mM EDTA pH 8.0, 0.5% SDS) with 1mg/mL Proteinase K (Ambion) and incubated at 55C in a water bath overnight. For organoids, growth media was isolated from wells and Cultrex and organoids were lysed in 500uL gDNA lysis buffer with 1mg/mL Proteinase K (Ambion) and incubated at 55C in a water bath overnight. gDNA from both FACS isolated cells and organoids was purified using a Quick DNA MicroPrep kit (Zymo) and assayed for concentration using the Qubit dsDNA High Sensitivity kit (Thermo Fisher) and a Qubit Fluorometer (Thermo Fisher).

RNA-seq: QC, library prep, and sequencing

RNA from 12,000 cells per Sox9^{EGFP} population was collected for each RNA-seq sample. 5uL of 15uL total from each sample was allocated for RT-qPCR validation. cDNA was prepared as described in "RT-qPCR" and sequential enrichment of *Sox9* mRNA across EGFP-sorted populations was validated by RT-qPCR with Taqman probes against *Sox9* (Life Technologies)

using *18S* as an internal control. To validate RNA quality, 1uL of 15uL total from each sample was subjected to Agilent Bioanalyzer analysis using the Total Eukaryotic RNA Pico kit (Agilent). All samples demonstrated significant enrichment of *Sox9* across EGFP population and had RIN values ≥ 8.0. Libraries were prepared using the SMARTer Stranded Total RNA-seq Kit v2 (Clontech) and 8uL of 15uL total RNA per sample, as per manufacturer instructions. Libraries were sequenced on a NextSeq500 (Illumina) with 75bp single-end reads, v2 chemistry. Each library was sequenced to ≥ 2.9 x 10^7 reads.

Omni-ATAC-seq: library prep and sequencing

12,000 cells per $Sox9^{EGFP}$ population were collected directly into 250uL 1X sort media, pelleted by centrifugation at 2,000g for 5min, and subjected to Omni-ATAC as previously described in (Corces et al., 2017). Libraries were sequenced on a NextSeq500 (Illumina) with 75bp paired-end reads, v2 chemistry. Each library was sequenced to ≥ 5.0 x 10^7 reads.

Nano-hmC-Seal-seq: library prep and sequencing

12,000 cells per $Sox9^{EGFP}$ population were collected via FACS (above). gDNA was isolated as described above (see: *gDNA isolation for Nano-hmC-Seal-seq and dot blot*) and subjected to Nano-hmC-Seal as previously described (Han et al., 2016). Libraries were sequenced on a NextSeq500 (Illumina) with 75bp single-end reads, v2 chemistry. Each library was sequenced to ≥ 2.2 x 10^7 reads.

RNA-seq: analysis

Quantification was performed using salmon (version 0.8.2) (Patro et al., 2017) and differential expression testing was performed using DESeq2 (version 1.2.22) with tximport (1.10.1) to summarize transcript data at the gene level in R. Expression values in the Sox9-defined populations were then compared to single cell data from two independent sources (Haber et al., 2017; Yan et al., 2017).

Omni-ATAC-seq: analysis

Paired-end 76 bp reads were trimmed using trim_galore (version 0.4.3), aligned using bowtie2 (version 2.3.4) with -X 2000. Then duplicates were removed using picard (version 2.10.3) and peaks were called with Macs2 (version 2.1.2) –q 0.001 –f BAMPE –keep-dup-all . For visualization, signal tracks were computed using Deeptools (version 2.5.2) bamCoverage –binSize 1 –extendReads –normalizeUsingRPKM –ignoreForNormalization chrM chrX chrY –minFragmentLength 40 –maxFragmentLength 130 . Comparison of replicates was performed using bamCompare in the deeptools package. Normalized Tn5 insertions were calculated using pyatac (Schep et al., 2015).

To identify differences in motif usage the R package chromVAR was used (Schep et al., 2017). To compare differences in accessibility at individual sites, UMAP dimensionality reduction was used (McInnes et al., 2018) (R package version 0.2.1.0. https://CRAN.R-

project.org/package=umap) and the gini index for each site across the 4 population (3 replicates each) was calculated using the gini function in the edgeR package. Motif counts were identified using the motifmatcher R package, using the default p-value cutoff of 5e-5 and projected into UMAP space.

Nano-hmC-Seal-seq: analysis

5hmc data were trimmed using trim_galore and aligned to mm9 using bowtie2 (version 2.3.4) –sensitive. Follwing duplicate removal using picard (version 2.10.3), signal tracks were generated using deeptools (version 2.5.2) bamCoverage –binSize 10 –extendReads 150 –normalizeUsingRPKM. Signal over genic regions was calculated using deeptools computeMatrix function and gencode annotations (vM1). Pearson correlation between expression and 5hmc levels was calculated in R using the cor.test function and statistical inferences were adjusted for multiple testing using the p.adjust function.

5hmC dot blot

gDNA prepared from control and dimethyl succinate-treated organoids was diluted to 400ng per sample in 20uL TE. 40uL 0.1M sodium hydroxide was added and samples were incubated at 95C for 10min in thermal cycler to denature DNA. 60uL ice cold 1M ammonium acetate was added and serial dilutions were made with 60uL sample and 60uL TE. Denatured gDNA was spotted onto Hybond N+ membranes (GE Amersham) using a BioDot SF Microfiltration apparatus (BioRad). Membranes were dried for 5min at 80C, UV crosslinked, and blocked overnight in 5% non-fat dry milk in PBS. Membranes were incubated in anti-5-hydroxymethylcytidine (Active Motif, mouse monoclonal, 59.1, 0.4ug/mL) for 3hr at RT in 5% non-fat dry milk in PBS, then rinsed three times in PBS. Secondary detection was carried out by incubating membranes in Gt-anti-Ms HRP (Jackson Immunoresearch, 1:5000) for 3hr at RT in 5% non-fat dry milk in PBS. Membranes were rinsed three times in PBS, HRP was detected using Clarity ECL (BioRad), and blots were imaged using a Protein Simple gel box/imaging station. Following 5hmC detection, blots were rinsed three times in PBS and incubated for 10min at RT in 0.02% methylene blue (Sigma) in 0.3M sodium acetate, pH 5.2 to detect total gDNA. Methylene blue-stained blots were rinsed 5 times in molecular grade H2O and imaged using a Protein Simple gel box/imaging station. Signal intensity for 5hmC was measured for 100ng dots in ImageJ and changes in signal for succinate-treated organoids was calculated as percent change relative to the untreated control for each biological replicate (n = 4 biological replicates).

