IDENTIFICATION OF TRANSCRIPTIONAL REGULATORY MECHANISMS MEDIATING HOST RESPONSES TO MICROBIOTA IN THE INTESTINAL EPITHELIUM

James M Davison

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the curriculum of Cell Biology and Physiology in the School of Medicine

Chapel Hill
2017

Approved by:

John F. Rawls
Patrick Brennwald
Douglas M. Cyr
Michael B. Major
Praveen Sethupathy

ABSTRACT

James M Davison: Identification of transcription regulatory mechanisms mediating host responses to the microbiota in the intestinal epithelium
(Under the Direction of John F. Rawls)

All organisms must detect and respond to environmental pressures or else risk death. For animals, these pressures include maintaining symbiosis with the microorganisms that dominate their world. Over the course of evolution, these intimate microbial relationships have influenced animal tissue function and cellular identities. Microbial impact on animal cellular identity is most salient in the intestinal epithelia which interfaces with the largest concentration of microorganisms on any animal surface. In this dissertation, I explore the genomic and transcriptional regulatory mechanisms that mediate microbial tuning of intestinal epithelial identities. The collection of microorganisms that reside in the intestine (the intestinal microbiota), contribute to host physiology by facilitating energy harvest, tuning metabolic programs, promoting epithelial barrier function, promoting epithelial renewal and promoting immune system development. In addition to these important roles in health, intestinal microbiota have been implicated in a growing number of human diseases associated with loss of intestinal epithelial identity and function like Inflammatory Bowel Diseases. The microbiota impact intestinal epithelial function in part by regulating the expression of hundreds of genes in intestinal epithelial cells. Extensive research has identified the downstream physiological consequences of this transcriptional control. However, there remains a significant gap in our understanding of the upstream molecular mechanisms that mediate these host transcriptional responses. I identified that zebrafish transcription factor Hepatocyte nuclear factor 4 alpha (Hnf4a) specifically binds and

activates a microbiota-suppressed intestinal epithelial transcriptional enhancer. Deletion of this transcription factor in zebrafish revealed that it activates nearly half of the genes that are also suppressed by the microbiota, suggesting that its activity is negatively regulated upon microbiota colonization. Experiments from intestinal epithelial cells from gnotobiotic mice revealed that microbiota colonization is associated with genome wide reductions in HNF4A DNA occupancy. Similarly, HNF4A binding sites were associated with hundreds of microbiota-activated or microbiota-inactivated enhancers. These data indicate HNF4A may be an important regulator in the host response to the microbiota. Together, these data provide a novel genomic mechanism for understanding how the microbiota tune intestinal epithelial transcription programs and may contribute to Inflammatory Bowel Diseases.

ACKNOWLEDGEMENTS

Foremost, I want to thank Dr. John Rawls, who accepted me into his lab and chose to mentor me under unusual circumstances. You gave me a wonderful opportunity to succeed in your research program and I am very grateful. As I move into the next phase of my career in science, I will continue to learn from your expertise.

I want to thank my thesis committee for their advice and criticisms throughout these past several years. Thank you to Dr. Praveen Sethupathy for your encouragement and interest in my project. Thank you to Dr. Ben Major for your attention to detail in experimental design as well as figure design. Thank you to Dr. Patrick Brennwald for your advice and concern about my graduate career. Thank you to Dr. Douglas Cyr for your critical assessment of my progress and interest in my success.

I want to thank Dr. Vytas Bankaitis, who showed initial interest in my education and helped springboard my graduate career. I am grateful for his interest in my success.

Many thanks to Dr. James Alb who helped guide me during a difficult transition period and maintained a constant level of care and interest in my personal welfare and my education. We met many years ago and throughout these years he has helped me grow as a scientist and as an connoisseur of fine drink.

I want to thank both past and present members of the Rawls lab: Gray Camp, Ivana Semova, Sandi Wong, Jordan Cocchiaro, James Minchin, Patrick Williams, Sheila Janardhan, Colin Lickwar, Lihua Ye, Jia Wen, Sol Gomez De La Torre Canny, Jessica McCann, Rebecca Graham, Caitlin Murdoch and Ted Espenschied. In particular, I want to thank Colin for his advice and teaching of the bioinformatics analysis. Our conversations of

science or otherwise have always been a regular source of inspiration. I also want to thank Caitlin and Ted for their collegiality.

I want to thank Dr. Greg Crawford and Dr. Ling Song for their support with my genomics projects. I also want to thank Dr. Ghislain Breton for the use of his Yeast-1-hybrid library. I thank Balfour Sartor, Scott Magness, Maureen Bower, and Ted Espenschied for assistance with gnotobiotic mice, and Wenbiao Chen and Stacy Horner for sharing reagents. I also thank the Genomic Sequencing Laboratory at HudsonAlpha Institute for Biotechnology and the Duke Sequencing and Genomic Technologies Facility. This work was supported by grants from the National Institutes of Health (R01-DK081426, U24-DK097748, P01-DK094779, R24-OD016761, and P30-DK34987).

I want to thank Chris Blowers. Your company has always provided an escape from the responsibilities of graduate school. I also want to thank Tony and Miranda Krueger for your friendship and support during these many years.

I also want to thank my parents, Bob and Margie, who moved to North Carolina to see and help with their grandson. Their love and guidance during my childhood laid the foundation for my successes in graduate school. I want to thank my sister for her support, understanding and love. I also want to thank my brothers- and sisters-in-laws for their interest and concern while I have been in graduate school. I want to thank my parents-in-law, Tim and Sandy Gallagher for their constant love and never begrudging me of moving their only daughter away from the rest of their family.

The two people I am most grateful for are my wife and son. You have deserved far more care and attention than I have provided these past several years. Nicole, thank you for the constant support and thank you for being the primary care giver to Seamus while I have been absent and preoccupied with this work. Seamus, thank you for being my inspiration these past 2 years. During the hardest times of my graduate career and research, you have been where my mind goes to maintain a level of happiness.

PREFACE

Essay in Idleness #127
By Kenkō
Translated by Donald Keene

It is best not to change something if changing it will not do any good

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF ABBREVIATIONS

ANGPTL4	Angiopoietin-like 4; a secrete peptide that inhibits lipoprotein lipase
ATOH1	Atonal homolog 1; a mammalian homolog of the *drosophila* transcription factor *atonal* – involved in secretory cell lineage differentiation
CV	Conventionalized; Animals that had been germ-free but have been colonized with a conventional microbiota
CDX2	Caudal-related homeobox 2; a mammalian homolog of the *drosophila* transcription factor *caudal* – involved in intestinal epithelial idenitity
DHS	Dnase hypersensitivity site
GF	Germ-free; animals that do not have a microbiota
H3K4me1	Hisone 3, Lysine 4 monomethylation; a post translational modification on a histone 3 tail on lysine 4
H3K27ac	Histone 3, Lysine 27 acetylation; a post translational modification on a histone 3 tail on lysine 27
HNF4A	Hepatocyte nuclear factor 4 alpha; a nuclear receptor transcription factor
IBD	Inflammatory Bowel Disease
Mut	Mutant
TF	Transcription factor
WT	Wildtype

CHAPTER 1: INTRODUCTION

All animals maintain intimate relationships with the microbial communities that reside on their surfaces. The largest of these microbial communities occupies the animal intestine and are named the gut microbiota. This microbial assembly of viruses, fungi, and bacteria interface with the intestinal epithelium. The intestinal epithelium is comprised of a single layer of columnar epithelial cells that harvest dietary nutrients from the lumen to maintain energy homeostasis for the animal. Furthermore, this single layer of cells also contains several specialized cell types that detect fluctuations in the environment of the intestinal lumen, including changes to the microbiota. These cells all contribute to maintaining an effective barrier against the microbiota residing in the lumen. Improper epithelial response to these fluctuations have the capacity to disrupt homeostasis with the microbiota or impair energy homeostasis. Therefore, the intestinal epithelia must be poised with a variety of mechanisms that restrict aberrant responses yet maintain sensitivities to the environment while preserving vital absorptive and barrier functions. Most of the known sensory and response mechanisms impact transcriptional regulation that facilitate appropriate handling of the environment and the microbiota.

There remains intense interest in understanding how the intestinal microbiota impact human health. The microbiota can reprogram metabolic homeostasis, edify the innate and adaptive immune systems, and fortify skeletal bone. However, the microbiota are also associated with several human diseases including Inflammatory Bowel Diseases (Ulcerative Colitis and Crohn’s Disease) and Metabolic Syndrome. We understand how the microbiota contribute to these processes through the help of gnotobiotic animal models. Gnotobiotics is a controlled experimental system that enables scientists to account for all organisms within

an environment. Using gnotobiotics, scientists have the capacity to add a controlled community of microbes to an animal model and observe how the microbes impact animal physiology. By using gnotobiotic systems, scientists now understand the profound rewiring of transcriptional programs that occurs in animal tissues (especially the intestinal epithelia) upon the introduction and establishment of a microbiota.

Recent advances in the microbiome field have utilized whole genome analysis in conjunction with gnotobiotic animal models to test how the microbiota impact host epigenetics. By performing these types of experiments, scientists hope to understand the genomic mechanisms that mediate the host response. By identifying specific genomic and transcriptional regulatory mechanisms, scientists can begin to identify the broader host signaling pathways and microbial factors that control the host response and human health. Throughout the dissertation, I use the term “transcription program”. to refer to the network of regulatory mechanisms that control the expression of genes. These mechanisms include but are not limited to nucleosome location, histone modification, transcription factor binding, cofactor interactions, RNA polymerase binding, microRNA and long-non-coding RNA activities. Together these mechanisms function to activate and/or deactivate transcription of genes that can help determine cellular function and identity. However, I bias the discussion toward the transcription factors that mediate these “transcription programs”.

A central question that I maintain as a theme in this dissertation is “How does the intestinal epithelia remain sensitive to the luminal environment without losing its identity?” In other words, if the microbiota are capable of modifying epithelial cell fate and cell decisions, how does the intestinal epithelia balance sensitivity and responses to the environment while maintaining epithelial function and intestinal homeostasis? This question is introduced and expanded in chapter 2 of this dissertation. I also discuss what we currently know about transcriptional programs and the transcription factors that control these programs in the intestinal epithelia. I then highlight the transcriptional regulatory mechanisms that tune

inflammatory, metabolic, cell fate, and circadian rhythm networks in response to the environment. In chapter 3 of this dissertation, I discuss the use of a yeast-1-hybrid assay to identify transcription factors that mediate epithelial response to the microbiota. I then use both gnotobiotic zebrafish and mice to validate the discoveries made using the yeast-1-hybrid as well as make new observations about the role of a transcription factor, called Hepatocyte Nuclear Factor 4 (HNF4A), in the epithelium's response to the microbiota. I conclude chapter 3 with multi-species meta-analysis to determine new links between the microbiota and HNF4A and their contributions to human diseases. In chapter 4 of this dissertation, I discuss the possible mechanisms that mediate microbial suppression of HNF4A. I explore the possible evolutionary advantages for both the host and the microbiota. I speculate that rather than a commensal relationship that drives this microbial suppression of HNF4A, it's an antagonist relationship prompted by the host to secure an advantage in the warfare for resources in the intestinal lumen. I discuss how the HNF4A-microbiota interaction may be linked to human diseases and expand upon findings in chapter 3 regarding human IBD. I conclude the dissertation with a discussion about the expansion of the nuclear receptor and their role in metazoan evolution. Together this dissertation provides a framework for how the luminal environment impacts intestinal epithelial transcriptional regulatory mechanisms.

CHAPTER 2: THE PARADOX OF MAINTAINING CELL IDENTITY WHILE REMAINING SENSITIVE TO A STOCAHSTIC AND COMPLEX ENVIRONMENT

2.1 Overview

All animal tissues and cells must maintain a set of transcriptional programs that define their function (identity). Dysregulation of these transcriptional programs can result in human diseases such as cancers or Inflammatory Bowel Diseases (IBDs). However, animal cells must also remain responsive to changes in their environment to proceed through development or maintain physiological homeostasis. Therefore, cells must balance between “rigid” and “flexible” transcription programs to maintain both cellular identity and sensitivity to the environment. No cellular environment may be more stochastic or challenging to respond to than the intestinal lumen that interfaces with the intestinal epithelia. The intestinal epithelia maintain identity despite its exposure to a battery of growth factors, organic molecules, electrolytes, minerals, and microorganisms that have the capacity to modulate its transcription programs. The epithelium is comprised of several different specialized absorptive and secretory cell types that maintain epithelial function and each have their own unique set of transcription programs that define their identities. Each of these cell types similarly, must respond appropriately to the luminal environment to maintain intestinal homeostasis. These responses include tuning of their transcriptional programs which can impact their development, inflammatory response and energy harvest. Gross dysregulation of these same transcriptional responses can lead to human diseases such as IBDs where cellular identities are modified and the epithelia fails to appropriately respond the luminal environment. Paradoxically, a highly plastic transcriptional regulation is paramount to epithelial function and maintaining intestinal homeostasis and identity.

2.2 Introduction

The intestinal epithelium is a rapidly renewing and remarkably resilient tissue that interfaces with a complex luminal environment. At this interface, specialized cells that comprise the epithelium establish defensive barriers between itself and the intestinal microbiota. Most epithelial cell types absorb dietary nutrients while other cells communicate nutritional conditions to extra-intestinal tissues. The identity of each of these cell types is determined by a set of transcription factors that coordinate activities to establish and maintain cellular function. However, the transcriptomes for individual cell types varies along the length of the intestine, demonstrating a plasticity of these transcriptional programs (or regulatory transcriptional networks). Indeed, these malleable transcription programs permit the different cellular functions along the length of the intestine required to maintain animal health. However, aberrant dysregulation of these transcription programs can lead to the onset of human diseases such as cancer [1] and are associated with inflammatory bowel diseases [2].

The common function of all intestinal tracts is to harvest usable energy necessary for animal growth and survival from exogenous dietary nutrients (food) while maintaining a barrier between the microbiota and the animal. Along the length of the intestine, both host and microbiota catabolize and solubilize complex macromolecules to usable energy sources such as peptides and fatty acids (Figure 2.1A-C). The intestinal epithelium absorbs these dietary nutrients to achieve energy homeostasis for the whole organism. Both the microbiota and the host secrete digestive enzymes and small molecules into the lumen, which assist in nutrient harvest. Together with the dietary nutrients, these secreted factors contribute to an already complex luminal environment that interface with and direct intestinal epithelial differentiation and processes (Figure 2.2A) [3-5].The relative concentrations of these dietary factors change along the length of the intestine and these correlate with epithelial function. For examples, the most proximal segments of the mammalian intestine, the duodenum and

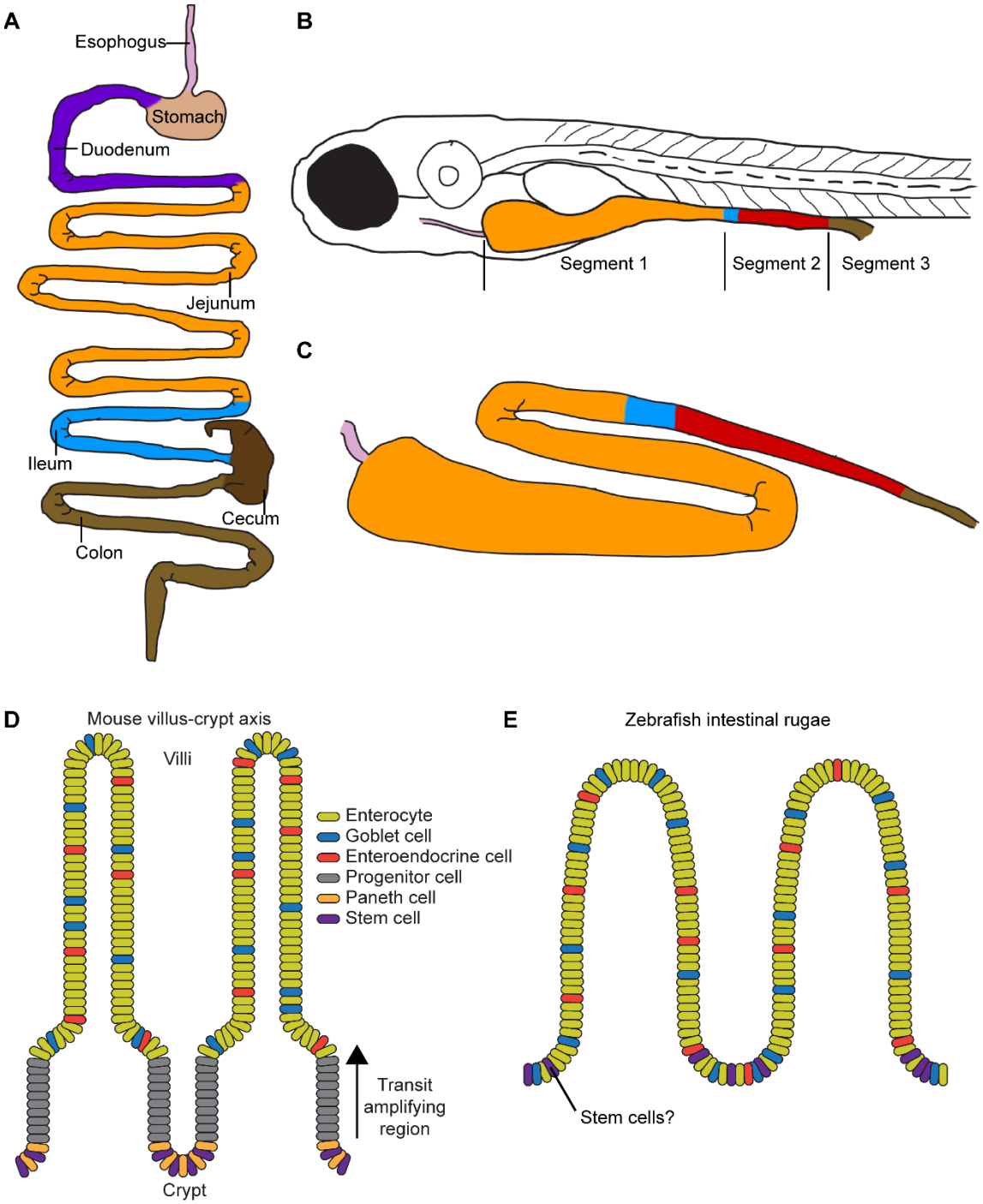


Figure 2.1: Macroscopic and microscopic morphology of the mouse and zebrafish intestine. (A) A schematic of the mouse gastrointestinal tract which is comprised of an esophagus, stomach, small intestine (duodenum, jejunum, and ileum), the cecum, and the colon. (B) A schematic of the larval zebrafish digestive tract. (C) A schematic of the adult zebrafish digestive tract. The coloration of the zebrafish intestines in the schematic represents the conserved regional specification shared between the zebrafish and mouse. The zebrafish lacks a stomach and the intestinal segments are classically broken up into three parts: Segment one, which based on gene expression resembles the mouse duodenum and jejunum segments; Parts of Segment 2 resembles the ileum in that a small section (highlighted in blue) expresses an ileum specific gene, fabp6, involved in bile acid absorption