RT-qPCR

cDNA was prepared using iScript cDNA Synthesis Kit (BioRad) and diluted 1:10 with molecular grade H2O (Corning) prior to use in RT-qPCR assays with SsoAdvanced Universal Master Mix (BioRad). Taqman primers (Life Tech) were used to detect: *Chga* (Mm00514341_m1), *Cox1* (Mm04225243_g1), *Defa-rs1* (Mm00655850_m1), *Dclk1* (Mm00444950_m1), *Lct* (Mm01285112_m1), *Lyz2* (Mm00727183_s1), *Muc2* (Mm00545872_m1), *Sis*

(Mm01210305_m1), *Sox9* (Mm00448840_m1), *Tac1* (Mm00436880_m1), *Tff3* (Mm00495590_m1). Fold change was calculated using $\Delta\Delta C_T$ with *18S* (Hs99999901_s1) as the internal reference (Pfaffl, 2001).

Immunofluorescence and image quantification

Sox9^{EGFP} jejunum was dissected, fixed overnight in 4% PFA, then transferred to 30% sucrose for an additional 16-18 hours prior to embedding in Tissue-Tek OCT (Sakura) and frozen at -80C. Immunofluorescence for IEC lineage markers was carried out on 10um sections. Tissue was permeabilized with 0.3% Triton-X 100 (Sigma) in PBS for 20 minutes at room temperature, followed by blocking in 1x Animal-Free Block (Cell Signaling Technologies) for 40 minutes. Primary antibodies (Supplemental Table 4) were diluted at specified concentrations in 1x Animal-Free Block and incubated for 2hr at RT. Secondary antibodies (Supplemental Table 4) were diluted 1:500 in 1x Animal-Free Block and incubated for 45min at RT. Nuclear counterstain was carried out by bisbenzimide (Sigma).

Images were taken as 1um optical sections using a Zeiss LSM 700 confocal microscope. EGFP fluorescence intensity was measured in ImageJ. Briefly, CD326 was used as an epithelial-specific membrane marker and cells positive for lineage specific markers were outlined using the freehand drawing tool for pixel intensity measurement. Background signal was measured and subtracted from the raw value of the measured cells from the same image. Enteroendocrine cells were marked by CHGA (n = 36), tuft cells were marked by DCLK1 (n = 33), absorptive enterocytes were marked by FABP1 (n=97), Paneth cells were marked by LYZ (n = 78), goblet cells were marked by MUC2 (n = 62), intestinal stem cells (ISCs) were marked by OLFM4 (n = 78), and transit-amplifying progenitor cells were considered to be in the +4 position and not marked by OLFM4 (n = 78). Statistical significance was assessed by one-way ANOVA with post-hoc comparisons by Tukey's test in Prism 8 (GraphPad).

Acknowledgements

The authors would like to thank Brian Golitz and Noah Sciaky for technical support with sequencing and deconvolution of sequencing files. Qiancheng You and Chuan He of the University of Chicago for technical input and training on Nano-hmC-Seal. Pablo Arial and the Microscopy Services Laboratory in the Department of Pathology and Laboratory Medicine at UNC Chapel Hill for assistance with confocal microscopy. Bailey Zwarycz and Othmane Jadi for helpful conversations and critical reading of the manuscript. Terry Magnuson and members of the Magnuson lab at UNC Chapel Hill for helpful conversations and feedback.

Funding

This work was funded by the National Institutes of Health (K01 DK111709 to ADG; R01 DK091427 to STM), the American Gastroenterological Association (Research Scholar Award to ADG), the Center for Gastrointestinal Biology and Disease (P30 DK34987 to ADG), and the Burroughs Wellcome Foundation (Collaborative Research Travel Grant to ADG). The Microscopy Services Laboratory, Department of Pathology and Laboratory Medicine, is

supported in part by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center.

Data availability

RNA-seq, Omni-ATAC-Seq, and Nano-hmC-Seal data are available in the Gene Expression Omnibus under accession number: GSE131442.

REFERENCES

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., *et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.

Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49-59.

Buczacki, S.J., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. Nature *495*, 65-69.

Carey, B.W., Finley, L.W., Cross, J.R., Allis, C.D., and Thompson, C.B. (2015). Intracellular alpha-ketoglutarate maintains the pluripotency of embryonic stem cells. Nature *518*, 413-416.

Chen, L., Toke, N.H., Luo, S., Vasoya, R.P., Fullem, R.L., Parthasarathy, A., Perekatt, A.O., and Verzi, M.P. (2019). A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. Nat Genet.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., *et al.* (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods *14*, 959-962.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., *et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

Formeister, E.J., Sionas, A.L., Lorance, D.K., Barkley, C.L., Lee, G.H., and Magness, S.T. (2009). Distinct SOX9 levels differentially mark stem/progenitor populations and enteroendocrine cells of the small intestine epithelium. American journal of physiology Gastrointestinal and liver physiology *296*, G1108-1118.

Gracz, A.D., Ramalingam, S., and Magness, S.T. (2010). Sox9 expression marks a subset of CD24-expressing small intestine epithelial stem cells that form organoids in vitro. American journal of physiology Gastrointestinal and liver physiology *298*, G590-600.

Gracz, A.D., Samsa, L.A., Fordham, M.J., Trotier, D.C., Zwarycz, B., Lo, Y.H., Bao, K., Starmer, J., Raab, J.R., Shroyer, N.F., *et al.* (2018). SOX4 Promotes ATOH1-independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates. Gastroenterology.

Gracz, A.D., Williamson, I.A., Roche, K.C., Johnston, M.J., Wang, F., Wang, Y., Attayek, P.J., Balowski, J., Liu, X.F., Laurenza, R.J., *et al.* (2015). A high-throughput platform for stem cell niche co-cultures and downstream gene expression analysis. Nature cell biology *17*, 340-349.

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.

Han, D., Lu, X., Shih, A.H., Nie, J., You, Q., Xu, M.M., Melnick, A.M., Levine, R.L., and He, C. (2016). A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. Mol Cell *63*, 711-719.

Kaaij, L.T., van de Wetering, M., Fang, F., Decato, B., Molaro, A., van de Werken, H.J., van Es, J.H., Schuijers, J., de Wit, E., de Laat, W., *et al.* (2013). DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. Genome Biol *14*, R50.

Kim, R., Sheaffer, K.L., Choi, I., Won, K.J., and Kaestner, K.H. (2016). Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation. Genes Dev *30*, 2433-2442.

Kim, T.H., Li, F., Ferreiro-Neira, I., Ho, L.L., Luyten, A., Nalapareddy, K., Long, H., Verzi, M., and Shivdasani, R.A. (2014). Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. Nature *506*, 511-515.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., *et al.* (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943-949.

Lien, W.H., Guo, X., Polak, L., Lawton, L.N., Young, R.A., Zheng, D., and Fuchs, E. (2011). Genome-wide maps of histone modifications unwind in vivo chromatin states of the hair follicle lineage. Cell Stem Cell *9*, 219-232.

McConnell, B.B., and Yang, V.W. (2010). Mammalian Kruppel-Like Factors in Health and Diseases. Physiol Rev *90*, 1337-1381.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods *14*, 417-419.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res *29*, e45.

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al.* (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Roche, K.C., Gracz, A.D., Liu, X.F., Newton, V., Akiyama, H., and Magness, S.T. (2015). SOX9 Maintains Reserve Stem Cells and Preserves Radio-resistance in Mouse Small Intestine. Gastroenterology.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res *25*, 1757-1770.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods *14*, 975-978.

Schmitt, M., Schewe, M., Sacchetti, A., Feijtel, D., van de Geer, W.S., Teeuwssen, M., Sleddens, H.F., Joosten, R., van Royen, M.E., van de Werken, H.J.G.*, et al.* (2018). Paneth Cells Respond to Inflammation and Contribute to Tissue Regeneration by Acquiring Stem-like Features through SCF/c-Kit Signaling. Cell Rep *24*, 2312-2328 e2317.

Suzuki, K., Harada, N., Yamane, S., Nakamura, Y., Sasaki, K., Nasteska, D., Joo, E., Shibue, K., Harada, T., Hamasaki, A.*, et al.* (2013). Transcriptional regulatory factor X6 (Rfx6) increases gastric inhibitory polypeptide (GIP) expression in enteroendocrine K-cells and is involved in GIP hypersecretion in high fat diet-induced obesity. J Biol Chem *288*, 1929-1938.

Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, J.H.*, et al.* (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell *18*, 203-213.

Yan, K.S., Janda, C.Y., Chang, J., Zheng, G.X.Y., Larkin, K.A., Luca, V.C., Chia, L.A., Mah, A.T., Han, A., Terry, J.M.*, et al.* (2017). Non-equivalence of Wnt and R-spondin ligands during Lgr5(+) intestinal stem-cell self-renewal. Nature *545*, 238-242.

Ye, D.Z., and Kaestner, K.H. (2009). Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice. Gastroenterology *137*, 2052-2062.

Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A., Maier, B., Wroblewska, A., Desland, F., Chudnovskiy, A., Mortha, A., Dominguez, C.*, et al.* (2019). The cis-Regulatory Atlas of the Mouse Immune System. Cell *176*, 897-912 e820.

Zwarycz, B., Gracz, A.D., and Magness, S.T. (2018). Organoid Cultures for Assessing Intestinal Epithelial Differentiation and Function in Response to Type-2 Inflammation. Methods in molecular biology *1799*, 397-417.

Figure 1

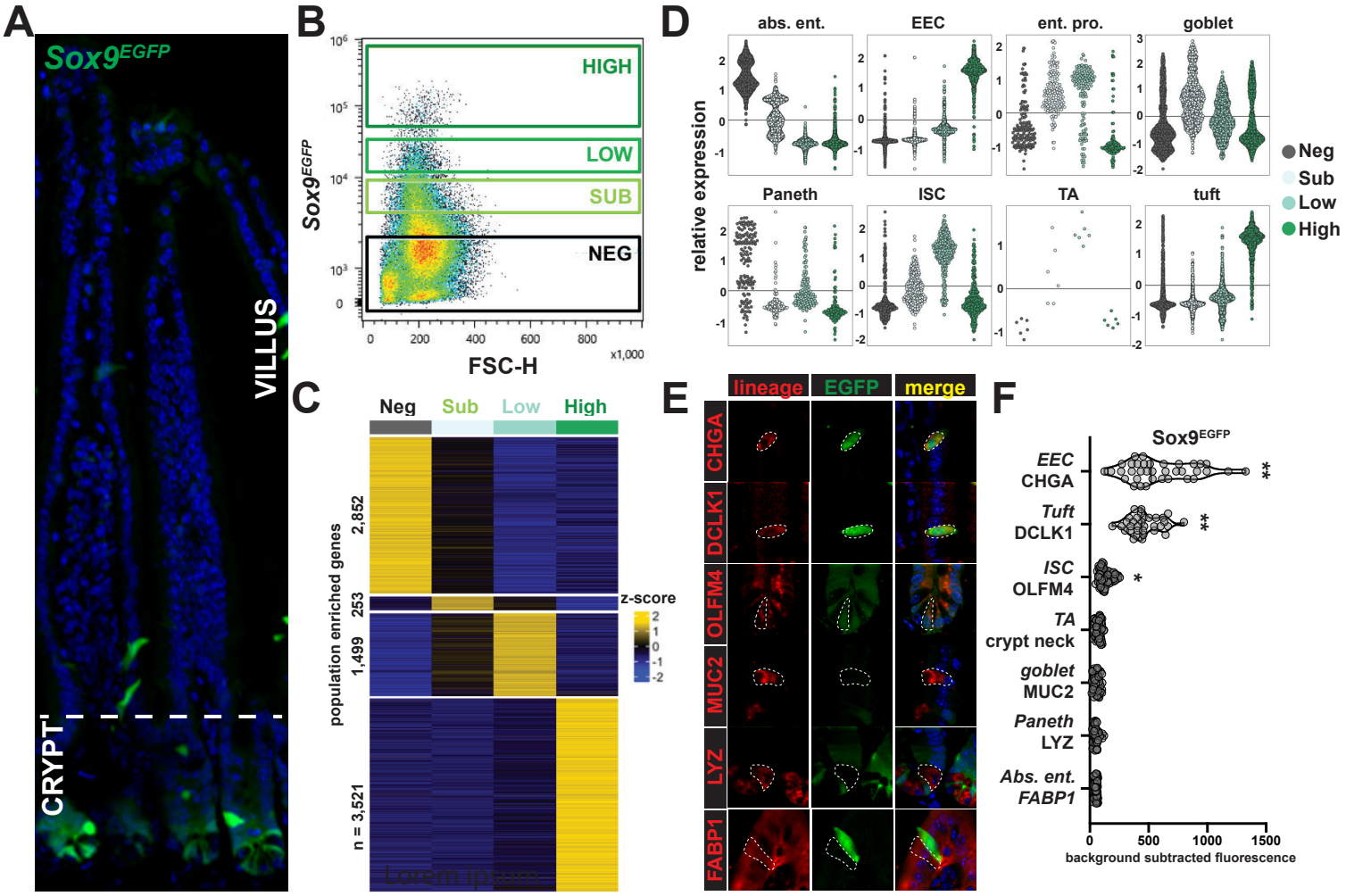

A
Sox9EGFP
VILLUS
CRYPT
B
HIGH
LOW
SUB
NEG
Sox9EGFP
FSC-H
x1,000
C
Neg
Sub
Low
High
population enriched genes
2,852
253
1,499
n = 3,521
z-score
Lorem ipsum
D
abs. ent.
EEC
ent. pro.
goblet
Paneth
ISC
TA
tuft
relative expression
Neg
Sub
Low
High
E
lineage
EGFP
merge
CHGA
DCLK1
OLFM4
MUC2
LYZ
FABP1
F
Sox9EGFP
EEC
CHGA
Tuft
DCLK1
ISC
OLFM4
TA
crypt neck
goblet
MUC2
Paneth
LYZ
Abs. ent.
FABP1
background subtracted fluorescence