(Lickwar et al; in revision)[6]. Segment 2 is also highly enriched in goblet cells [7]. Further the epithelial cells contain large vacuoles that are associated with lysosomal membrane proteins cells [7, 8]. Based on gene expression and the larger concentration of bacterial load in the intestinal lumen [9], Segment 3 resembles the mammalian colon. (D) Schematic of the mouse small intestinal epithelia showing the villus-crypt axis. Most epithelial cells are represented in the schematic; however, Tuft cells, mention briefly in this chapter are not depicted. Programmed cell "death", termed anoikis, occurs most frequently at the villus apex [10]. (E) Schematic of the zebrafish intestinal epithelia. Zebrafish do not have crypts or villi. Instead, their intestinal morphology resembles the mammalian stomach with epithelial folds termed rugae. The zebrafish have many of the same cell types including enterocytes, enteroendocrine cells and goblet cells. There is no evidence for Paneth cells. There is recent evidence for fish having a proliferative cell compartment and perhaps also stem cells near the base of the rugae [6, 11, 12].

jejunum, function as the primary location for protein, fatty acid and simple carbohydrate absorption (Figure 2.1A and Figure 2.2B). Bile acids enter the intestinal lumen through the common bile duct in the duodenum and facilitate fatty acid digestion by breaking down large dietary fat droplets into micelles that are digestible by secreted lipases. This process primarily occurs in the duodenum and jejunum where these fatty acids are absorbed and then transported to other tissues in the body for energy or storage. Bile acids are reabsorbed in the subsequent segment of the small intestine, the ileum, distal to the site of fatty acid absorption (Figure 2.2B). These bile acids are transported back to the liver and again stored in the gall bladder until they are again secreted into the duodenal lumen following a meal [13]. This enterohepatic bile acid circuit represents the logical layout of the intestinal tract, where molecules are absorbed after they facilitate nutrient harvest in previous segments. Water and electrolytes are similarly absorbed in the most distal part of the intestine [14]. The zebrafish and mouse intestinal epithelia demonstrate conserved regional and cellular specification that each participate in key aspects of energy harvest and digestive physiology (Lickwar et al, in revision)[6, 12] (Figure 2.1).

The cellular specification and luminal environment not only changes along the proximal-distal axis of the intestine, but a gradient of mucus and antimicrobial peptides that protects the intestinal epithelia from the microbiota contributes to distinct microenvironments within the invaginations of the epithelia called the intestinal crypts. The multipotent intestinal stem cells (ISCs) reside at the base of these intestinal crypts (Figure 2.1D). As ISCs rapidly

and asymmetrically divide, they give rise to undifferentiated progenitor cells that migrate away from the base of the crypts staying along the epithelial layer. These progenitor cells migrate through the transient amplifying region where rapid cell division and the initiation of specific transcription programs begin to determine their cellular function (Figure 2.1D). The luminal environment is known to direct these transcription programs and control aspects of epithelial identity.

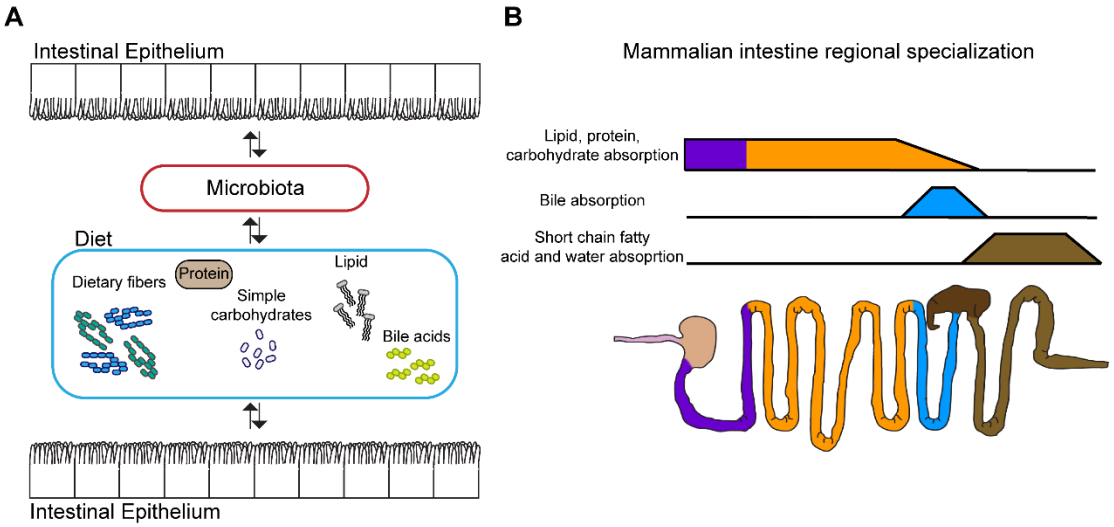


Figure 2.2: The intestinal luminal environment changes along the length of the intestine. (A) A diagram illustrating the interaction between the microbiota, the diet and other molecules within the intestine and the intestinal epithelia. (B) A diagram depicting the functional specification along the intestine and the major processes that occurs along each segment in the mouse intestine.

Loss of cellular identity promotes the onset of human diseases like cancers [15]. Each animal tissue maintains cellular identity using a variety of different strategies. Some cells progress slowly, if at all, through the cell cycle which reduces the rates of mutations generated during DNA replication [16]. Interestingly, most genetic mutations that manifest in human disease are in non-coding regions of the genome, indicating that non-coding transcriptional regulatory mechanisms like transcription factor binding or non-coding RNAs are critical aspects of disease etiology [17, 18]. Activating or repressing transcription factors recognize and bind specific DNA sequences within *Cis*-regulatory regions (CRRs), which

can be located proximal to the transcription start sites (TSS), within gene bodies, as well as in intergenic regions [19]. CRRs are generally depleted of nucleosomes ("accessible chromatin"), which can be experimentally captured by hypersensitivity to DNase I cleavage [20] (Figure 2.3). CRRs are also generally associated with specific post-translational modification of histone proteins within adjacent nucleosomes. For example, mono-methylation of lysine 4 (H3K4me1) and acetylation of lysine 27 of histone H3 (H3K27ac) distinguishes between CRRs that act as poised and active enhancers respectively [21]. CRRs associated with H3K27ac marks are "permissive chromatin", meaning the chromatin arrangement promotes transcription factor binding and induces transcription of the regulated genes. Genome wide binding locations of transcription factors and locations of H3K27ac modifications can be captured using ChIP-seq (Figure 2.3). Specific types of transcription factors, termed pioneer factors or master transcription factors, orchestrate the accessible and permissive chromatin landscape by initially displacing nucleosomes and recruiting histone modifying enzymes to regulatory sites [22-25]. These processes enable the recruitment of other transcription factors to the sites of open and permissive chromatin and promote transcription of the regulated gene. Similarly, competing transcription factors may bind to the same site and repress transcription of the regulated gene [26].

The intestine-specific master transcription factor, Caudal type homeobox 2 (CDX2), functions as a pioneer factor that imprints and retains accessible chromatin in ISC and progenitor cell types within the intestinal epithelium [27-31]. Aberrant activity of CDX2 in esophageal epithelia is an early marker in Barrett's disease where esophageal epithelial cells take on a small intestine epithelial identity [32, 33]. Similarly, loss of CDX2 activity is associated with the development of a specific type of colorectal cancer [1, 34]. Interestingly, changes in CDX2 activity is also associated with molecules in the luminal environment. Increased bile acid concentrations in the esophagus induces CDX2 expression through activity of another transcription factor, NF-κB [35]. This aberrant CDX2 expression imprints a

chromatin architecture in esophageal epithelia that promotes small intestinal epithelial transcription programs and loss of esophageal identity. Similarly, the short chain fatty acid, butyrate, stimulates CDX2 expression in colon cancer cell lines which may protect against oncogenesis and thus preserve epithelial identity [36]. To suppress disease onset, it may be beneficial for the intestinal epithelia to adopt a rigid transcription program that would be less sensitive to environmental stimuli. However, the primary role of the intestine is to harvest nutrients for the organism to maintain energy homeostasis. To accomplish this goal, the epithelia may need to detect and respond to the millions of molecules in the intestinal lumen. How does the intestinal epithelium maintain cellular identity while remaining sensitive to its environment? What are the transcriptional mechanisms that permit both sensitivity and retain cell identity in the intestine? In this chapter, I will explore the transcription programs that control cellular identity followed by the programs that permit cellular sensitivity. I then explore the overlap of these transcription programs and how dysregulation may lead to the onset of inflammatory bowel diseases.

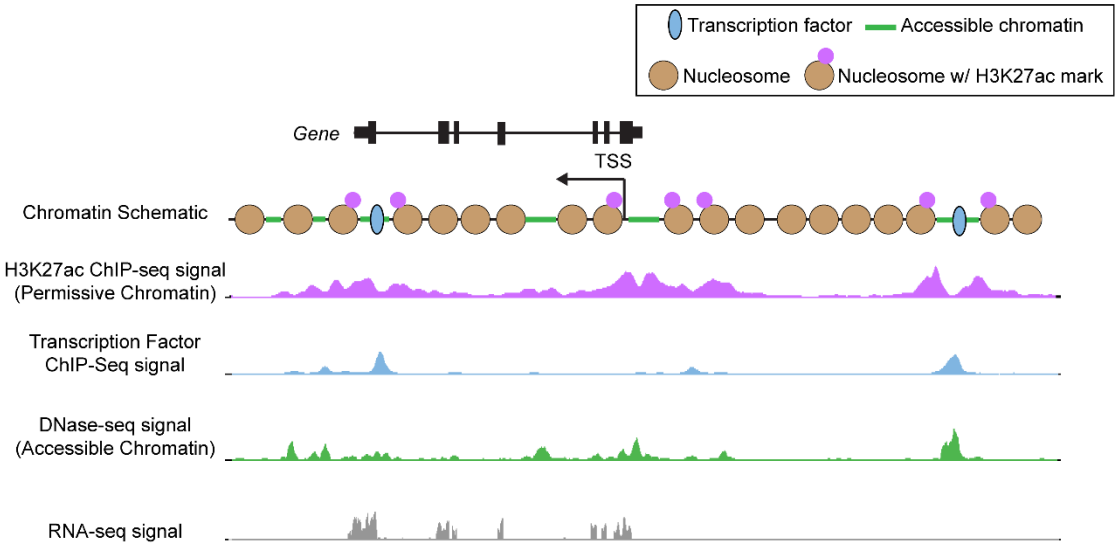


Figure 2.3: Genome wide sequencing techniques used to identify mechanisms of transcriptional regulation. Illustrated in this figure is a gene locus, three genome browser tracts each depicting different types of chromatin architecture, and a cartoon that summarizes what those genome tracts indicate.

2.3 The basic architecture of the transcription programs that maintain intestinal epithelial identity

To maintain homeostasis with the microbiota and the luminal environment, the mammalian intestine has evolved several different cell types that maintain a barrier between the exogenous environment and the host. These cells are programmed by transcription networks that determines their specific cellular function and identity. These cell types include enterocytes, enteroendocrine cells (EECs), cup cells, tuft cells, Paneth cells and goblet cells (Figure 2.1D). Each one of these cells types differentiates from the ISCs residing in the base of the intestinal crypts [37]. In the next subsections, I will explore the transcriptional programs that control cell identity for the majority of intestinal epithelial cell types and during their “lifespan” on the epithelial layer.

Throughout this chapter, I concentrate on the transcription factors that mediate intestinal epithelial cell function. These transcription factors activate or deactivate transcription of coding genes as well as non-coding RNAs like microRNAs and long-non-coding RNAs that have been shown to similarly control cellular identity. Recent studies have shown microRNAs in particular play an important role in tuning the host transcriptional response to the microbiota in multiple cell types [38-41]; however, for the following sections, I primarily discuss the transcription factors that control the transcription programs.

2.3.1 The intestinal stem cell and the journey toward anoikis

The intestinal epithelia have an underlying transcriptional program that maintains their identity despite rapid cell turnover. These transcription programs are defined during differentiation as cells migrate out of the intestinal crypts along the epithelial barrier. Most of the intestinal epithelia self-renews every 4 – 7 days and this process is driven by the rapidly dividing multipotent intestinal stem cells residing in the base of these crypts [42]. Unlike other mammalian stem cell types which are thought to remain in mainly quiescent states to

preserve their long-term proliferation and guard against DNA replication errors that can be passed onto daughter and progenitor cells [43, 44], the intestinal stem cell rapidly self-renews, producing daughter cells and progenitor cells that eventually differentiate into a specialized intestinal epithelial cell type [44]. Furthermore, unlike other stem cells, the intestinal stem cell is exposed to a potentially harsh external environment, and therefore the risk of injury and DNA damage is potentially greater [37]. As such, the intestinal stem cells are protected in specialized niches that maintain the proliferative potential while minimizing cellular and DNA damage. In both the large and small intestine, the ISCs are protected by a thick layer of mucus, primarily secreted by goblet cells, that blocks the invasion of potentially dangerous molecular signals or pathogens [45]. These niches are also composed of a subepithelial stromal microenvironment, and an epithelial/luminal microenvironment which is mainly supported by the adjacent cells [37]. In the small intestine, these adjacent cells include the specialized Paneth cells that sense and suppress microbiota infiltration and they nurse stem cells with pro-proliferative growth factors and signals to maintain the ISC proliferative potential [46, 47]. In the small intestine, Paneth cells are immediately apposed to these stem cells and as such are thought to be the primary cell type that sustains stem cell multipotency and maintain the stem cell niche in the luminal microenvironment. Colonic crypts lack Paneth cells. However colonic ISC are protected by sentinel goblet cells that never migrate out of the colon and secrete mucus upon the detection of harmful molecules [48]. In the small intestine, both the stromal microenvironment and Paneth cells are individually sufficient to maintain a functional stem cell niche and are further discussed in these reviews [37, 49, 50]. Wnt signaling plays a critical role in the maintenance of intestinal stem cells and crypts compartments. Mice lacking Tcf4, a transcription factor that becomes active after forming a complex with beta-catenin downstream of Wnt signaling, lack actively dividing cells in the base of their crypts. [51].

Progenitor cells differentiate into the specialized cell types that both survey the microenvironment and absorb exogenous and endogenously derived molecules as well as maintain homeostasis between the host and microbiota. These cells types fall into two major lineages: secretory cell lineages (Paneth cells, tuft cells, goblet cells, and enteroendocrine cells) and absorptive cell lineages (enterocytes). Following asymmetric stem cell division, progenitor cells begin a journey of further differentiation that proceeds along the villus-crypt axis through the transit amplifying compartment (Figure 2.2D) [37, 42]. During this process of differentiation, progenitor cells maintain an accessible and permissive chromatin architecture that resembles the intestinal stem cell [31, 52]. These studies indicate that distinct differentiation programs are not predefined by nucleosome location. Instead, cell type specific transcription factors likely bind to the accessible chromatin and direct differentiation. Conversely, mouse intestinal epithelial cells (IECs) exhibit distinct accessible chromatin profiles across segments of the intestine [53]. However, we do not know if these differences in chromatin accessibility are inherent to the stem cell populations from which they are derived or if the chromatin accessibility is modified during terminal differentiation to achieve a cellular identity that permits appropriate regional activity.

Fully differentiated enterocytes and goblet cells diverge from this ISC-like chromatin landscape [31, 52]. These data indicate that these cells express transcription factors that have the capacity to function as pioneer factors and govern terminal differentiation [54]. During intestinal epithelial differentiation, cell-type specific transcription factors, like HNF4A or RBPJ, bind to the open chromatin as determined by pioneer factors or other nucleosome displacement factors and initiate different transcription programs that drive lineage differentiation [23]. The transcription factor ATOH1 activates the transcription of genes that are involved in secretory cell differentiation [55, 56]. HES1 drives absorptive cell differentiation by suppressing ATOH1 and consequently suppressing secretory lineage differentiation [55]. Once fully differentiated, epithelial cells (with the exception of Paneth

cells and sentinel goblet cells) continue to migrate up the crypt (and along the villus in the small intestine) and end their journey in the zone of extrusion where mature 4 – 7 day old epithelial cells undergo apoptosis and leave the epithelium through a process called anoikis [10]. Anoikis is an important function in the repertoire of small and large intestinal epithelial cell identity. Aberrant repression of E-cadherin transcription in colon epithelial cells reduces anoikis and promotes tumor growth [57, 58]. Thus, the intestinal epithelia must maintain tight transcriptional control during cellular migration toward the zone of extrusion to maintain epithelial identity and homeostasis.