Figure 2

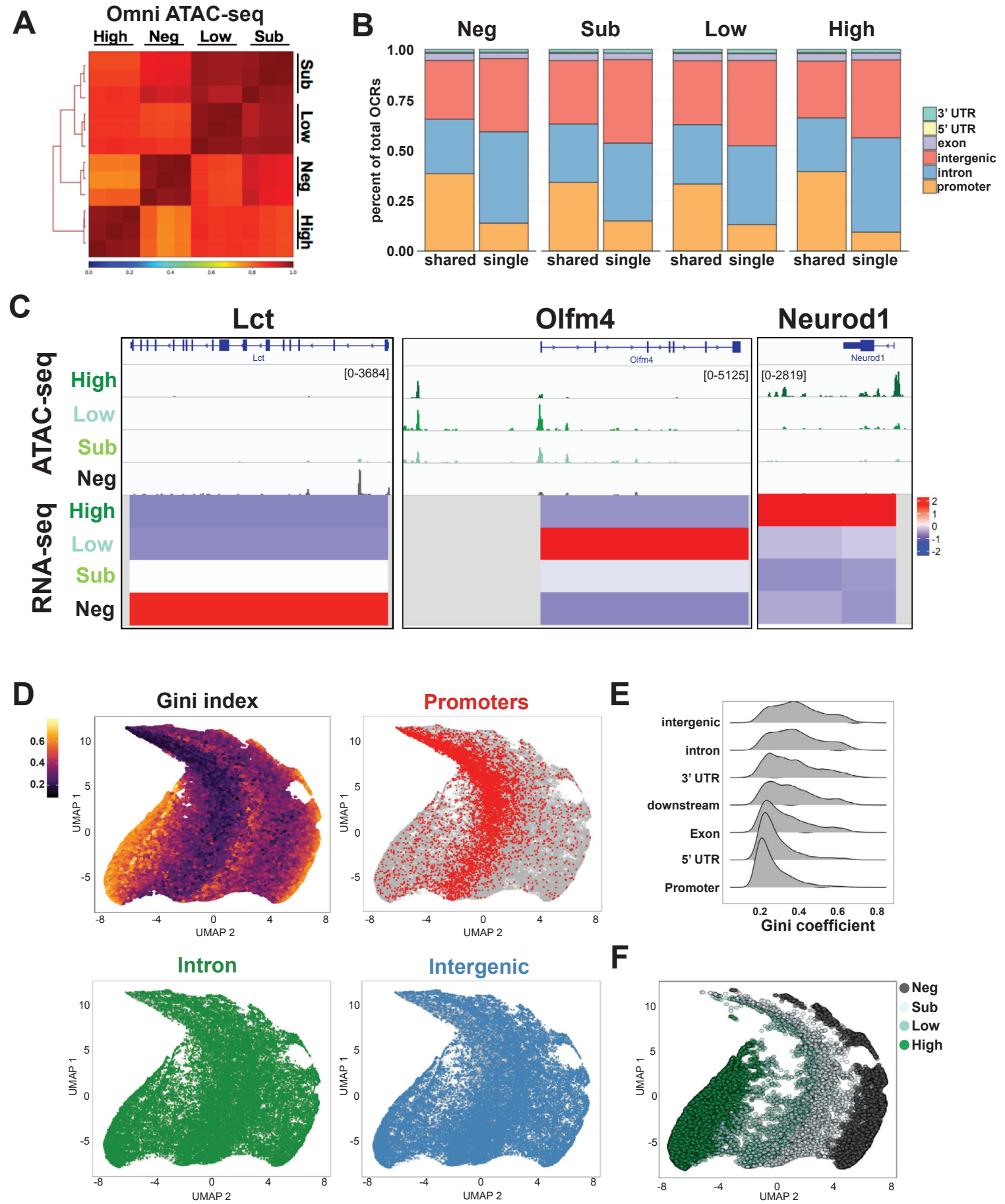

A
Omni ATAC-seq
High
Neg
Low
Sub
B
Neg
Sub
Low
High
percent of total OCRs
1.00
0.75
0.50
0.25
0.00
shared
single
3' UTR
5' UTR
exon
intergenic
intron
promoter
C
Lct
Olfm4
Neurod1
ATAC-seq
RNA-seq
[0-3684]
[0-5125]
[0-2819]
D
Gini index
Promoters
Intron
Intergenic
UMAP 1
UMAP 2
E
intergenic
intron
3' UTR
downstream
Exon
5' UTR
Promoter
Gini coefficient
F
Neg
Sub
Low
High

Figure 3

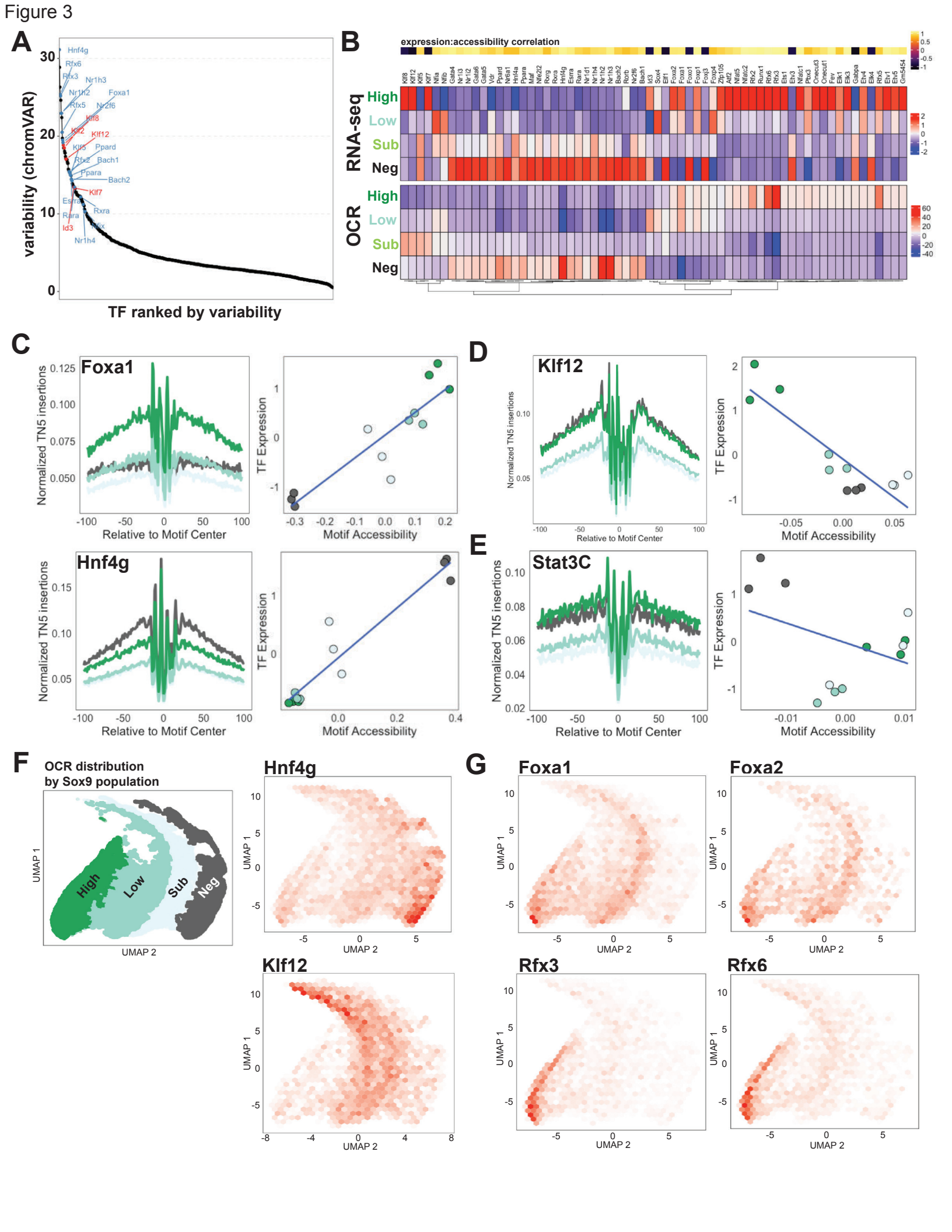
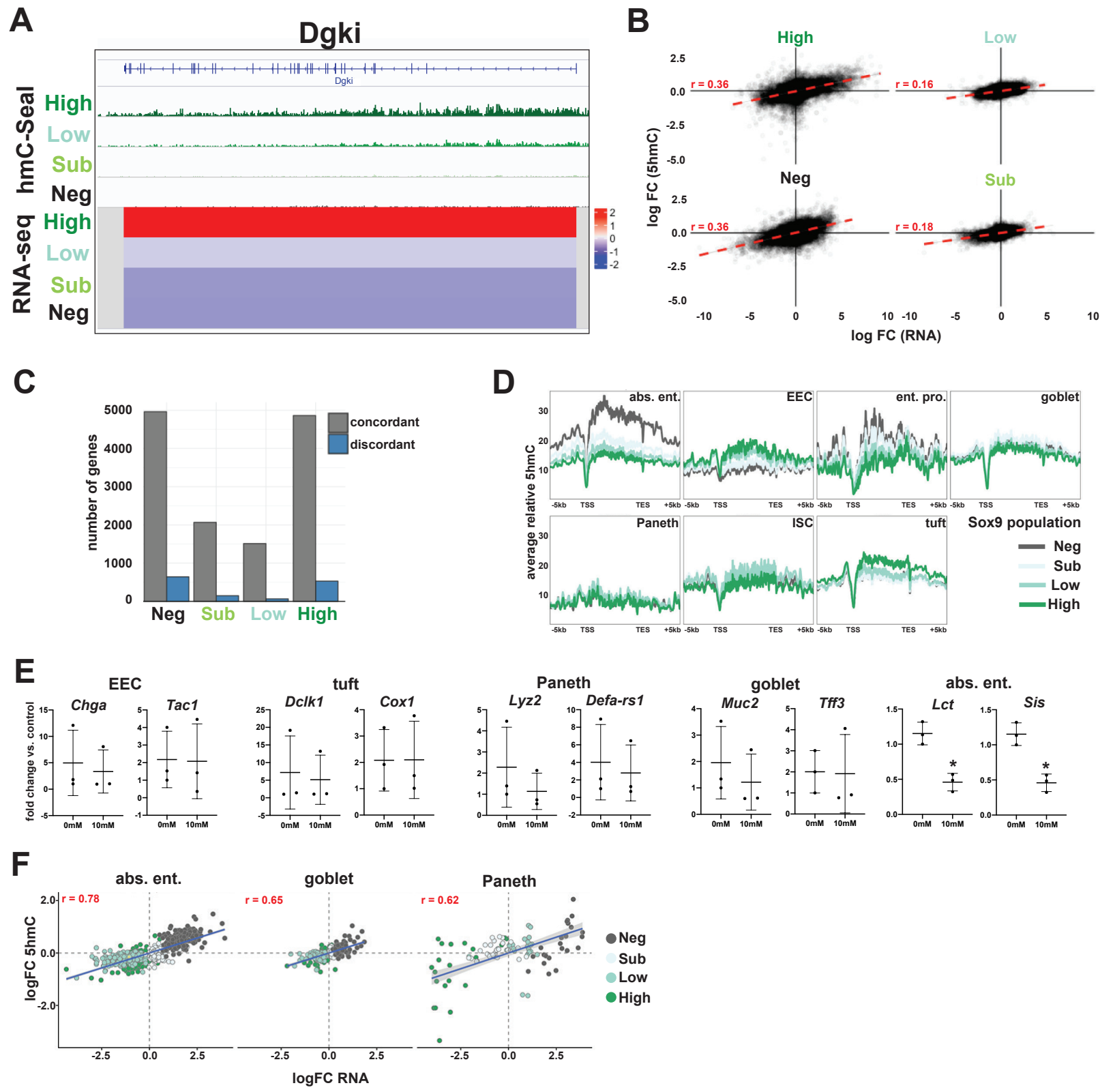