2.3.2 Secretory cell lineages

The transcription factor ATOH1 mediates differentiation of all secretory cell lineages [56]. Its activity and binding to different promoters in mouse colonic crypts is associated with the induction of other transcription factors including *Neurog3*, *Gfi1*, *Sox9*, *Creb3l142* and *Spedf* [59]. The activities of these transcription factors support differentiation into goblet cells, enteroendocrine cells, and Paneth cells. The roles of these transcription factors in secretory cell differentiation are each discussed in more detail below. Interestingly, ATOH1 also binds directly to the genes of the Notch signaling ligands *Dll1* and *Dll4*. Previous reports have indicated that ATOH1 expression induces lateral inhibition of secretory cell differentiation by inducing Notch signaling in neighboring cells and directing neighboring cells toward absorptive cell differentiation [31] (discussed in more detail below in 2.3.3). These studies demonstrate how ATOH1 activity helps determine secretory cell identity while also directing the absorptive cell identity in neighboring cells.

Enteroendocrine cells (EECs) comprise the largest population of chemosensory cells in the intestine although represent only a small fraction of total epithelial cells [42]. These cells detect dietary molecules within the intestinal lumen including amino acids, free fatty acids, bile acids and many more. During fasting and upon nutrient detection, EECs release

hormones like PYY and somatostatin that fine-tune both intestinal and systemic responses to the nutrient status [60]. Activity of the transcription factor NEUROG3 is necessary for differentiation of the EEC fate [61, 62]. Following expression of the *Neurog3* gene, EECs express transcription factors like NEUROD1, ISL1 and PDX1 which promote its endocrine signaling activities. For example, transcription factor PDX1 binds and activates the promoter of Glucose-dependent-insulintrophic polypeptide (GIP) [63], a signaling peptide that is secreted by EECs upon sensing dietary fat or carbohydrates and induces insulin release from pancreatic beta cells [64, 65]. This group of transcription factors also regulates transcription programs involved in pancreatic islet development which highlights the similarities between these hormone producing cell types [66]. The nuclear receptor transcription factor HNF4G similarly participates in glucose tolerance through regulating Glucose-like peptide (*GLP*) expression and secretion by EECs. [67] The transcription factors FOXA1 and FOXA2 have also been shown to regulate the transcription of several EEC-secreted hormones. These factors function downstream of NEUROG3 activity, but it is unknown if NEUROG3 regulates their transcription directly [68].

Goblet cells are professional mucus producing cells that maintain a thick mucus layer between the microbiota in the luminal environment and the intestinal epithelia [69]. These cells produce a highly glycosylated secretory mucin called MUC2 which is tightly folded, packaged and processed in the endoplasmic reticulum and Golgi networks and stored in secretory granulae at the cell membrane. Following secretion, mucus expands ~1,000 fold [70], forming large nets that in the colon have been shown to attenuate Brownian motion of molecules like microbe-associated molecular patterns (MAMPs) and blocks the invasion of the majority of the microbiota [71]. Like other secretory cell lineages, Goblet cell identity is primed by ATOH1 activity. ATOH1 induces expression of SPDEF and GFI1. GFI1 suppress *Neurog3* expression, thereby blocking differentiation toward an EEC identity [72]. Knockout mouse models indicate both SPDEF and GFI promote goblet cell differentiation [72, 73].

Similarly, both FOXA1 and FOXA2 knockout animals have reduced goblet cell differentiation [74]. The transcription factors FOXA1, FOXA2 and CDX2 have been shown to bind the promoter of *Muc2* and activate its transcription [68, 74, 75]. In gastric and lung epithelial tissues, the pioneer factor SOX2 promotes goblet cell differentiation [76, 77]; however, aberrant activity of SOX2 activity in the small intestine reprograms the epithelial identity into gastric-like epithelia [78].

Deletion of the *Sox9* transcription factor gene in the mouse intestinal epithelia results in attenuated differentiation of goblet cells, and complete ablation of the Paneth cell lineage, indicating a requirement for *Sox9* in Paneth cell fate determination [79, 80]. In cell culture, SOX9 suppresses WNT signaling by physically interacting with beta-catenin which results in degradation of both proteins [81]. Furthermore, in the intestine, SOX9 *trans*-activates the expression of Groucho-related inhibitors of the beta-catenin-TCF pathways and transcriptionally suppresses proliferation markers like *Cyclin-D1* and *c-Myc* [79]. Suppression of these pathways may indicate why secretory cells make up a small proportion of epithelial cells. Terminally differentiated Paneth cells do not express *Sox9*, indicating SOX9 activity is only required for fate determination and not function following differentiation [82]. Paneth cells differentiate amongst other progenitor cells and migrate down toward the stem cell niche where they reside in the crypt for 2 – 3 months [83, 84]. They contain large cytoplasmic secretory granulae that contain antimicrobial peptides and proteins like lysozyme [47, 85-87]. Paneth cells secrete these granulae into the crypt lumen to protect the precious stem cells and progenitor cells from certain microbiota that may threaten intestinal homeostasis. These cells also maintain stem cell longevity and promote progenitor differentiation by producing EGF, Notch, and WNT ligands [37]. Furthermore, the gradient of these ligands in the epithelial microenvironment control transcription programs in progenitor cells that ultimately influence cellular identity [46, 88]. The nuclear receptor transcription factor HNF4A may regulate Paneth cell function, following differentiation. The intestine

specific HNF4A knockout mouse has equivalent numbers of Paneth cells compared to control mice; however, the granulae within the Paneth cells of the knockout mouse fail to stain for lysozyme. This study indicates HNF4A may play important roles in maintaining Paneth cell function [89].

Tuft cells represent a minor percentage of secretory cells and seem to function as chemosensory cells [62]. Tuft cells differentiation requires ATOH1 activity, suggesting it belongs to the secretory cell lineages [62]. These cells detect pathogens in the luminal environment and respond by signaling to neighboring epithelial cells and cells residing in the lamina propria. These signals launch a type 2 immune response and induce tuft cell expansion [90]. Tuft cell biology remains relatively unstudied compared to other intestinal epithelial cell lineages. Only recently have researchers identified novel transcriptional programs that mediate Tuft cell differentiation. Indeed, deletion of the transcription factor POU2F3 in mouse results in failed Tuft cell differentiation and a deficiency to initiate the type 2 immune response to intestinal parasites [91, 92]. Each of these secretory cell identities are programed and maintained by unique transcription factor networks.

2.3.4 Absorptive cell lineages

Several of these transcription factors that participate in secretory cell function also participate in absorptive cell lineages. However, a master transcription factor, HES1, controls the initial “decision” to become either a secretory cell or an absorptive cell during early progenitor differentiation. HES1 expression is upregulated by active Notch signaling. The ligands Dll1 and Dll4 bind and activate the Notch transmembrane receptor which results in proteolytic cleavage of the intracellular-domain (NICD). The NICD translocates to the nucleus and forms a transcriptional complex with RBPJ and activates transcription of the *Hes1* gene [93]. Notch signaling along with enhanced beta-catenin signaling promotes rapid proliferation of these progenitor-absorptive cells [94, 95]. The transcription factor HES1

represses ATOH1 expression and promotes absorptive cell (enterocyte) differentiation [55]. In the small intestine, GATA transcription factors function in tandem to tune Notch signaling and promote enterocyte differentiation [96]. For example, GATA4 binds near the transcription start site of the Notch ligand *Dll1* [97]; And, GATA4 knockout mice have reduced progenitor proliferation and reduced *Dll1* expression, suggesting GATA4 directly activates *Dll1* transcription and promotes Notch signaling. Similarly, GATA6 promotes progenitor proliferation and GATA4 and GATA5 promote enterocyte identity by upregulating terminal differentiation genes in the intestinal villi [96]. Deletion of GATA4 and GATA6 in the mouse intestine results in decreased proliferation, reduced enterocyte enteroendocrine cell numbers and increased Goblet cell numbers [97, 98]. As mentioned early, ATOH1 binds directly to the genes of the Notch signaling ligands *Dll1* and *Dll4*, supporting its role in lateral inhibition of secretory cell differentiation of neighboring cells [59].

Enterocytes comprise the largest cell population in the small intestinal epithelia [99]. These cells are professional absorptive cells that transport dietary molecules across their plasma membrane. Amino acids and monosaccharides are transported across the plasma membrane by sodium dependent transporters [100]. However, the majority of water absorption in the colon is facilitated by electrochemical gradients generated by transport activity of short chain fatty acids generated by the microbiota [101]. Uptake of dietary fats occur only after triglycerides are cleaved to form free fatty acids and 2-monoglycerides which can enter the enterocyte by simple diffusion or through the help of long chain fatty acid transporters [102]. Once in the enterocyte, fatty acids can be reassembled into triglycerides and packaged in the endoplasmic reticulum to form lipid droplets for temporary storage or packaged in chylomicrons and trafficked into the body through the basolateral membrane [102]. Chylomicrons are large lipoproteins particles. The apolipoproteins APOA4 and APOC2 are highly expressed by enterocytes and secreted as surface competents on newly synthesized chylomicrons. APOA4 and APOC2 stimulate lipoprotein lipase activity

[103], an enzyme that is tethered to endothelial cells at sites throughout the body and cleaves triglycerides into free fatty acids permitting their absorption into the nearby cells for storage or energy consumption. Thus, these apolipoproteins promote serum triglyceride clearance. The transcription factor CREBH/CREB3l3 activates transcription of APOA4 and APOC2 in mice and supports serum triglyceride clearance [104]. Similarly, overexpression of *Crebh/Creb3l3* represses transactivation by binding directly to promoters of genes that mediate intestinal cholesterol absorption like *Npc1l1* [105]. Overexpression of *Crebh* also represses *Srebp2* expression, a transcription factor that has previously been shown to activate transcription of genes involved in lipogenesis and may mediate cholesterol homeostasis by binding the promoter and activating transcription of *Npc1l1* [106, 107]. In zebrafish, the nuclear receptor transcription factor LXRa was shown to regulate the delivery of ingested lipid to circulation, perhaps through its transactivation of a gene involved in biogenesis and growth of lipid droplets [108]. The transcription factor HES1 directs absorptive cell differentiation, however, once differentiated, several other transcription factors tune its absorptive function. Many of the enterocyte transcription factors have been shown to be regulated by the influx of nutrients into the lumen and these will be discussed in section 2.4.

2.3.5 Regional specification

Aside from cellular specialization along the villus-crypt axis, transcription networks generate distinct differences in cellular function along the proximodistal axis of the small intestinal tract. For instance, GATA4 suppresses ileal transcription programs and promotes the transcription of the fatty acid metabolism and absorption genes in the jejunum [97, 109]. Likewise, the nuclear receptor FXR is most highly expressed in the ileum and selectively regulates genes involved in bile acid absorption [110, 111]. Furthermore, FXR binds bile acids directly and activates transcription of *Fgf19 [112].* FGF19 (FGF15 in mouse) is an ileal

secreted hormone that regulates bile acid synthesis in the liver hepatocyte by inhibiting the enzyme that controls the rate limiting step of bile acid synthesis, CYP7A1 [112, 113]. Regionalization also impacts cell identity as well as gross morphology of the intestine. Indeed, the distal end of the small intestine has greater number of goblet cells compared to the proximal end, perhaps to protect the epithelia from the larger concentration of microbes in the posterior intestine [114, 115]. The primary secreted hormones by enteroendocrine cells changes along the length of the small intestine [60]. The duodenum harbors the highest concentration of EECs that produce hormones like secretin and motilin which mediate water homeostasis and peristalsis, respectively. Although EECs in the distal part of the small intestine have the capacity to secrete these hormones, they primarily secrete hormones that regulate the dopamine pathway (neurotensin) and glucogaon- and insulin-like peptides (GLP-1, GLP-2, and INSL5) [60]. Furthermore, the duodenal and jejunal segments have longer villi compared to the ileal segment [115, 116], perhaps because these regions are both exposed to a higher concentration of dietary nutrients and are responsible for the majority of the nutrient uptake including fatty acid absorption [114]. The longer villi in the anterior intestine optimizes surface area and exposure to the environment which increases the capacity of nutrient absorption. Intestine-specific deletion of the nuclear receptor transcription factor HNF4A reduces villus size and impairs fatty acid absorption in the jejunum and impaired colonocyte differentiation [89, 117, 118]. Since HNF4A is most highly expressed in the duodenum and jejunum and given the intestine specific knockout phenotype, transcription programs controlled by HNF4A may help define regional specification [28, 119]. Lastly, chromatin accessibility likely contributes to regional specification since accessible chromatin marginally differs in epithelial cells from anterior versus posterior segments of the small intestine [53]. Collectively, these insights into the transcriptional programs that contribute to intestinal identity provide an essential frame of

reference for how the epithelia tune these programs to effectively handle its changing environment.

2.4 The transcription programs that permit intestinal epithelial sensitivity

Microbiota colonization of germ free animals is associated with robust changes to intestinal transcription programs, resulting in differential regulation of hundreds of genes (see Chapter 3 and [119-121]). Similarly, the diet has also been shown to drastically change intestinal transcriptional programs [122]. Host dietary habits control the nutritional content within the lumen and it has direct effects on the microbial communities and the taxa present in the lumen [123-128]. Similarly, microbes in the intestine have the capacity to modify the dietary contents. Indeed, some bacterial taxa like *Rosburia intestinalis* encode genes that generate isomers of poly-unsaturated fatty acids which have been shown to affect host physiology [129] and the large cohort of anaerobic bacteria in the colon have the capacity to catabolize dietary fibers that generate short-chain fatty acids [130]. Diet and microbiota mediate activity of transcription programs that fall into four categories: 1) developmental programs, 2) metabolic programs, and 3) immune/inflammatory programs 4) circadian rhythm programs [120, 121]. This categorization of transcription is not entirely accurate since the activation of one of these programs may result in activation or suppression of the others [131-135] and therefore they are not truly distinct transcriptional programs. Instead, their regulatory mechanisms often interact to define cellular function (Figure 2.4). Furthermore, activation of these transcription programs are selectively modified by the presence of specific bacterial taxa and the factors these bacteria produce [136-139], indicating IECs respond to specific cues in the environment. Interestingly, mouse and zebrafish intestinal epithelial cells share a conserved transcriptional response to microbiota colonization indicating the same regulatory pathways may be mediating these responses [137, 140].

Only recently have studies identified the epigenetic and genomic changes that underlie some of these transcriptional and downstream responses. Surprisingly, microbiota colonization does not significantly impact chromatin accessibility in the intestinal epithelia, suggesting that microbiota modulate epithelial transcription by modifying the activity of specific transcription factors [53]. However, recent studies suggest histone modifying and DNA-methylation enzymes play key roles in the epithelial response to diet, the microbiota, and antibiotic treatment [141-144]. We still lack an understanding if microbiota regulate transcription factor binding genome wide.

Identifying the transcription factors that mediate epithelial response has become an intense field of study due the microbiome's relationship with human diseases such as metabolic syndrome and inflammatory bowel diseases. The transcription factors involved in these pathways may represent novel therapeutic targets. Strategies for identifying these transcription factors have included identifying microbiota-responsive *cis*-regulatory regions proximal to differentially regulated genes [53, 145]. The *cis*-regulatory regions contain

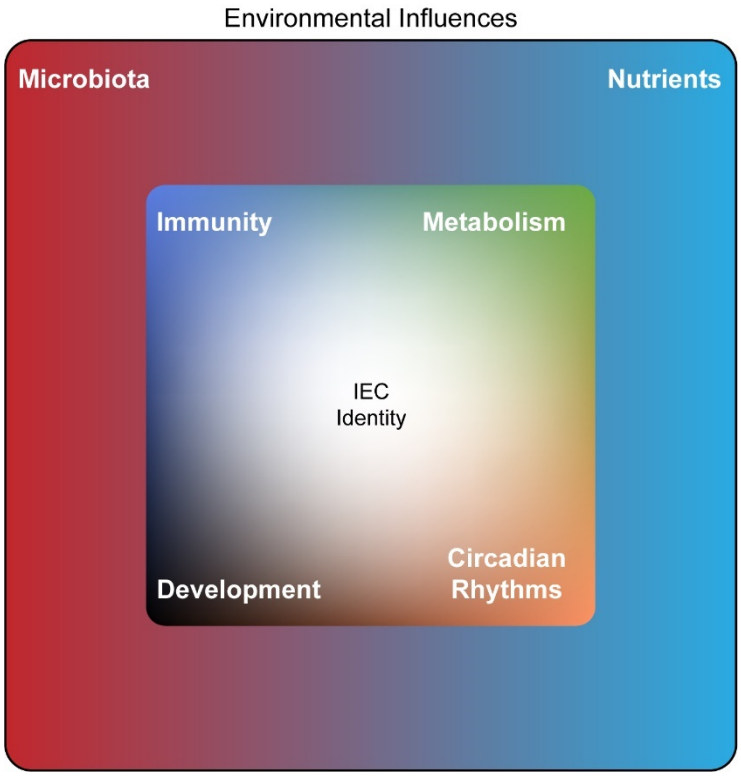


Figure 2.4: IEC identity is determined by the blend of environmental factors and transcriptional programs. The microbiota and nutrition within the intestinal lumen influence each other's molecular makeup. This blend of environmental factors influences transcription programs that similarly blend together to modify the environment and establish an IEC identity.

specific sequences that transcription factors recognize and bind to activate or repress a gene. Other transcription factors have been identified simply by their association with differentially regulated genes. Similarly, genetic analysis and CRR reporter constructs in *in vivo* systems have also provided important insight into transcription factors that mediate the epithelial response [146].

2.4.1 Metabolic programs

Microbiota colonization elicits a strong response in transcription of genes involved in metabolic processes in the epithelia. Following microbiota colonization, the transcriptional programs in the jejunum shift over a 4 day period from oxidative phosphorylation to an anabolic metabolism [120]. Fatty acid mobilization genes, like *Apoc3* and *Cd36*, and genes involved in the processing of fatty acids for energy, like *Cpt1b* and *Pdk4*, remained suppressed by the microbiota even after 30 days post colonization [120]. The microbiota similarly suppress the expression of the transcription factors that regulate these metabolic pathways [53, 119, 120]. These transcription factors include the GATA motif binding transcription factors GATA4 and GATA5 and nuclear receptors transcription factors like PPARA, CAR, LXR, and MCR.

Enterocytes detect nutrient availability through metabolite sensors that activate and deactivate transcription networks that impact cellular and systemic processes. Nuclear receptors represent a key set of transcription factors that are expressed in the enterocytes and serve as elegant detection and response modules. Several nuclear receptors bind metabolites and metabolic intermediates and initiate or suppress transcription programs [147]. Nuclear receptors are animal innovations, first emerging in sponges [148]. This superfamily of transcription factors includes ~50 members. The most studied of these transcription factors are the steroid-binding receptors; however, the metabolite-binding receptors have recently received much more attention [147]. Bile acids serve as an

activating ligand for some nuclear receptors, including FXR [149]. Microbiota regulate FXR activity in the ileum, the site of bile acid uptake in the small intestine and as a consequence regulates systemic metabolic processes like fat storage [113, 150-153]. Furthermore, the microbiota modify primary bile acids, adding different chemical groups to the structure of these molecules to generate secondary bile acids [113, 154]. Different bile acid species bind and activate FXR at varying capacities [149]. These data permit the hypothesis that the microbiota may selectively tune FXR activity by modulating the levels of bile acid species. This mechanism of transcriptional regulation represents an attractive model for how the microbiota mediate epithelial gene transcription programs. The microbiota also modify or produce other ligands for nuclear receptors. PPARG transcriptional activity is enhanced by SCFAs, which are generated by the microbiota in the colon [155]. Similarly, the microbiota express enzymes that modify long-chain fatty acids which might serve as ligands for other nuclear receptors like HNF4A [129].

One of the best studied genes suppressed by the microbiota in small intestine is *Angptl4*, which inhibits lipoprotein lipase activity and blocks serum triglyceride clearance [119, 120, 156]. Microbial suppression of intestinal *Angptl4* leads to decreased serum triglyceride levels and increased epididymal fat pads in murine models. Interestingly, the probiotic bacterium *Lactobacillus paracasei* induces expression of Angptl4 in colon cell lines and reduces fat storage, indicating this bacterium may have the capacity to directly regulate the transcription of this hormone peptide [157]. This gene is regulated by the nuclear receptors PPARG in IECs of the colon [158], and Glucocorticoid receptor in hepatocytes and adipocytes [159, 160]. Perhaps microbial regulation of these transcription factors or another nuclear receptor in the small intestine mediates microbial control of *Angptl4*.

The primary location of SCFA production occurs in the lumen of the mammalian colon that harbors anaerobic bacteria that thrive in this low O_2 concentration environment [161]. These SCFA are the primary source of energy for colonic epithelia. Consumption of

these SCFA by the epithelia decreases O_2 concentrations within the cell which promotes stable activation of the transcription factor Hypoxia induced factor (HIF) [162]. HIF activity results in transcriptional activation of defensing-b, an antimicrobial peptide, MUC3 and ITF, which all help maintain epithelial barrier function [163-165]. HIF similarly induces transcription of the antimicrobial peptide LL77 which protects the host the opportunistic pathogen *Candida albicans .* However, it remains unknown if hypoxic environment induces HIF transactivation of LL77 [166].

Colonization of GF animals with a microbiota leads to increased energy harvest [167, 168] and changes in metabolic homeostasis including decreased AMPK activity in skeletal muscle and liver [169]. Low-dose penicillin treatment similarly rewires metabolic transcriptional programs and persists long after microbiota communities recover from penicillin treatment [170, 171]. Together, these data indicate reorganization of the luminal environment through microbial colonization or antibiotic treatment alter metabolic transcriptional programs indirectly through induced alterations in energy homeostasis. In *Drosophila*, reorientation of the metabolic processes in the intestinal epithelia in response to microbiota colonization may be in part mediated by NF-κB signaling, a transcription factor that regulates immune/inflammation programs [172]. Maintaining an adaptive transcriptional program provides the flexibility the epithelia need in order to respond to fluctuations in the nutrition in the intestinal lumen and maintain energy homeostasis.

2.4.2 Immune/inflammation programs

Epithelial cells are equipped with a variety of signaling receptors called Pattern Recognition Receptors (PRRs) which detect Pathogen-associated molecular patterns (PAMPs), Damage-associated molecular patterns (DAMPs), or Microbial-associated molecular patterns (MAMPs); These three classes of molecules are derived from microbes in the environment (MAMPs or PAMPs) or damaged host cells (DAMPs) [173]. Upon

detection of these molecular patterns, PRRs induce innate immune responses which protect the host from infection. Two commonly studied PRRs expressed in the small intestinal epithelia are Tol-like receptors (TLRs) and NOD-like receptors (NLRs). TLRs reside on the cell surface (both basolateral and apical membranes) and promote homeostasis with the microbiota [174]. NOD2 polymorphisms are associated with impaired innate immunity, reduced expression of antimicrobial peptides and deregulation of immune tolerance to the microbiota [175-179]. Loss of TLR signaling can result in impaired response to infection and closer epithelial association with microbiota [180, 181]. Conversely, aberrant activation of these PRRs can similarly result in inflammatory bowel diseases, representing the important balance of inflammation that must be maintained to prevent disease [177, 182]. TLRs bind a specific type of molecular pattern; for instance, TLR4 bind lipopolysaccharide (LPS), TLR5 binds flagellin, and TLR1, TLR2 and TLR5 bind lipopeptides [183]. TLR signaling includes two types of signaling cascades: those that require the adapter protein MyD88, and those that require the adaptor protein TRIF [184-187]. Mouse models of infection have demonstrated *Myd88* is required for protective innate immune responses [188]. Signal transduction through these surface receptors results in activation of the transcription factors NF-κB and JUN which regulate transcription of genes involved innate immunity and proinflammatory cytokines [173, 189-191].

Signaling through MyD88 dependent pathways induces transcription of “early-phase” activation of the transcription factor complex NF-κB which promotes transcription of the innate immune response and proinflammatory cytokines [190]. The MyD88 independent pathway induces “late-phase” activation of NF-κB which promotes transcription of interferon genes which may induce caspase activation and apoptosis [192]. NF-κB represents a central factor that maintains the appropriate balance of pro- and anti-inflammatory programs in the intestinal epithelia. In macrophages, NF-κB has been shown to regulate the transcription of IL-1b and TNF-alpha, two proinflammatory cytokines that can induce

apoptosis [193]. NF-κB activity promotes transcription of serum amyloid a (*saa*) and stimulates neutrophil migration to the intestine in the zebrafish [194]. NF-κB activity similarly regulates beta-defensin transcription, a primary antimicrobial peptide used as a defense mechanism to maintain mucosal homeostasis with the microbiota. Some bacterial species suppress NF-κB trans-activity and thereby evade these antimicrobial defense mechanisms [195]. In the *Drosophila* intestine, the transcription factor caudal directs NF-κB activity toward just a subset of NF-κB targets including antimicrobial peptides [196]. Suppression of caudal activity in the presence of a *Drosophila* pathogen results in intestinal epithelial apoptosis, impaired epithelial barrier function and an epithelium that resembles inflammatory bowel disease due to reduced antimicrobial peptide transcription [196]. Similarly, inhibition of NF-κB activity by deletion of its activating kinase NEMO in mouse IECs results in chronic inflammatory response in intestinal epithelia [197]. Loss of NF-κB activity in IECs resulted in increased apoptosis, reduced expression of antimicrobial peptides and increased bacterial infiltration into the mucosa. In macrophages, NF-κB suppresses inflammation by activating transcription of the antiapoptotic proteins *PAI-2* and *Bcl-XL* which inhibit activity of caspases. NF-κB also down regulates itself by participating in an inflammation negative feedback loop where it activates transcription of its own suppressor, *IkBa* and the anti-inflammatory protein *Tnfiap3 [193, 198]*.

Microbial stimulation may result in MAP kinase signal transduction and the activation of the transcription factor ATF2. ATF2 forms homodimers or heterodimers with c-JUN [199] and binds to the promoter and activates transcription of DUOX [200]. In the *Drosophila* intestine, this transcriptional activity has been shown to induce cell proliferation (see below in Developmental programs). Transcription of reactive-oxygen species (ROS) generating enzymes, like DUOX or NOS2, represents another innate immune response [201]. The transmembrane proteins DUOX1 and DUOX2 generate ROS in the intestinal lumen which promotes mucosal defense against microbes [202]. A c-JUN/ATF2/IRF3 complex has been

captured on an enhancer that regulates IFN-beta, an important cytokine in viral defense [199].

Janus Kinase-Signal Transduction And Transcription (JAK-STAT) signaling also plays important roles in the response to the microbiota. Activation of the JAK-STAT signaling pathway in the intestine by a cytokine results in phosphorylation of STAT transcription factors by JAKs. Once phosphorylated, STATs homo or heterodimer with other STAT family members and translocate to the nucleus to bind to CRRs and activate transcription of innate immune genes like NOS2 [203]. Upon microbiota colonization and infection of pathogenic bacteria, Paneth cells become filled with granulae that are filled with lysozyme and antimicrobial peptides [121]. These antimicrobial peptides include alpha-defensins and the C-type lectins, like RegIIIg, both of which may disrupt microbial membrane integrity as a mechanism to kill bacteria [204]. Symbiotic bacteria induce expression of RegIIIg in Paneth cells [205], but the transcription factors that mediate its activity in this cell type remains unknown. In the lung and gastric epithelia, STAT3 binds the promoter of RegIIIg and drives transcription of the gene [204, 206]. Microbiota colonization similarly induces *Stat3* expression, so perhaps the same transcriptional mechanisms are shared between these other epithelial cell types and the intestinal epithelia. The transcription factor TCF-4 mediates transcription of the alpha-defensins in Paneth cells, indicating the microbiota may impact TCF-4 activity. Indeed, microbiota mediate beta-catenin/Wnt signaling (see below in Developmental programs) and cellular gradients of these pathways are linked to TCF-4 activity [207].

2.4.4 Developmental programs

Following microbial colonization, an intense restructuring of the small intestinal mucosa occurs that results in decreases in villus length and increases in crypt depth [120, 121]. Microbiota colonization results in transient repression of transcripts involved in Notch

signaling and a rapid expansion of goblet cells [120, 121]. Microbiota colonization also results in sustained suppression of *Neurog3* transcription [120], the transcription factor that drives the enteroendocrine cell differentiation. Indeed, microbiota colonization results in reduced EEC populations as determined by *Sox9:GFP*^{high} transgenic mice [38]. Similarly, microbiota colonization also induces an expansion of mature goblet cells in rats [208]. Microbiota colonization similarly promotes goblet cell differentiation in zebrafish [209]. Since microbiota colonization results in changes to host metabolic homeostasis [156, 210], perhaps microbiota influence goblet cell differentiation through the activation of the transcription factor NFAT5 which suppresses the metabolic regulator mTORC1 and suppresses Notch signaling, which results in an expansion of goblet cells and MUC2 expression [211, 212].

Experiments using larval zebrafish have demonstrated that a dominant member of the zebrafish microbial community secretes a factor that promotes proliferation through activation of the beta-catenin/Tcf4 transcription complex. Furthermore, these experiments demonstrated that the innate immune response, but not inflammation, mediated the microbial induction of IEC proliferation [136]. In drosophila, JAK-STAT signal transduction mediate an oxidative burst that drives epithelial renewal in response to the microbiota and infection [213]. Similarly, In cell culture studies, STAT3 regulates progression through the cell cycle through upregulation of cyclins D2, D3 and A, as well as *Cdc25a* and down regulates cell cycle regulators p21 and p27 [214]. Conversely, antibiotic treatment results in cell cycle arrest in colon epithelial cells [215], suggesting that the loss of microbial communities impairs epithelial proliferation.

Several proinflammatory cytokines and interferons are capable of inducing apoptosis including TGFb, TNFa, and IL-1b [193, 216]. Microbial induction of the proinflammatory cytokines and chemoattractant TNFa results in epithelial apoptosis and increased cell shedding [217]. ETS-factor transcription factors can bind the promoter of this gene and

activate its transcription [218]. ETS-factors represent a novel set of transcription factors that have been implicated in mediating epithelial response to the microbiota [53]. However, little is known about either genomic or molecular mechanisms that mediate its response to microbial stimuli.

2.4.5 Circadian rhythms

Feeding and the circadian rhythms in the intestinal epithelia are tightly connected [133]. Circadian rhythms coordinate several intestinal epithelial functions including nutrient absorption, nutrient trafficking and cell proliferation [219]. Microbiota colonization results in transcriptional suppression of several key circadian rhythm transcription factors including *Arntl2*, *Per1*, *Per2* and *Cry1* in the jejunum. This suppression persists over several weeks and does not appear to recover after the epithelial immune response reaches homeostasis [120]. Since this expression pattern resembles the metabolic reorientation, perhaps the fluctuations in energy homeostasis upon colonization regulate transcription of these circadian rhythm genes [120].

Surprisingly, antiphasic and oscillating expressions of the nuclear receptors RORa and ReverbA coordinate a rhythmic pattern of TLR expression. This rhythmic pattern of TLR expression translates an arrhythmic signal from the microbiota to oscillating signal resulting in a circadian rhythm output of AP-1 and NFkB signaling [131]. The diurnal oscillations of TLR signaling may direct the microbiota rhythmic oscillations which undergo phasic changes to community composition and metabolic function [141, 220]. These diurnal oscillations impact histone modifications at oscillating genes and maintain intestinal homeostasis [141, 220].

2.5 Loss of intestinal identity and inflammatory bowel diseases

Prolonged loss of the homeostasis between the intestinal epithelia and the luminal environment can result in inflammatory bowel diseases [221]. Interactions between human genetics the microbiota and the diet have been implicated in Crohn's disease (CD) and ulcerative colitis (UC) [221-224]. Indeed, dietary interventions are early steps used to treat IBDs [225-227] and fecal (microbiota) transplantations can ameliorate IBD symptoms [228, 229]. However, a variety of genetic and genomic factors within the human intestinal epithelia also control disease pathology and progression [2, 230-234]. Indeed, dysregulation of the acetylome by Hdac1 and Hdac2 disrupts epithelial homeostasis [235]. Furthermore, the intestinal epithelia require Hdac3 for maintenance in the presence of the microbiota [236]. Interestingly, the transcriptional pathways that respond to microbiota colonization overlap the transcriptional pathways that are dysregulated in IBDs [40, 237, 238]. Another fascinating aspect of inflammatory bowel diseases is the aberrant loss of epithelial identity. Previous transcriptome analysis from patients with ileal Crohn's disease and colonic Crohn's diseases indicate a subset of these patients have tissue identities that correspond to the wrong intestinal segment [2]. That is, some transcriptomes from colonic Crohn's disease biopsies identified more with healthy ileal transcription programs than colonic programs. Similarly, ileal transcriptomes from ileal Crohn's disease biopsies identified more with healthy colonic programs. However, the tissue complexities in human biopsy samples limit the conclusions about the role of how specific cell types contribute to disease progression.

Transcriptome analysis of patients with UC, ileal CD (iCD) and colonic CD (cCD) revealed genes that are differentially regulated in IBD compared to healthy patients. Analysis of these transcriptomes identified 5 transcription factors that serve as central regulators of the genes differentially expressed in all three types of IBD. Two of these transcription factors, NFkB and STAT1, activate transcription of genes differentially upregulated in IBD. These TFs in the context of IBD have been discussed extensively in

other reviews [239-241]. The three transcription factors that mediate expression of genes commonly downregulated in IBD are excitingly nuclear receptors: HNF4A, PPARG, and NR3C1 (GR). Despite all three TFs being implicated in IBD pathology, their roles in disease progression remains unknown. Furthermore, the genes consistently downregulated in IBD are genes involved in metabolic pathways (ex. oxidative reduction, fatty acid trafficking, intestinal absorption) and how suppression of these pathways impacts impact chronic inflammation remains relatively unstudied. Furthermore, other studies have indicated that inflammatory programs and metabolic transcription programs are mutually exclusive [132]; however, underlying causes remain unclear.

The three transcription factors that regulate genes that are suppressed in all three types of IBD are all Nuclear receptors. Aside from these three nuclear receptors, several other nuclear receptors have been implicated in UC or CD or both [242], indicating the importance these direct-environment-sensing transcription factors play in intestinal homeostasis. Several nuclear receptors have been shown to suppress inflammation in multiple tissues: Estrogen Receptor [243], Glucocorticoid Receptor , Vitamin D Receptor [244], PPARs [245, 246], LXR [247], HNF4A [248-251]. GR, PPAR, and LXR work synergistically to suppress TLR signaling in macrophages [252]. Perhaps these nuclear receptors and others function similarly in intestinal epithelia to suppress inflammation. To support this hypothesis, loss of FXR function in mice increases susceptibility to DSS and TNBS induced colitis [253] Indeed, FXR has been shown to modulate proinflammatory responses by forming a repressor complex with NCOR on the NF-κB response element on the IL-1B promoter [254]. Similarly, LXR activation suppressed the expression of the pro-inflammatory marker TNFa in the colon following DSS treatment. LXR deletion resulted in enhanced migration of immune cells to a DSS damaged colon [255].