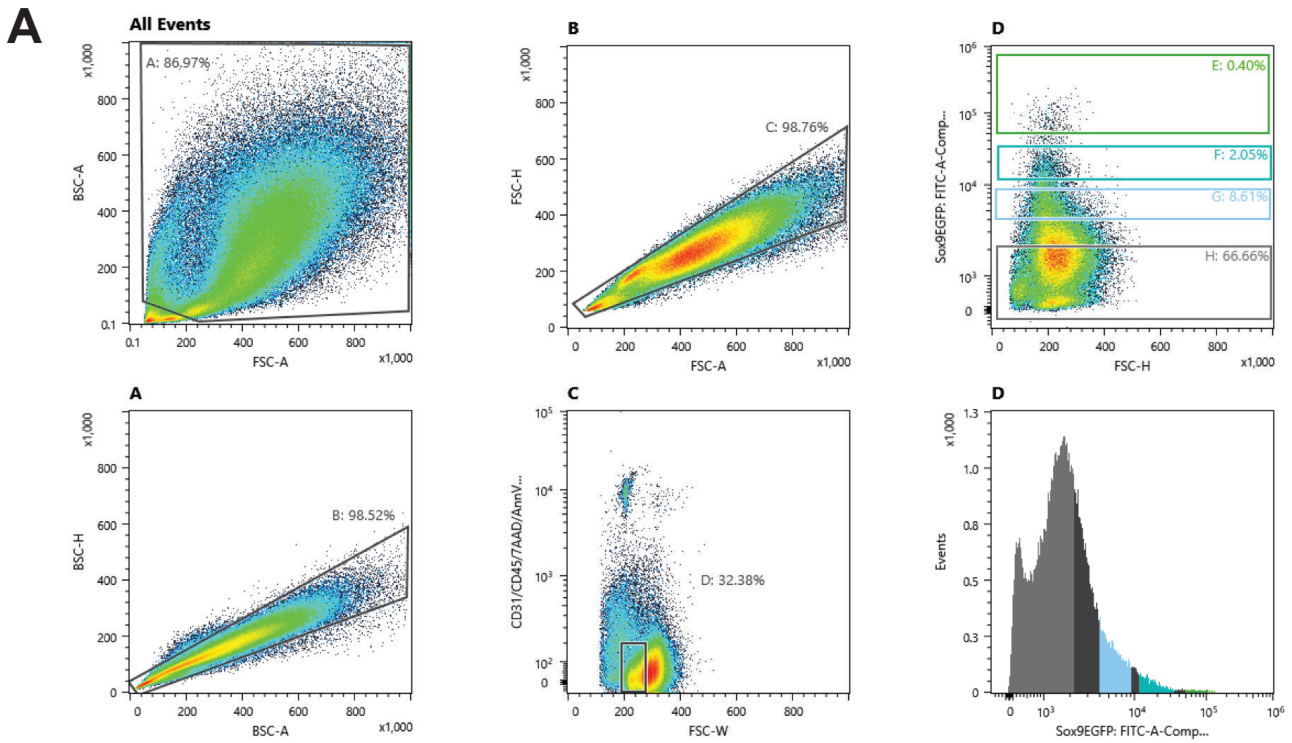
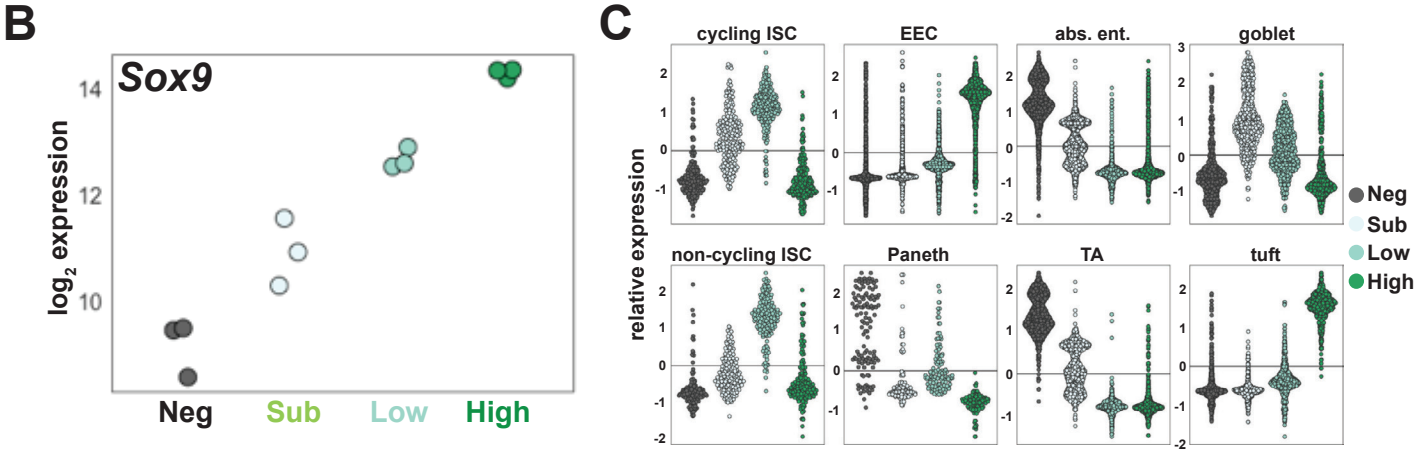