Three independent GWAS studies have identified variants of the HNF4A risk loci for CD and UC [232, 233, 256] and another study indicated HNF4G may also be a risk locus for

UC [257]. Importantly, differentially activate enhancers in IBD have predicted HNF4A binding sites [230, 231]. Furthermore, HNF4A expression is reduced in biopsies from UC and IBD patients [258]. IEC-specific knockout of mouse *Hnf4a* results in spontaneous intestinal inflammation, resembling human IBD [251, 259]. The IEC-specific knockout mice of *Hnf4a* have reduced lysozyme staining in the granulae of Paneth cells, but unaltered Paneth cell number compared to WT [89]. These data indicate Paneth cells in the mutant animals produce fewer antimicrobial peptides. Therefore, one possible reason for spontaneous colitis in IEC-specific mutants could be an overgrowth of pathogenic bacteria. Similarly, DSS-induced colitis and inflammation suppresses HNF4A nuclear localization and expression [260]. In the liver, knockdown of HNF4A expression induces a proinflammatory feedback circuit that continues to repress HNF4A expression and upregulate proinflammatory genes [250]. Indeed, HNF4A activates the transcription of miR-124 in liver cells [250]. mir-124 has reduced expression in biopsies from pediatric CD patients compared to healthy patients and may protect from CD by silencing STAT3 [261]. Together, these studies suggest HNF4A protects against inflammation and IBDs. Therefore, HNF4A target genes may also protect from IBD. If HNF4A regulates mir-124 in the intestine like it does in the liver, then this scheme provides one possible way HNF4A protects from IBD. Similarly, biopsies from IBD patients demonstrate decreased APOA1 expression, a target of HNF4A. Furthermore, injection of APOA1 mimetic peptide rescues experimental colitis [262] and induces tissue repair in endothelial cells through the Akt/AMPK/eNOS pathways [263].

Genetic variants at human *PPARG* are associated with increased risk for both UC and CD [242, 264, 265]. UC is associated with decreased PPARG expression [266]. Although the exact mechanism in IBDs remains unknown, PPARG has been shown to suppress TLR signaling by exporting a subunit of NF-κB, RelA, out of the nucleus and inhibiting transcription of proinflammatory genes [267]. Induction of PPARG activity by chemical ligands, rosiglitazone and troglitazne, suppresses colonic inflammation associated

with DSS administration in mice [268, 269]. These studies provide insight into how dietary interventions may help alleviate IBD symptoms. Both PPARG and HNF4A bind nutrients like long chain fatty acids and short chain fatty acids and these dietary molecules have been used to relieve IBD-associated inflammation [270]. For example, conjugated linoleic acid supplementation suppresses the immune response of patients with Crohn's disease [271]. Chemical models of inflammation in zebrafish indicate linoleic acid, a poly-unsaturated fatty acid and the endogenous ligand of HNF4A, is protective of inflammation. This result contrasts with treatment with two other fatty acids, palmitoleic acid and palmitic acid, which exacerbated the inflammatory response [272].

Nuclear receptors represent an attractive class of transcription factors for therapeutic design. Activation of glucocorticoid receptor (GR, NRC31) suppresses inflammation by activating transcription of anti-inflammatory genes like IL-10 [273]. GR also inhibits transcription of the proinflammatory cytokine IL-11 [274]. In may perform these anti-inflammatory roles by suppressing AP-1 and NF-κB transcriptional activity. Indeed, glucocorticoids suppress NF-κB and AP-1 activity and increased GR expression is associated with decreased NF-KB activity [275, 276]. Genetic variants at human *NR3C1* are associated with risk for CD [277],but genetic variation may have more important role in choice of therapy. Glucocorticoid treatment represents a cornerstone therapy for inflammatory conditions [278]. As such, it remains an obvious choice for inflammatory bowel disease treatment. However, Genetic variants at *NR3C1* are associated with hormone resistant therapies in UC and CD [279, 280]. Indeed, augmented expression of GR-beta, a ligand binding isoform of GR and a putative dominant negative regulator of GR activity, is associated with unresponsive UC to glucocorticoid treatment [281]. Both villus and crypt epithelial cells express GR, however, aside from maintaining ionic gradients across the epithelia [282, 283], its role in maintaining the intestinal epithelia remains relatively unstudied.

2.6 Chapter Conclusion:

Cells must balance between “flexible” and “rigid” transcription programs to maintain both cellular identity and sensitivity to the environment. No cellular environment may be more challenging to maintain this balance than the intestinal lumen that interfaces with the intestinal epithelia. The epithelia tune its transcription programs to maintain homeostasis with the microbiota. Dysregulation of these transcription programs can promote the onset of human diseases, like cancers and inflammatory bowel diseases [221, 284]. However, these diseases represent extreme circumstances and do not reflect the normal symbiotic relationships that have been maintained between animals and their microorganisms for over 650 million years. Above, I have provided individual stories of how particular transcription factors mediate the expression of a handful of genes and how these transcription factors function to maintain epithelial identity. However, microbiota colonization is associated with the induction and suppression of hundreds of genes in the epithelia. And to date, no study has identified how the microbiota tune entire regulatory networks. Therefore, there remains a significant gap in understanding how the host normally perceives and responds to the microbiota. We do not have a strong understanding of how the “rigid” transcriptional programs overlap with these “flexible” transcription programs that permit sensitivities to the intestinal environment. In the next chapter, I will provide the first genome wide evidence of how microbiota colonization impacts histone modifications and transcription factor binding to regulate host gene expression.

CHAPTER 3: MICROBIOTA REGULATE INTESTINAL EPITHELAIL GENE EXPRESSION BY SUPPRESSING THE TRANSCRIPTION FACTOR HEPATOCYTE NUCLEAR FACTOR 4 ALPHA

3.1 Overview

Microbiota influence diverse aspects of intestinal physiology and disease in part by controlling tissue-specific transcription of host genes. However, host genomic mechanisms mediating microbial control of intestinal gene expression are poorly understood. Hepatocyte nuclear factor 4 (Hnf4) is the most ancient family of nuclear receptor transcription factors with important roles in human metabolic and inflammatory bowel diseases, but a role in host response to microbes is unknown. Using an unbiased screening strategy, we found that zebrafish Hnf4a specifically binds and activates a microbiota-suppressed intestinal epithelial transcriptional enhancer. Genetic analysis revealed that zebrafish *hnf4a* activates nearly half of the genes that are suppressed by microbiota, suggesting microbiota negatively regulate Hnf4a. In support, analysis of genomic architecture in mouse intestinal epithelial cells disclosed that microbiota colonization leads to activation or inactivation of hundreds of enhancers along with drastic genome-wide reduction of HNF4A and HNF4G occupancy. Interspecies meta-analysis suggested interactions between HNF4A and microbiota promote gene expression patterns associated with human inflammatory bowel diseases. These results indicate a critical and conserved role for HNF4A in maintaining intestinal homeostasis in response to microbiota.

3.2 Introduction

All animals face the fundamental challenge of building and maintaining diverse tissues while remaining sensitive and responsive to their environment. This is most salient in

the intestinal epithelium which performs important roles in nutrient absorption and barrier function while being constantly exposed to complex microbial communities (microbiota) and nutrients within the intestinal lumen. The presence and composition of microbiota in the intestinal lumen influence diverse aspects of intestinal development and physiology including dietary nutrient metabolism and absorption, intestinal epithelial renewal, and edification of the host immune system. Abnormal host-microbiota interactions are strongly implicated in the pathogenesis of inflammatory bowel diseases (IBD), including Crohn's disease (CD) and ulcerative colitis (UC) [221]. Studies in mouse and zebrafish models of IBD have established that impaired intestinal epithelial cell (IEC) responses to microbiota are a key aspect of disease progression [223, 285, 286]. Improved understanding of the molecular mechanisms by which microbiota evoke host responses in the intestinal epithelium can be expected to lead to new strategies for preventing or treating IBD and other microbiota-associated diseases.

The ability of IEC to maintain their physiologic functions and respond appropriately to microbial stimuli is facilitated through regulation of gene transcription. Genome-wide comparison of transcript levels in intestinal tissue or isolated IEC from mice reared in the absence of microbes (germ-free or GF) to those colonized with a microbiota (conventionalized or CV) have revealed hundreds of genes that have significantly increased or decreased mRNA levels following microbiota colonization [53]. Interestingly, many mouse genes that are transcriptionally regulated by microbiota have zebrafish homologs that are similarly responsive, suggesting the existence of evolutionarily-conserved regulatory mechanisms [137]. For example, the protein hormone Angiopoetin-like 4 (ANGPTL4, also called FIAF) is encoded by a single ortholog in the mouse and zebrafish genomes, and microbiota colonization results in significant reductions in transcript levels in the small intestinal epithelium of both host species [145, 156]. Whereas these impacts of microbiota on host IEC transcriptomes and their downstream consequences have been extensively

documented, the upstream transcriptional regulatory mechanisms remain poorly understood.

Specification and tuning of gene transcription proceeds in part through interactions between transcription factors (TFs) and their sequence-specific binding to *cis*-regulatory DNA. *Cis*-regulatory regions (CRRs) harbor binding sites for multiple activating or repressing TFs and are generally associated with nucleosome depletion and specific post-translational modifications of histone proteins within adjacent nucleosomes when acting as poised (H3K4me1) or active (H3K27ac) enhancers [21]. Antibiotic administration can impact transcript levels and histone modifications in IECs [141], however it's unclear if these changes are indirect effects caused by alterations to microbiota composition, direct effects of the antibiotic on host cells, or by the effects of remaining antibiotic-resistant microbiota [287]. Previous studies have shown that histone deacetylase 3 is required in IECs to maintain intestinal homeostasis in the presence of microbiota [236], and that overall histone acetylation and methylation in the intestine is altered by microbiota colonization [142]. However, the direct and specific effects of the microbiota on host CRRs and subsequent transcriptional responses in IECs remain unknown.

Our previous studies predicted key roles for one or more nuclear receptor TFs in microbial down regulation of IEC gene expression [53], but the specific TF(s) were not identified. Nuclear receptors are ideal candidate TFs for integrating microbe-derived signals, since for many their transcriptional activity can be positively or negatively regulated by the binding of metabolic or hormonal ligands [288]. However, the roles of nuclear receptors in host responses remain poorly understood, and no previous study has defined the impact of microbiota on nuclear receptor DNA binding. Nuclear receptors are a metazoan innovation. The earliest animals encoded a single nuclear receptor orthologous to Hepatocyte nuclear factor 4 (HNF4; nuclear receptor subfamily NR2A) [148]. Despite subsequent duplication and diversification, distinct HNF4 TFs remain encoded in extant animals including mammals

(HNF4A, HNF4G) and fishes (Hnf4a, Hnf4b, Hnf4g) (Supplemental Figure 3.S1G). HNF4A serves particularly important roles in IECs, where it binds CRRs and activates expression of genes involved in IEC maturation and function [289]. IEC-specific knockout of mouse *Hnf4a* results in spontaneous intestinal inflammation similar to human IBD [251]. In accord, genetic variants at human *HNF4A* are associated with risk for both UC and CD as well as colon cancer [232, 233, 256, 290]. HNF4A is predicted to bind a majority of IBD-linked CRRs and to regulate IBD-linked genes [231, 237]. Similarly, genetic variants near human *HNF4G* have been associated with obesity and CD [257, 291]. Importantly, these diverse roles for HNF4 TFs in host physiology have only been studied in animals colonized with microbiota. Therefore, the role of Hnf4 in host-microbiota interactions and the implications for human IBD remain unknown.

3.3 Results

3.3.1 *hnf4a* is essential for transcriptional activity from a microbiota-suppressed cis-regulatory DNA region

To identify transcriptional regulatory mechanisms underlying microbial control of host gene expression, we took advantage of a previously identified microbiota-responsive CRR termed in3.4 located within the third intron of zebrafish *angptl4* (Figure 3.1A). A GFP reporter construct under control of in3.4 termed *in3.4:cfos:gfp* drives tissue specific expression of GFP in zebrafish IEC and is suppressed by microbiota colonization, recapitulating the microbial suppression of zebrafish *angptl4* [145]. However, the factor(s) that mediate microbial suppression of in3.4 were unknown. Using a yeast one-hybrid (Y1H) assay, we tested the capacity of 150 TFs expressed in the zebrafish digestive system to bind in3.4 (Supplemental Figure 3.S1A,B) and detected an interaction only with *hnf4a*, *hnf4b*, and *hnf4g* (Figure 3.1B). When either of two predicted Hnf4 motifs in in3.4 are mutated, the Hnf4-in3.4 interaction in the Y1H assay and intestinal GFP expression in

in3.4:cfos:gfp zebrafish is strongly reduced (Supplemental Figure 3.S1C-F). Interestingly, while *gata4*, *gata5*, and *gata6* have predicted motifs in in3.4 [145] these TFs did not interact in the Y1H assay. This suggests that HNF4 TFs are capable of binding in3.4 directly and HNF4 binding sites are necessary for directing in3.4-based transcription *in vitro* and in the intestine.

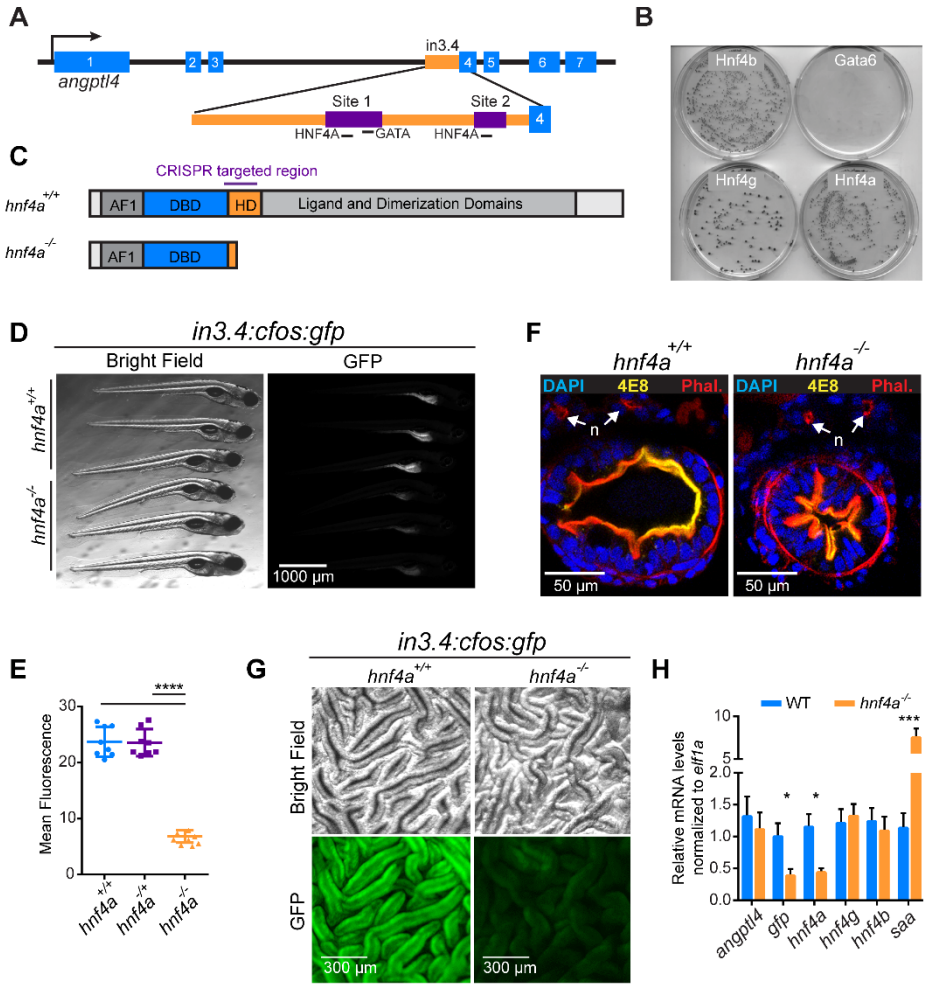


Figure 3.1: Zebrafish *hnf4a* is required for robust in3.4:cfos:gfp activity (A) Schematic of the microbiota-suppressed zebrafish enhancer, in3.4, highlighting the regions required for intestinal activity (purple) which both contain putative HNF4 binding sites (Site 1 and Site 2) [145]. (B) Image of 4 plates from the Y1H assay showing the hnf4 family of transcription factors capable of binding in3.4 and driving expression of the antibiotic resistance reporter gene. (C) Hnf4a+/+ and Hnf4a-/- protein cartoons showing the DNA binding domain (DBD) and hinge domain (HD). We characterized the two with the largest lesions, a -43 deletion in the hinge domain and a +25 insertion in the hinge domain which both result in frame-shift early-stop codons and significantly reduced transcript. (D) Stereofluorescence GFP and bright field microscopy showing representative hnf4a+/+ (top 3) and hnf4a-/- (bottom 3) 6dpf in3.4:cfos:gfp zebrafish. Genotype was blinded and samples were arranged by intensity of GFP fluorescence. (E) GFP fluorescence (mean ± sem) in hnf4a+/+ (n = 8), hnf4a+/- (n = 8) and hnf4a-/- (n = 8) 6dpf in3.4:cfos:gfp zebrafish (Two-tailed t-test: t = 17.84, 16.51, respectively,

df = 14, and **** p < 0.0001). (F) Confocal microscopy showing representative axial cross sections in 6dpf hnf4a+/+ (n = 4) and hnf4a-43/-43 (n = 4) larval zebrafish. 4e8 antibody (yellow) labels the intestinal brush border, DAPI (blue) and phalloidin (red), and nephros (n). (G) Bright field microscopy (top) and stereofluorescence GFP (bottom) for representative hnf4a+/+ (n = 3) (left) and hnf4a-/- (n = 3) (right) dissected intestinal folds from adult in3.4:cfos:gfp zebrafish. (H) Relative mRNA levels (mean ± sem) in hnf4a+/+ (n = 3) and hnf4a-/- (n = 3) adult zebrafish intestinal epithelial cell as measured by qRT-PCR. Two-tailed t-test: t = 0.93, 5.22, 6.56, 10.65, 0.75, 0.94 respectively, df = 4, and * p <0.05, *** p < 0.001). See also Supplemental Figures 3.S1 and 3.S2.

We hypothesized that the *hnf4* transcription factor family is required to mediate microbial suppression of in3.4 activity. Although the Y1H assay demonstrated all 3 zebrafish Hnf4 members are capable of binding in3.4, we concentrated our efforts on understanding the function of *hnf4a* because it is the most highly conserved Hnf4 family member (Supplemental Figure 3.S1G) and has well-documented roles in intestinal physiology [292]. To that end, we generated *hnf4a* mutant zebrafish using the CRISPR/Cas9 system (Figure 3.1C; Supplemental Figure 3.S2A-C,E). Whole-animal *Hnf4a* knockout mice die during early embryogenesis due to failure to develop visceral endoderm [293], but zebrafish and other fishes do not develop that extra-embryonic tissue. We found that zebrafish homozygous for a non-sense mutation in *hnf4a* are viable and survive to sexual maturity (Supplemental Figure 3.S2D) providing new opportunities to study the roles of HNF4A in host-microbiota interactions.

To determine if *hnf4a* is essential for in3.4 activity, we crossed mutant *hnf4a* alleles to the *in3.4:cfos:gfp* transgenic reporter line. GFP expression was significantly reduced in the absence of *hnf4a* suggesting that *hnf4a* activates in3.4 (Figure 3.1D,E,G,H). This loss of GFP expression in *hnf4a*$^{-/-}$ mutants was not associated with overt defects in brush border development or epithelial polarity in larval stages (Figure 3.1F), nor in the establishment of intestinal folds during adult stages (Figure 3.1G). However, intestinal lumen of mutant larvae was reduced in size at 6 days post fertilization (dpf) compared to WT siblings (Figure 3.1F; Supplemental Figure 3.S2F). Together, these data indicate *hnf4a* is essential for robust activity of a microbiota-suppressed CRR. Unlike *in3.4:cfos:gfp*, *angptl4* is expressed in

multiple tissues and cell types [145]. To determine if intestinal *angptl4* expression is dependent on *hnf4a* function, we isolated RNA from IECs from *hnf4a*$^{+/+}$ and *hnf4*$^{-/-}$ adult *in3.4:cfos:gfp* zebrafish and performed qRT-PCR. Adult IECs (AIECs) from *hnf4a*$^{-/-}$ have significant reductions in mRNA for *gfp*, *fabp*2 (a known HNF4A target in human cell lines) [294], and *hnf4a* compared to *hnf4a*$^{+/+}$ controls. However, *angptl4* expression remained unchanged in *hnf4*$^{-/-}$ AIECs compared to WT, suggesting *angptl4* transcript levels in the adult intestine are regulated by additional mechanisms and not solely from in3.4 or Hnf4a (Figure 3.1H). Transcript levels for *hnf4g* and *hnf4b* in *hnf4a*$^{-/-}$AIEC were also unchanged. Together, these results establish that Hnf4a is required for in3.4 activity in IECs and raises the possibility that Hnf4a may have broader roles in mediating host transcriptional and physiological responses to microbiota.

3.3.2 Hnf4a activates transcription of genes that are suppressed upon microbiota colonization

To better define the roles of *hnf4a* in microbiota response and other aspects of digestive physiology, we used RNA-seq to compare mRNA levels from digestive tracts isolated from *hnf4a*$^{+/+}$ and *hnf4a*$^{-/-}$ zebrafish larvae in the presence (CV) or absence of a microbiota (GF; Figure 3.2A). Consistent with our previous studies [137, 146], comparison of wildtype zebrafish reared under CV vs GF conditions revealed differential expression of 598 genes that were enriched for processes such as DNA replication, oxidation reduction, and response to bacterium (Figure 3.2B,D; Supplemental Figure 3.S3D). Strikingly, disruption of the *hnf4a* gene caused gross dysregulation of the transcriptional response to microbiota colonization, with the total number of microbiota responsive genes (CV vs GF) increasing to 2,217. Furthermore, comparison of the *hnf4a* mutant (Mut) vs wild type (WT) genotypes revealed differential expression of many genes in the CV condition (2,741 genes) and GF condition (1,441 genes) that inform a general role for Hnf4a in regulating genes in the

intestinal tract (Figure 3.2D,E). Principal components analysis (Supplemental Figure 3.S3A) and hierarchical clustering (Figure 3.2B) of FPKM values indicated that *hnf4a* genotype had a complex contribution to regulating genes involved in both responses to the microbiota and digestive physiology.

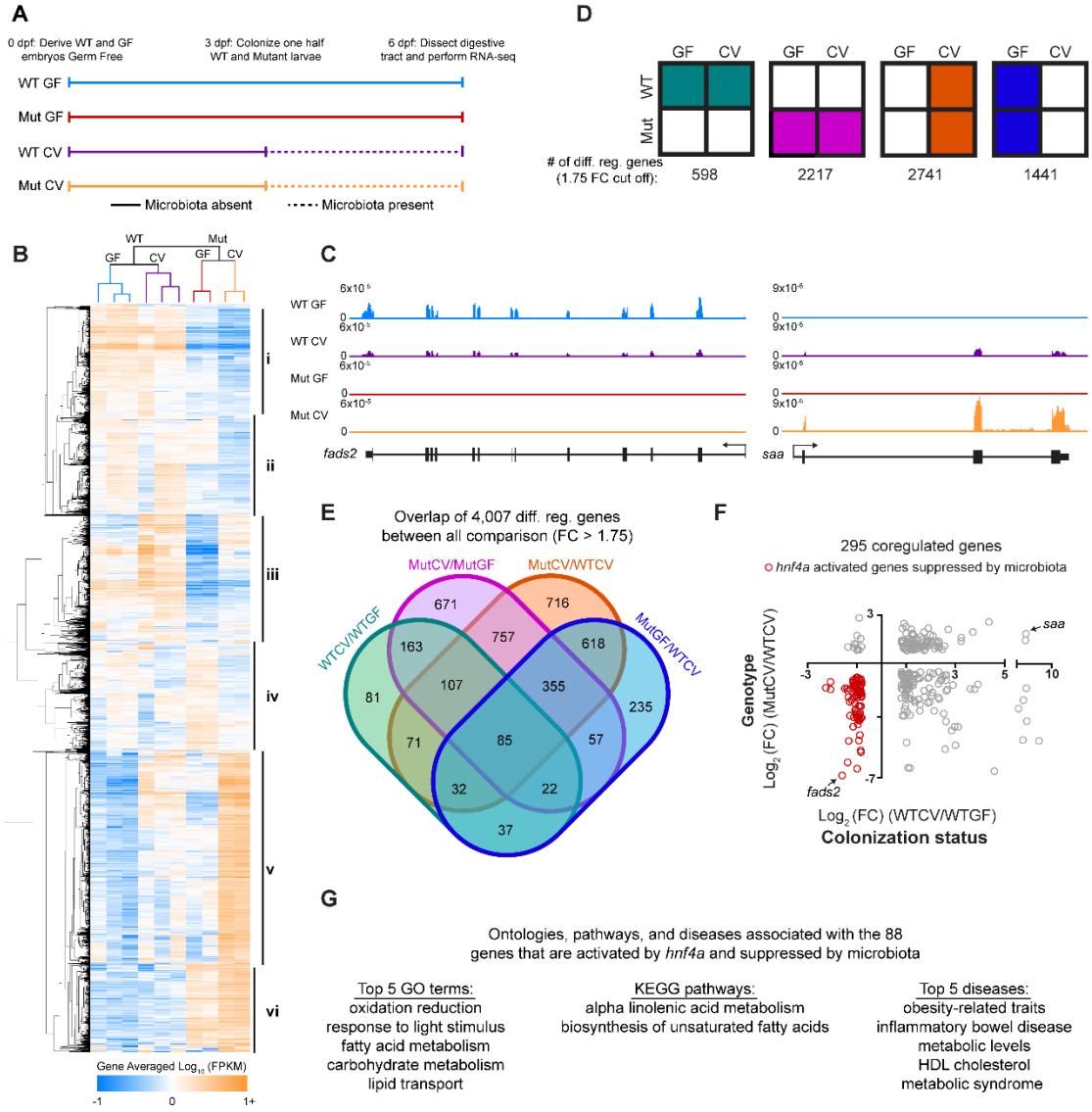


Figure 3.2: Hnf4a activates the majority of coregulated genes that are suppressed by the microbiota. (A) Schematic showing the experimental timeline for zebrafish digestive tract GF and CV hnf4a+/+ and hnf4a-/- RNA-seq experiment (n = 3 for WTCV and WTGF and n = 2 for MutCV and MutGF). (B) Hierarchical relatedness tree and heatmap of differentially regulated genes in mutant and gnotobiotic zebrafish digestive tracts. Gene averaged log10 FPKMs for the biological replicates are represented for each of the 4,007 differentially regulated genes. (C) Representative RNA-seq signal tracks at fatty acid-desaturase 2 (fads2), serum amyloid a (saa) loci. (D) Summary of the total number of differentially expressed genes between indicated conditions (GF and CV) and genotype (WT and hnf4a-/- (Mut)). (E) 4-way Venn diagram showing overlaps between all 4,007 differentially regulated

genes. (F) The 295 coregulated genes were plotted using the log2 (FC) calculated in the WTGF/WTCV comparison (X-axis) and WTCV/MutCV (Y-axis). The 88 out of 98 genes that are activated by hnf4a but suppressed by the microbiota are highlighted (red) and (G) their GO term, KEGG pathway and disease associations are listed. See also Supplemental Figure 3.S3.

Because we found that *hnf4a* activates the microbiota-suppressed intestinal CRR, in3.4, we hypothesized that this may represent a general regulatory paradigm for other microbiota-influenced CRRs and genes across the genome. When we compared the 598 genes that were microbiota responsive in wildtype digestive tracts with the 2,741 genes that *hnf4a* regulates in CV digestive tracts we found these lists shared 295 genes that included *fads2* and *saa*, both of which have human orthologs that are either implicated (*FADS1/2*) or markers (*SAA*) of IBD [224, 295] (Figure 3.2C-F). While loss of Hnf4a could be pleiotropic, strikingly, the overlap between these subsets reveals that a disproportionate 88 of the 98 (~90%) microbiota-suppressed genes are activated by *hnf4a* (Figure 3.2F). These 88 genes represent almost half of all 185 genes suppressed by the microbiota. These data suggest, like its role at in3.4, *hnf4a* plays a critical role in directly activating a large percentage of genes that are suppressed by microbial colonization. This set of *hnf4a*-activated microbiota-suppressed genes is enriched for ontologies and pathways involved in lipid and carbohydrate metabolism, suggesting microbiota might regulate these processes through suppression of Hnf4a (Figure 3.2G). Interestingly, the top 2 diseases associated with this gene set were obesity-related traits and IBD (Figure 3.2G). Based on these results, we hypothesized that Hnf4a DNA binding is lost upon microbial colonization within CRRs associated with microbiota-suppressed genes.

3.3.3 HNF4A binding sites are enriched in promoters near genes associated with microbiota-regulated H3K27ac marks

Previous attempts to identify microbial responsive enhancers genome-wide were complicated by the lack of significant changes in chromatin DNase accessibility between GF and CV IECs from mouse colon and ileum [53]. These previous findings suggested other

chromatin dynamics may be involved in regulating the IEC response to microbiota. We therefore sought to provide a genomic context for understanding how the microbiota alter HNF4A activity and chromatin modifications in IECs by performing RNA-seq, DNase-seq, and ChIP-seq for the enhancer histone modifications H3K4me1 and H3K27ac, and the Hnf4 TF family members HNF4G and HNF4A in CV and GF conditions totaling 35 datasets. We conducted these experiments in jejunal IEC from gnotobiotic mice because: (1) ChIP-grade antibodies for mouse HNF4A and HNF4G are available, (2) the larger organ size in mice provided sufficient numbers of IECs for ChIP-seq experiments, and (3) we speculated that the roles of HNF4A in host response to microbiota may be conserved to mammals. We first performed DNase-seq in jejunal IEC from mice reared GF or colonized for two weeks with a conventional mouse microbiota (CV) to determine the impact of microbiota colonization on chromatin accessibility (Figure 3.3A). In accord with previous studies that that tested for chromatin accessibility in ileal or colonic IECs from GF or CV mice [53], we similarly found no differential DNase hypersensitivity sites (DHSs) in GF or CV jejunum (data not shown, but see Supplemental Figure 3.S4A). These data indicate that gross accessibility changes in chromatin do not underlie the transcription of microbiota-responsive genes in IECs.

To test if other metrics of chromatin utilization were dynamically regulated by microbiota, we performed ChIP-seq from GF and CV mouse jejunal IECs for histone marks H3K4me1 and H3K27ac that are enriched at poised enhancers and active enhancers, respectively (Figure 3.3B). By determining the single-nearest gene TSS within 10kb of the differential histone marks and overlaying these data with our new RNA-seq datasets, we found that regions that gain poised (H3K4me1) and activated (H3K27ac) enhancers upon colonization are associated with genes that have increased transcript levels upon colonization (Figure 3.3C,H-K; Supplemental Figure 3.S4I). Similarly, regions that lose poised and active enhancers upon colonization are associated with microbiota-suppressed genes (Figure 3.3C,G,I,J,L; Supplemental Figure 3.4J). A two-sided Kolmogorov-Smirnov

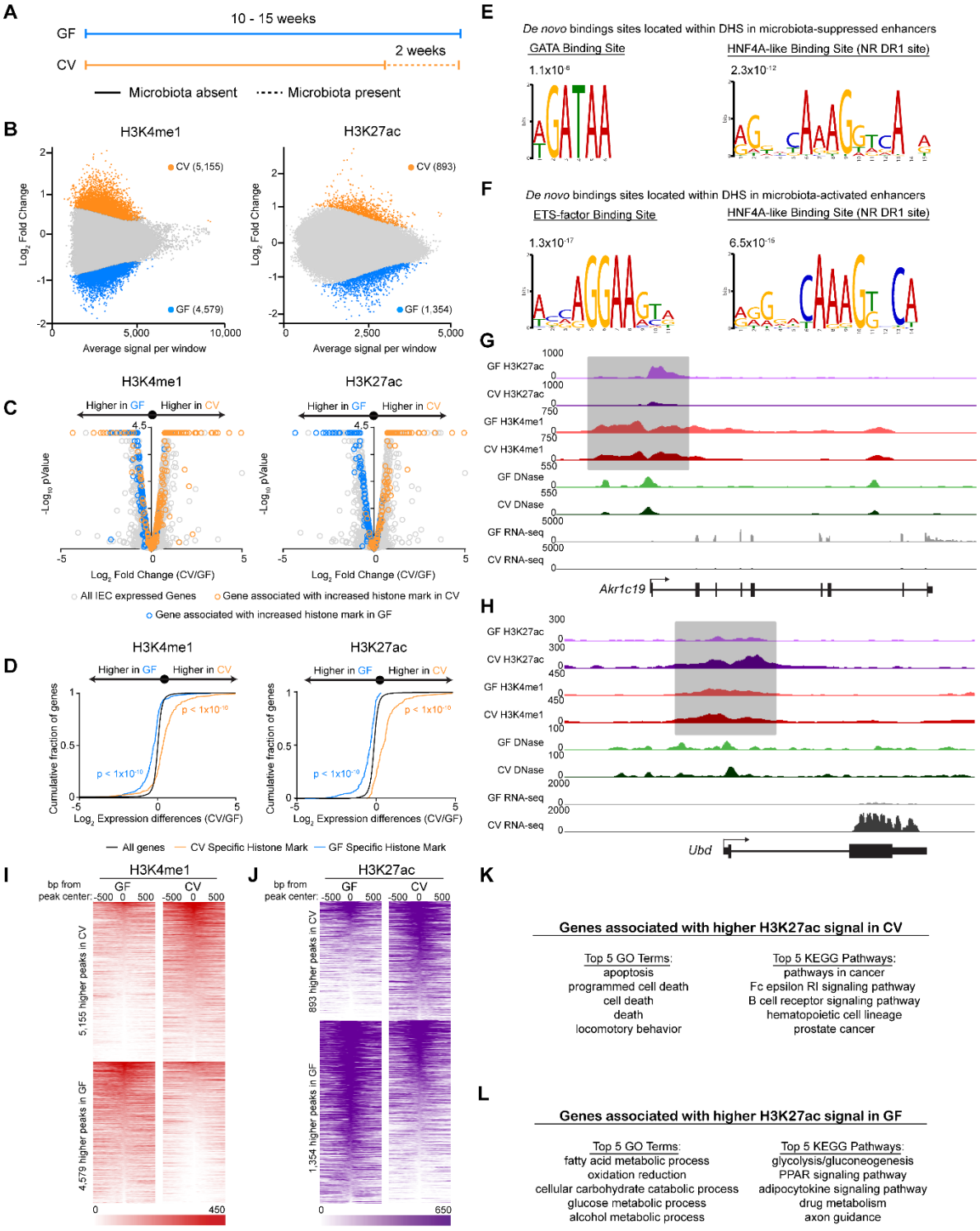


Figure 3.3: Microbiota selectively induce enhancer activity near genes that are upregulated upon microbiota colonization. (A) Schematic showing the gnotobiotic experimental timeline for testing mRNA levels and chromatin architecture in GF and CV. (B) MA plots from DESeq2 analysis (FDR < 0.01) of H3K4me1 (n = 3 per condition) (left) and H3K27ac (n = 2 per condition) (right) ChIP-seq from GF and CV mouse jejunal IECs. Colored dots signify regions significantly enriched for a histone mark in GF (blue) or CV (orange). We found 4,579 unique H3K4me1 and 1,354 unique

H3K27ac peaks in GF and 5,155 unique H3K4me1 and 893 unique H3K27ac peaks in CV. (C) Volcano plots showing pairwise comparison of RNA expression between GF (n = 2) and CV (n = 2) jejunal IECs. Blue and orange dots represent genes associated with a region enriched for H3K4me1 (left) or H3K27ac (right) signal in GF or CV. (D) Two-sided Kolmogorov-Smirnov goodness-of-fit test shows a positive relationship on average between the presence of a region enriched for H3K4me1/H3K27ac signal in a specific colonization state and increased transcript abundance of a neighboring gene in that same colonization state. (E) Top de novo binding site motifs found in DHSs that are flanked by regions enriched with H3K27ac signal in GF (E) or CV (F). Representative ChIP-seq tracks highlighting a microbiota-regulated gene associated with differential histone marks in GF (G) (Akr1c19, Aldo-keto reductase 1c19) or CV (H) (Ubd, Ubiquitin D). Heatmaps showing the average GF and CV H3K4me1 (I) or H3K27ac (J) signal at the 1000 bp flanking differential sites. (K-L) GO terms and KEGG pathways enriched in genes associated with differential H3K27ac sites shown in J. See also Supplemental Figure 3.S4.

goodness-of-fit test shows a positive relationship between differential H3K4me1/H3K27ac region and increased transcript abundance of nearby genes in the same colonization state (Figure 3.3D). Collectively, we identified for the first time a genome-wide map of hundreds of newly identified microbial regulated CRRs, suggesting that microbiota regulation of host genes is mechanistically linked to histone modifications changes more than gross chromatin accessibility changes [53].