A
variability (chromVAR)
TF ranked by variability
Hnf4g
Rfx6
Rfx3
Nr1h3
Nr1h2
Foxa1
Rfx5
Nr2f6
Klf8
Klf2
Klf12
Klf5
Ppard
Rfx2
Bach1
Ppara
Bach2
Klf7
Esrra
Rxra
Rara
Id3
Rfx
Nr1h4
B
expression:accessibility correlation
RNA-seq
OCR
High
Low
Sub
Neg
C
Foxa1
Normalized TN5 insertions
Relative to Motif Center
TF Expression
Motif Accessibility
Hnf4g
D
Klf12
E
Stat3C
F
OCR distribution
by Sox9 population
UMAP 1
UMAP 2
High
Low
Sub
Neg
Hnf4g
Klf12
G
Foxa1
Foxa2
Rfx3
Rfx6

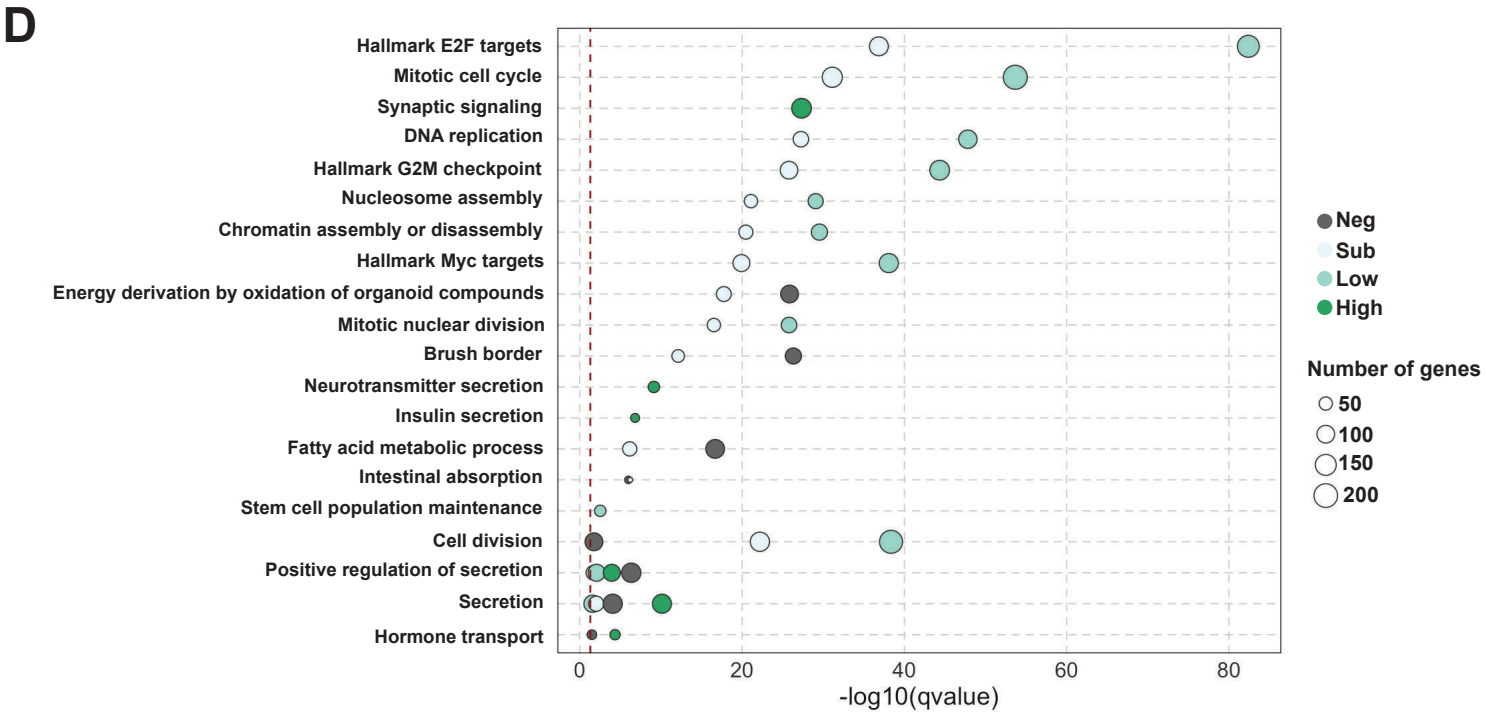
Figure 4


A
Dgki
hmC-Seal
High
Low
Sub
Neg
RNA-seq
High
Low
Sub
Neg
B
High
Low
Neg
Sub
r = 0.36
r = 0.16
r = 0.36
r = 0.18
log FC (5hmC)
log FC (RNA)
C
number of genes
concordant
discordant
Neg
Sub
Low
High
D
abs. ent.
EEC
ent. pro.
goblet
Paneth
ISC
tuft
average relative 5hmC
Sox9 population
Neg
Sub
Low
High
E
EEC
tuft
Paneth
goblet
abs. ent.
Chga
Tac1
Dclk1
Cox1
Lyz2
Defa-rs1
Muc2
Tff3
Lct
Sis
fold change vs. control
0mM
10mM
F
abs. ent.
goblet
Paneth
r = 0.78
r = 0.65
r = 0.62
logFC 5hmC
logFC RNA
Neg
Sub
Low
High