We leveraged this novel atlas of microbiota-regulated enhancers and accessible chromatin to determine which TFs are predicted to bind to these regions. An unbiased analysis found that three HNF4A binding site motifs were significantly ($p < 1e-45$, $p < 1e-28$, and $p < 1e-13$) enriched in promoters of genes associated with microbiota-suppressed enhancers (Supplemental Figure 3.S4E), and STAT1 binding site motifs were significantly ($p < 1e-16$) enriched in promoters of genes associated with microbiota-activated enhancers (Supplemental Figure 3.S4F). Interestingly, DHS sites associated with differentially active enhancers were enriched for two different sets of TF binding sites. DHSs flanked by microbiota-inactivated enhancers were enriched for nuclear receptor DR1 sites, which can be recognized by HNF4A [296], and GATA binding sites ($p = 2.3e-12$ and $1.1e-6$ respectively) (Figure 3.3E). DHS sites associated with microbiota-activated enhancers were similarly enriched for the nuclear receptor DR1 binding sites but also for STAT/IRF-like and ETS binding sites ($p = 6.5e-15$ and $1.3e-17$ respectively) (Figure 3.3F). These data suggest

that nuclear receptors like HNF4A may play a central role in IEC responses to microbial colonization.

3.3.4 Microbiota colonization is associated with a reduction in HNF4A and HNF4G cistrome occupancy

To directly evaluate the impact of microbiota on HNF4A activity, we tested the plasticity of the genome wide distribution of HNF4s in response to microbial colonization. HNF4A bound 28,901 and HNF4G bound 21,875 across the genome in GF conditions in jejunal IECs with ~80% of these sites being bound by both TFs. In striking contrast, the number of sites bound by HNF4A and HNF4G in CV conditions was ~10 fold less (Figure 3.4A,B; Supplemental Figure 3.S5A-D). Of the 3,964 HNF4A binding sites detected in CV there were only 267 HNF4A sites that were specific to the CV condition (Supplemental Figure 3.S6A,C). Yet, the genes associated with these HNF4A sites that are retained in CV are enriched for ontologies and pathways fundamental to intestinal epithelial biology (Supplemental Figure 3.S6B). Surprisingly, we found HNF4A sites are equally distributed between genes that are upregulated in both GF and CV conditions (Supplemental Figure 3.S5E). However, we did find that the average CV HNF4A signal strength was significantly increased at HNF4A sites associated with microbiota-induced genes relative to those HNF4A sites associated with microbiota-suppressed genes, suggesting HNF4A may play a limited role in genes upregulated by colonization (Supplemental Figure 3.S6F). In contrast, GF HNF4A ChIP signal was equivalent at HNF4A sites associated with microbiota-suppressed and induced genes (Supplemental Figure 3.S6F). Interestingly, we found that HNF4A sites were significantly correlated with increased H3K27ac, H3K4me1 and DHS signal in GF compared to these same chromatin marks in CV (Supplemental Figure 3.S6G). We do not believe that the reduction of HNF4A binding is the result of chromatin quality in a particular condition since there are genomic locations where GF and CV HNF4A sites appeared to have

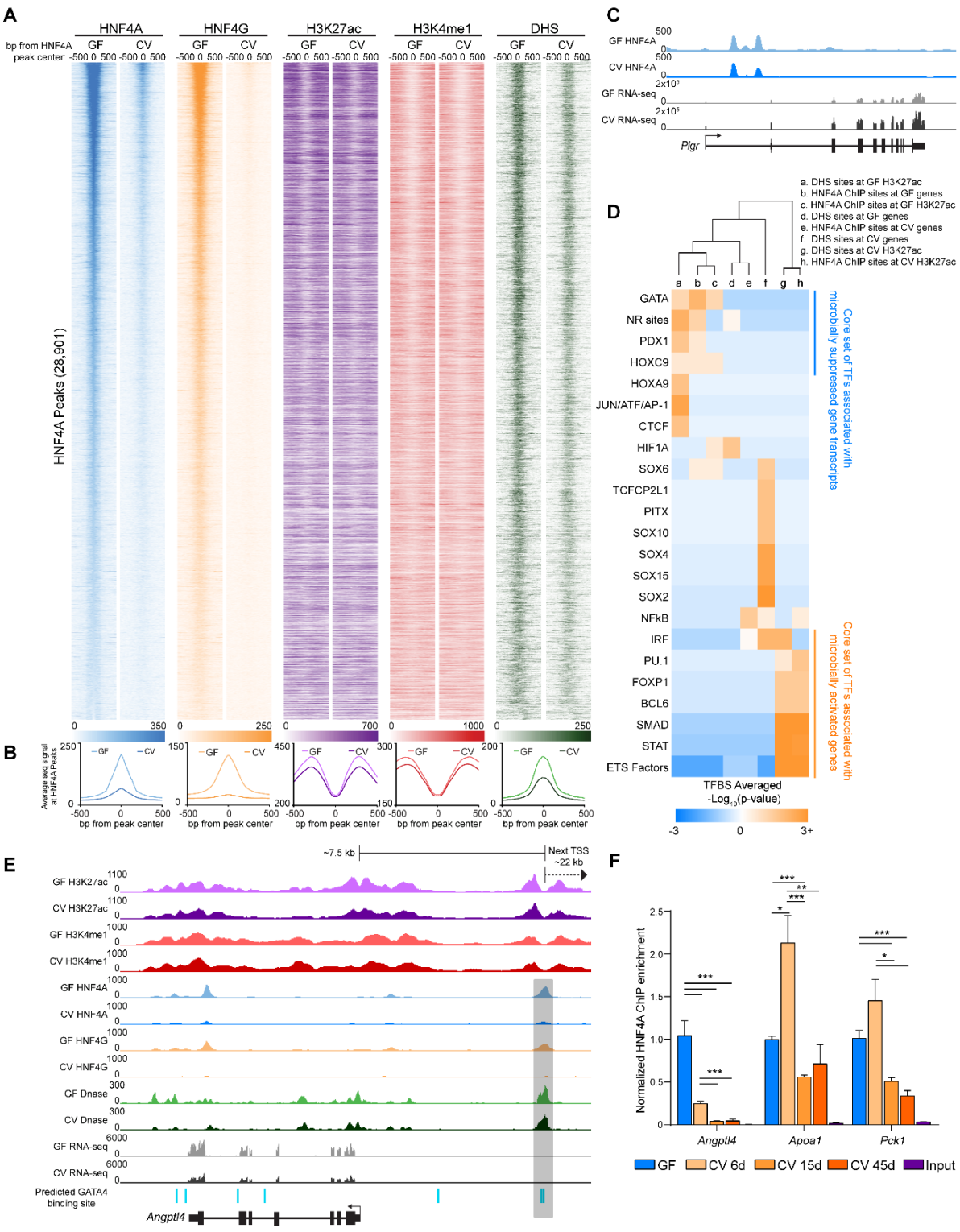


Figure 3.4: Microbiota colonization results in extensive loss of HNF4A DNA binding in IEC
(A) Heatmaps showing the average GF and CV ChIP-seq or DNase-seq signal at the 1000 bp flanking HNF4A sites found in GF.(B) Line plots showing the average GF (light-colored line) and CV (dark-colored line) ChIP-seq and DNase-seq RPKM-normalized signal for the indicated TF, histone mark or DHS at the 1000 bp flanking HNF4A sites found in GF (HNF4A: n = 3 per condition; HNF4G:

n = 4 per condition; H3K27ac n = 2 per condition; H3K4me1: n = 3 per condition; DNase: n = 3 for CV, n = 2 for GF) . (C) Representative signal tracks highlighting a microbiota-induced gene (Pigr, Polymeric immunoglobulin receptor) that is associated with an HNF4A peak with similar signal in both GF and CV jejunal IECs. (D) Heatmap showing the enrichment of TFBS motifs within 50 bp of the DHS or HNF4A peak maxima. (E) Representative signal track at Angptl4 highlighting two GATA4 sites within an HNF4A bound region. (F) Bar graph showing HNF4A ChIP-PCR results at Angptl4, Apoa1 and Pck1 loci from jejunal IECs from mice colonized for 0 (n = 2), 6 (n = 3), 15 (n = 2) and 45 (n = 3) days. Data are relative to the GF condition and normalized against a negative control locus (Neurog1) * $p < 0.5$, ** $p < 0.005$, *** $p < 0.0005$. See also Supplemental Figures 3.S5 and 3.S6.

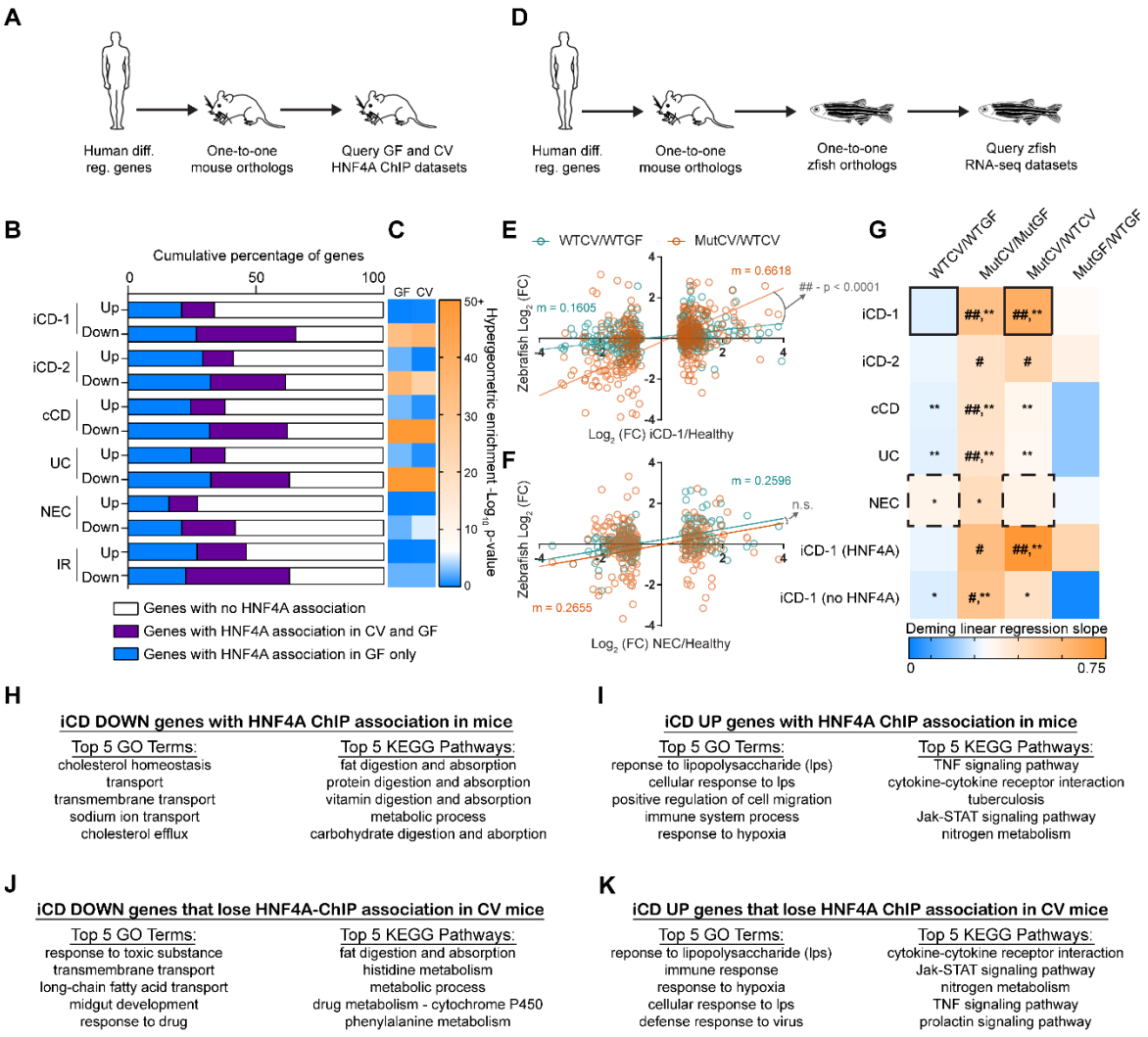
equivalent signal (Figure 3.4C). Furthermore, ChIP enrichment in these IEC preparations for another zinc finger TF, CTCF, was unaffected by microbiota colonization (Supplemental Figure 3.S6D). This indicates that the observed reduction of Hnf4 ChIP signal in CV IECs is a result of microbiota on HNF4 binding, and is not the result of altered ChIP efficiency or sample quality in the different conditions. To test if microbial suppression of HNF4A occupancy is persistent, we performed ChIP-PCR from ex-GF mice that were colonized with microbiota for 6, 15 or 45 days. We found that even after 45 days post-colonization, HNF4A occupancy at binding sites was significantly reduced compared to GF (Figure 3.4F). The dramatic loss of HNF4A and HNF4G DNA binding upon colonization is consistent with HNF4A acting as a potent activator of microbiota-suppressed genes.

We further speculated that certain coregulatory sequence-specific transcription factors may also contribute to regulating transcription with HNF4 at these sites. To explore this possibility, we searched for TF motifs associated with HNF4A ChIP sites and found an enrichment of putative binding sites for TFs known to be involved in small intestinal physiology (GATA and HOXC9) as well as nutrient metabolism (PDX1) at both HNF4A bound regions associated with genes and enhancers suppressed by microbes (Figure 3.4D). We similarly found GATA sites located within an HNF4A-bound CRR near murine *Angptl4* (Figure 3.4E), similar to the coincident HNF4 and GATA motifs in in3.4 [145]. Furthermore, binding sites for TFs known to be involved in cell proliferation and cell death (ETS transcription factor family) are enriched near HNF4A bound regions that intersect microbiota-induced enhancers (Figure 3.4D). Collectively our integrative analyses of these

novel ChIP-seq, DNase-seq, and RNA-seq datasets identifies a core set of putative microbiota-responsive TFs that may interact with HNF4A to mediate microbial control of IEC gene expression. These results suggest HNF4A plays a major role in integrating microbial signals to regulate gene expression, and raise the possibility that this novel microbiota-HNF4A axis might contribute to human disease.