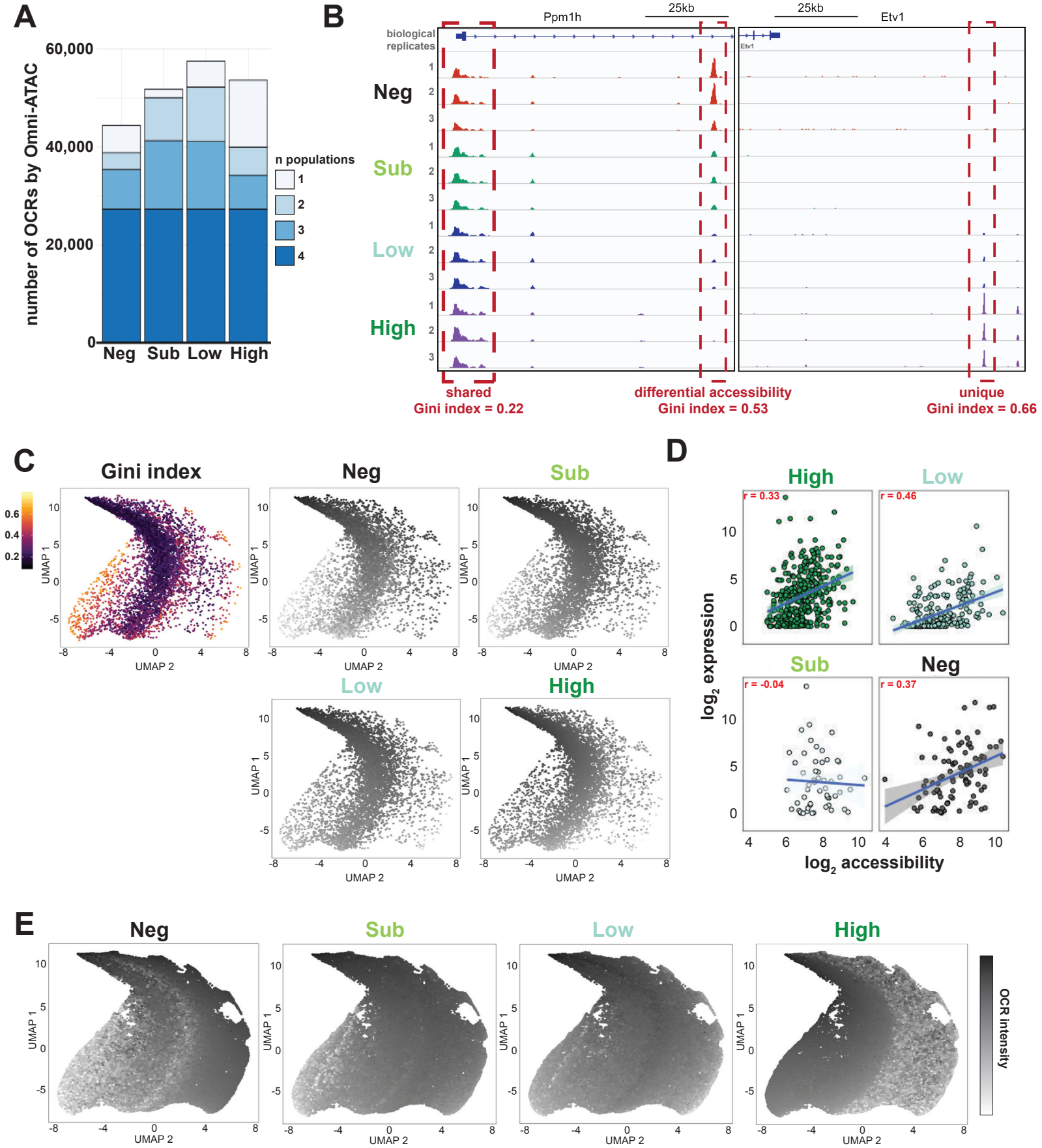
Supplemental Figure 1



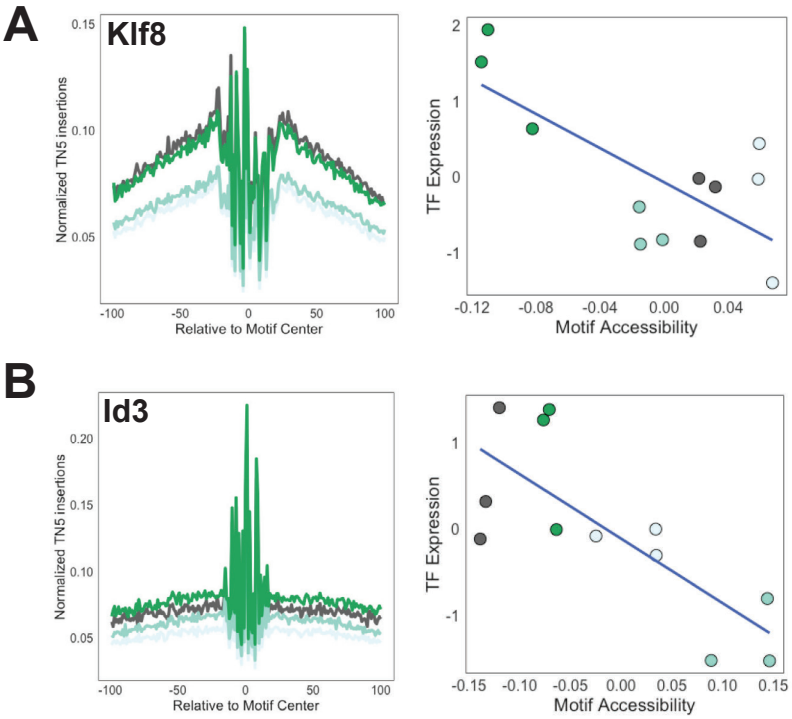




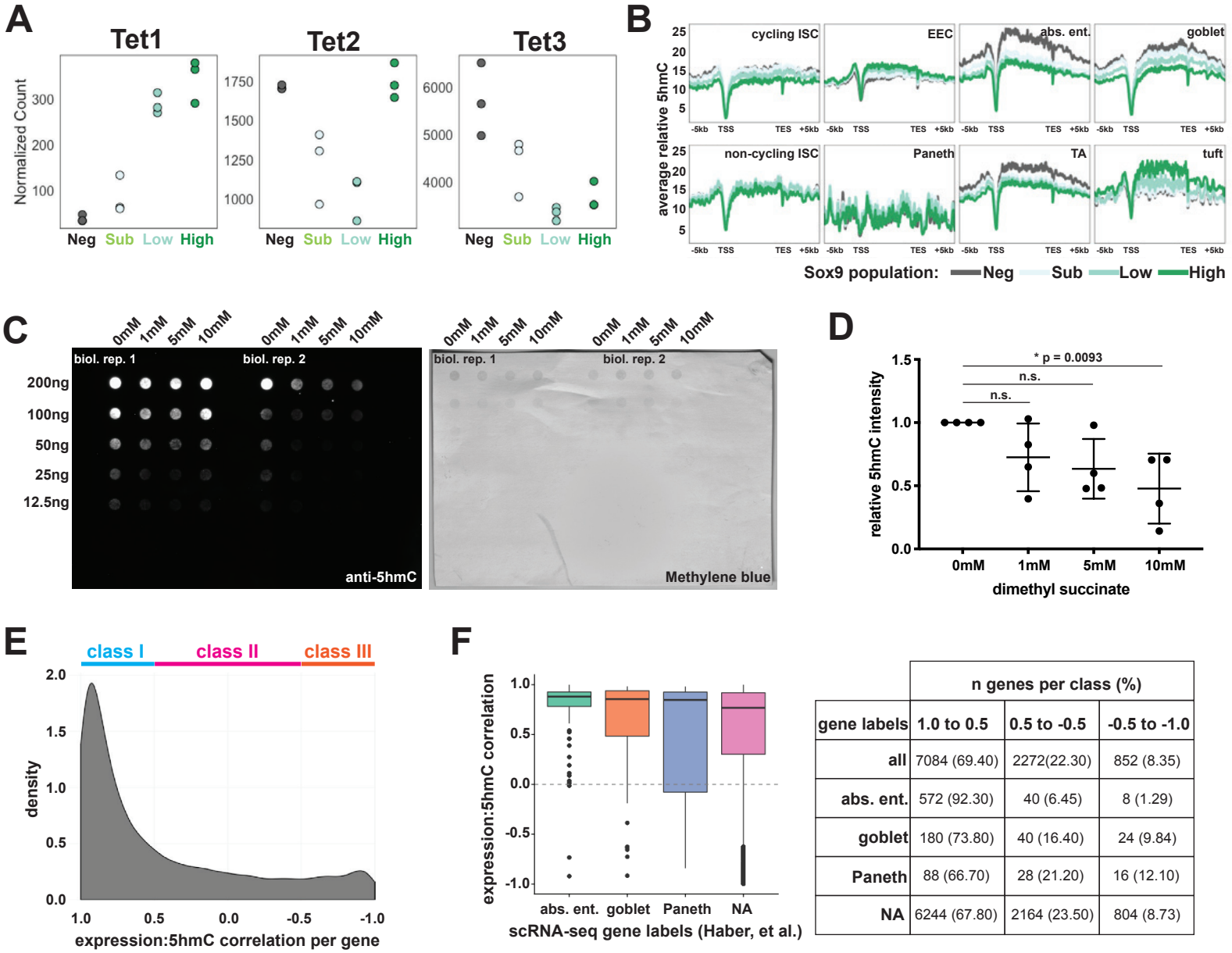
Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4



	n genes per class (%)		
gene labels	1.0 to 0.5	0.5 to -0.5	-0.5 to -1.0
all	7084 (69.40)	2272(22.30)	852 (8.35)
abs. ent.	572 (92.30)	40 (6.45)	8 (1.29)
goblet	180 (73.80)	40 (16.40)	24 (9.84)
Paneth	88 (66.70)	28 (21.20)	16 (12.10)
NA	6244 (67.80)	2164 (23.50)	804 (8.73)