3.3.5 Microbiota-mediated suppression of HNF4A may contribute to gene expression profiles associated with human IBD

Both HNF4A and the intestinal microbiota have been separately implicated in the pathogenesis of the human IBDs Crohn's disease (CD) and ulcerative colitis (UC) [221, 258]. However, a mechanistic link between microbiota and HNF4A in the context of IBD pathogenesis has not been established. Previous transcriptomic studies have identified genes differentially expressed in ileal (iCD) and colonic CD (cCD) and UC [237, 238] biopsies. We queried these human gene lists to identify one-to-one orthologs in mice, and referenced them against our new gnotobiotic mouse jejunal HNF4A ChIP-seq data (Figure 3.5A). Strikingly, the majority of human genes downregulated in each of these IBD datasets have mouse orthologs that are associated with an HNF4A-bound region (Figure 3.5B,C) Focusing on the iCD dataset from the largest of these previous studies [237], we found differential iCD genes associated with HNF4A sites are enriched for distinct ontologies and pathways that are dysregulated in IBD (Figure 3.5H-K). In contrast to IBD, analysis of intestinal transcriptomic datasets from human subjects with necrotizing enterocolitis (NEC) [297] or insulin-resistance (IR) [298] did not reveal strong enrichment of HNF4A-bound regions near downregulated genes (Figure 3.5C). Notably, in each of these CD, UC, NEC, and IR datasets, a greater percentage of downregulated genes were linked to HNFA-bound regions compared to upregulated genes (Figure 3.5B). These data suggest microbiota-dependent and microbiota-independent suppression of HNF4A activity in the intestine might



H

iCD DOWN genes with HNF4A ChIP association in mice

Top 5 GO Terms:	Top 5 KEGG Pathways:
cholesterol homeostasis	fat digestion and absorption
transport	protein digestion and absorption
transmembrane transport	vitamin digestion and absorption
sodium ion transport	metabolic process
cholesterol efflux	carbohydrate digestion and aborption

I

iCD UP genes with HNF4A ChIP association in mice

Top 5 GO Terms:	Top 5 KEGG Pathways:
reponse to lipopolysaccharide (lps)	TNF signaling pathway
cellular response to lps	cytokine-cytokine receptor interaction
positive regulation of cell migration	tuberculosis
immune system process	Jak-STAT signaling pathway
response to hypoxia	nitrogen metabolism

J

iCD DOWN genes that lose HNF4A-ChIP association in CV mice

Top 5 GO Terms:	Top 5 KEGG Pathways:
response to toxic substance	fat digestion and absorption
transmembrane transport	histidine metabolism
long-chain fatty acid transport	metabolic process
midgut development	drug metabolism - cytochrome P450
response to drug	phenylalanine metabolism

K

iCD UP genes that lose HNF4A ChIP association in CV mice

Top 5 GO Terms:	Top 5 KEGG Pathways:
reponse to lipopolysaccharide (lps)	cytokine-cytokine receptor interaction
immune response	Jak-STAT signaling pathway
response to hypoxia	nitrogen metabolism
cellular response to lps	TNF signaling pathway
defense response to virus	prolactin signaling pathway

Figure 3.5. Microbiota suppression of HNF4A activity is highly correlated with genes and intestinal processes suppressed in human IBD and conserved in zebrafish. (A) Flow chart showing the experimental design and filters used to identify IBD or NEC gene orthologs associated with mouse HNF4A ChIP sites. (B) Bar chart showing the proportion of HNF4A associations in GF and CV mouse jejunal IECs near human-to-mouse one-to-one gene orthologs differentially regulated in human pediatric ileal Crohn's Disease (iCD-1), adult iCD (iCD-2), adult colonic Crohn's Disease (cCD), adult ulcerative colitis (UC), neonatal necrotizing enterocolitis (NEC) or insulin-resistance (IR). (C) Heatmap representing the -Log10 (pValue) of the enrichment of GF or CV HNF4A associated genes that are differentially regulated genes in the indicated IBD datasets. Log10 p-values were calculated using a hypergeometric enrichment analysis and converting all HNF4A ChIP associated mouse genes to human orthologs (GF = 5863 genes and CV = 2119 genes). (D) Flow chart showing the experimental design and filters used to identify correlations between gnotobiotic WT or mutant zebrafish gene expression and gene orthologs differentially expressed in human IBD or NEC. Because loss of hnf4a function in zebrafish appeared to more closely resemble iCD signature than cCD or UC, we performed pairwise comparisons of gene orthologs that are (1) differentially regulated in human iCD and (2) have a mouse HNF4A ChIP association. Example of Deming linear regression analysis showing the correlation of Log2 (FC) between WTCV/WTGF (E) or MutCV/WTCV (F) zebrafish and pediatric iCD or NEC. m = slope of the line. (G) Heatmap representing slopes of Deming linear regression lines showing positive correlative relationships between the log2 gene expression fold changes of one-to-one orthologs from human diseases compared to log2 fold

changes in zebrafish WTCV/WTGF, MutCV/MutGF, MutCV/WTCV, and MutGF/WTGF. Hash signs indicate slope of Deming linear regression lines is significantly greater than WTCV/WTGF comparison (#, $p < 0.05$; ##, $p < 0.0001$). Asterisks indicate slope of Deming linear regression line is significantly greater than MutGF/WTGF (*, $p < 0.05$; **, $p < 0.001$). Solid boxes correspond to slope of lines in panel 5D, and dashed boxes correspond to slope of lines in panel 5E. (H-K) The top 5 GO terms and the top 5 KEGG pathways for indicated gene lists.

play an important role in IBD pathologies. To assess if microbiota suppression of HNF4A activity regulates genes differentially expressed in IBD, we queried the published human IBD and NEC gene expression datasets to identify human-mouse-zebrafish one-to-one-to-one orthologs that were differentially expressed in our RNA-seq analysis of gnotobiotic zebrafish *hnf4a* mutants (Figure 3.5D). We found ortholog expression fold changes in human IBD/healthy comparisons most closely resembled the expression fold changes of MutCV/MutGF and MutCV/WTCV (Figure 3.5E-G). Neither the WTCV/WTGF nor the MutGF/WTGF comparisons faithfully recapitulate the expression profiles of IBD/healthy comparisons. This indicates that both the microbiota and loss of *hnf4a* function in zebrafish are necessary to induce a gene expression profile that resembles human IBD. Strikingly, the positive correlation and significant resemblance to the iCD-like gene signatures in the colonized *hnf4a*$^{-/-}$ compared to colonized *hnf4a*$^{+/+}$ zebrafish digestive tracts become even stronger when we limited our analysis to one-to-one orthologs that have an association with an HNF4A bound region in mouse IECs (Figure 3.5G). Together, these results indicate that intestinal suppression of HNF4A target genes is a prevalent feature of human CD and UC, and suggests a model wherein HNF4A maintains transcriptional homeostasis in the presence of a microbiota and protects against an evolutionarily-conserved IBD-like gene expression signature.

3.4 Discussion

Over the course of animal evolution, the intestinal epithelium has served as the primary barrier between animal hosts and the complex microbial communities they harbor.

IECs maintain this barrier and perform their physiological roles in nutrient transport and metabolism through dynamic transcriptional programs. The regulatory mechanisms that orchestrate these transcriptional programs represent potential therapeutic targets for a variety of human intestinal diseases including IBD. Here we discovered that HNF4A activity and its transcriptional network are suppressed by microbiota. HNF4A is the oldest member of the nuclear receptor TF family [148], and our findings in fish and mammals suggest that microbial suppression of HNF4A may be a conserved feature of IEC transcriptional programs present in the common ancestor.

We discovered HNF4A as a microbiota-suppressed transcription factor by demonstrating it specifically binds to a microbiota-suppressed *cis*-regulatory element, in3.4, which is located at the zebrafish gene *angptl4*. This finding combined with our zebrafish RNA-seq data (Fig. 2FG) revealed a broad role for HNF4A in activation of microbially-suppressed transcripts. Though *hnf4a* mutant zebrafish have reduced in3.4 activity in the intestinal epithelium based on transgenic reporter assays, the transcript levels of the endogenous zebrafish *angptl4* gene appears unaffected in both larval digestive tracts and adult IECs. The zebrafish genome encodes two additional HNF4 family members (*hnf4b*, *hnf4g*), and previous studies in mammals have shown *Angptl4* can be regulated by other metabolically-activated nuclear receptors [158, 299]. We hypothesize that loss of HNF4A function may lead to a metabolic imbalance leading to atypical or compensatory activation of other *trans*- and *cis*-factors that control expression of *angptl4* and other genes in the intestine.

Our results suggest new links between HNF4A and microbiota in the context of human IBD. IBD patients, particularly those suffering from Crohn's disease, often present with decreased serum low-density lipoprotein levels and reduced total cholesterol levels compared to healthy individuals [300, 301]. These serum levels are consistent with reduced transcript levels for genes involved in intestinal absorption and transport of lipid and

cholesterol in ileal and colonic biopsies from UC and CD patients [237, 238]. Transcription factors, including nuclear receptors like HNF4A and FXR, are known to regulate bile acid production, lipid and cholesterol absorption and have already been implicated in IBD [258, 302]. Previous studies have shown that some IBD-associated H3K27ac activated regions that also overlap with an IBD-associated SNP contain HNF4A binding sites [230]. This is consistent with our findings and supports a role for HNF4A in regulating gene expression and inflammation in the context of IBD. However, our work is the first to demonstrate a role for microbiota in suppressing HNF4A, and to implicate microbiota-HNF4A interactions in driving an IBD-like gene expression signature (Fig. 5). In addition to IBD, human *HNF4A* variants are associated with metabolic syndrome [303] and type 2 diabetes [304]. Interestingly, microbiota have also been implicated in both of these diseases [305, 306] raising the possibility that microbiota suppression of HNF4A *trans* activity could play a role in these diseases as well. Indeed, we find that genes down regulated in intestinal tissue from IR obese patients have increased HNF4A binding associations compared to up-regulated genes [298], similar to the enrichment of HNF4A associations at down-regulated genes in IBD (Fig. 5B,C). Interestingly, up-regulated genes in these IR-obese patients were enriched for pro-inflammatory markers. This underscores the relationship between metabolic impairments and inflammation in the intestine, and prompts further investigation of how HNF4A might contribute. HNF4A has been shown to play key roles in anti-oxidative and anti-inflammatory defense mechanisms [307] so aberrant microbial suppression could promote an inflammatory state. HNF4A target genes are downregulated in human IBD [237, 238] and mouse experimental colitis [260], and the HNF4A target *APOA1* has been shown to be protective against intestinal inflammation in mice [262]. We speculate that the genes governed by this novel microbiota-HNF4A axis may include additional anti- and pro-inflammatory factors that could provide new targets for IBD therapy.

Our results reveal similar effects of microbiota colonization and experimental colitis on HNF4A cistrome occupancy in the intestine, but the underlying molecular mechanisms are unresolved. DSS induced colitis results in reduced HNF4A protein levels and altered cellular localization [260], however our results indicate the microbiota neither reduce HNF4A protein levels nor impact its nuclear localization in jejunal IECs two weeks after colonization (Supplemental Fig. 65H,I). Colonization of GF mice with microbiota initiates a transcriptional adaptation in the intestine that progresses for several weeks before reaching homeostasis [121]. However, our data indicate HNF4A suppression is achieved within 15 days and persists through at least 45 days after colonization. These data collectively suggest that microbiota suppress HNF4A activity in the jejunum through mechanisms distinct from those utilized in DSS induced colitis.

HNF4A has been characterized as a master metabolic regulator for its conserved roles in gluconeogenesis, glucose homeostasis, and fatty acid metabolism [118, 308, 309]. Despite its clear importance in metabolic health, relatively little insight into its regulation in a biological context has been reported. *In vitro* and cell culture studies have identified possible suppressors and activators of HNF4A including acetylation by CREB-binding protein (CBP), which has been shown to induce HNF4A activity [310, 311]. The nuclear receptor cofactor and master regulator of mitochondrial biogenesis PGC-1A binds HNF4A and promotes activation of HNF4A target genes [312]. Colonization of GF animals with a microbiota leads to increased energy harvest [167, 168] and changes in metabolic homeostasis including decreased AMPK activity in skeletal muscle and liver [169]. AMPK activates PGC-1A [313], therefore, microbiota might suppress HNF4A activity indirectly through induced alterations in metabolic homeostasis. Other studies have shown that HNF4A activity is controlled through use of alternative promoters which generate different isoforms [314]. However, we did not detect differential *Hnf4a* exon usage by DEXseq [315] in our RNA-seq data from GF and CV IECs (data not shown). Another facet of HNF4A biology that remains unresolved is the

identity of its endogenous ligand(s). Although historically considered an orphan nuclear receptor, several fatty acids, including linoleic acid, have been identified as ligands for HNF4A [308, 316, 317]. Fatty acids are an attractive class of putative regulators of HNF4A since the microbiota are known to regulate FA absorption in zebrafish IECs [168]. Further, specific bacterial taxa are known to modify the structure of polyunsaturated FAs (PUFAs) and these native and modified PUFAs have distinct impacts on animal health [318] and may serve as therapeutics for IBD [319].

In our attempt to understand how the microbiota regulate HNF4A activity and host gene transcription, we were motivated to investigate if microbiota impact histone modification and chromatin accessibility in the mouse jejunum. Our findings support the model that microbiota alter IEC gene expression by affecting TF binding and histone modification at tissue-defined open chromatin sites [53]. We provide the genomic addresses of hundreds of microbiota-regulated enhancers as well as the genes associated with these enhancers and HNF4A binding sites. Similar to other findings in intraepithelial lymphocytes [320], our work demonstrates a clear microbial contribution to the modification of the histone landscape in IECs and provides another important layer of regulation that orchestrate microbiota regulation of host genes involved in intestinal physiology and human disease. We were also able to establish a link between microbiota-regulated genes and enhancers and NR binding sites. These NR binding sites are coincident with a core set of TFs that are enriched near microbiota-suppressed enhancers/genes (GATA) or induced enhancers/genes (ETS-factors and IRF) (Supplemental Fig. S7). GATA4 was previously shown to be a positive regulator of genes suppressed by microbiota in the mouse jejunum ([132], supporting potential coregulatory interactions with HNF4A. Coregulation by other TFs represents one possible mode of HNF4A regulation by which the microbiota could suppress HNF4A activity without impacting the gene transcription of all HNF4A -associated genes.

3.5 Methods

3.5.1 Zebrafish Husbandry:

Tg(in3.4:cfos:gfp) (Camp et al., 2012) stable transgenic lines were maintained on a TL/Tü background using established protocols approved by the Animal Studies Committee at the University of North Carolina at Chapel Hill and Duke University School of Medicine. Conventionally raised zebrafish were reared and maintained as described [321]. Production, colonization, maintenance, and sterility testing of gnotobiotic zebrafish were performed as described [322].

3.5.2 Mouse Husbandry:

All mouse husbandry was performed as described in [53] using established protocols approved by the Animal Studies Committee at the University of North Carolina at Chapel Hill and Duke University School of Medicine with the following exceptions. All mice used in this study were 10 – 12 week old male C57BL/6J, housed on Alpha-dri bedding (Shepherd) and fed 2020SX diet (Envigo) *ad libitum*. To generate conventionalized mice, germ-free mice were colonized with a conventional microbiota from by receiving a 200 uL oral gavage of 20% glycerol stock containing 1:1 w/v fecal sample collected from adult SPF C57BL/6J mice collected over 2 weeks and homogenized in reduced PBS.

3.5.3 Yeast 1-Hybrid ORFeome Screen:

The yeast 1-hybrid ORFeome screen was performed using the Clontech MatchmakerTM Gold Yeast One-hybrid Library Screening System (cat. 630491) protocol with the following exceptions: The Y1HGold yeast strain was transformed using standard yeast transformation procedures with BstBI digested pBait-AbAi containing either the WT or a SDM in3.4 or the p53 binding site sequence, and positive transformants were selected on

SD/-URA media. In addition, a ORFeome library consisting of 148 zebrafish transcription factors cloned from adult zebrafish liver into pDEST22 prey vectors containing an N-terminal GAL4-activation domain was utilized [323]. Each plasmid was individually transformed in the yeast strains Y1HGold[in3.4/AbAi] or Y1HGold[p53/AbAi] and positive transformants were selected on SD -URA -TRP. The primary screen to test for positive interactions between the prey transcription factor and the bait sequence was tested twice in the laboratory by pipetting 10 uL of transformed yeast onto SD/-URA/-TRP with AbA (125ng/ml) agarose plates. The secondary screen to test for positive interactions was performed by streaking individual colonies from the primary screen onto SD/-URA/-TRP with AbA (125ng/ml) agarose plates. Zebrafish *hnf4a* and *hnf4g* cDNAs (see Supplemental Table 1 for primers used for amplification) were cloned into a custom pENTR plasmid (termed pENTR-Ale1) using In Fusion (Takara Bio 638909) and inserted into pDEST22 using LR clonase (Invitrogen 12538120). These newly cloned transcription factors were tested for a positive interaction with in3.4 using the same procedures as above.

3.5.4 Site Directed Mutagenesis:

Site directed mutagenesis was performed using the primers found in Supplemental Table 1. A 40 cycle PCR reaction was performed using iProof HiFi Polymerase (Biorad 1725301). Newly synthesized plasmids were digested with DpnI (New England Biolabs R0176L) overnight to digest template DNA and transformed into DH5a *E. coli*. SDM Vectors were confirmed by Sanger sequencing.

3.5.5 Zebrafish Transgenesis and Imaging:

Co-injections of Tol2 SDM or WT in3.4:cfos:gfp plasmid and transposase mRNA were performed as described (Camp et al., 2012) with the following exceptions: 50 – 100

zebrafish embryos were injected at the 1–2 cell stage with approximately 69 pg of plasmid DNA at a DNA:transposase ratio of 1:2. At least 9 - 18 fish/construct were imaged on a Leica M205 FA with a Leica DFC 365FX camera at the same magnification and exposure time and densitometric measures were quantified in 8-bit gray scale images using FIJI software. Three mosaic patches within a given tissue of an imaged fish were quantified for mean fluorescence intensity and averaged. Statistical significance was analyzed using Kruskal-Wallis one-way analysis of variance and Dunn's multiple comparison test using GraphPad Prism software.

3.5.6 Zebrafish Mutagenesis:

Targeted gene deletion of the *hnf4a* gene was performed using CRISPR/Cas9 nuclease RNA-guided genome editing targeting the fourth exon of *hnf4a*. The guide RNA sequences were designed using "CRISPR Design Tool" (http://crispr.mit.edu/). Guide RNAs (Supplemental Table 1) were generated from BamHI (New England Biolabs R0136L) digested pT7-gRNA plasmid (a gift from Wenbiao Chen and available from Addgene: http://www.addgene.org/46759/) and by performing an *in vitro* transcription reaction using MEGAshortscript T7 kit (Ambion/Invitrogen AM1354) [324]. Cas9 mRNA was generated from XbaI (New England Biolabs R0145S) digested pT3TS-nls-zCas9-nls plasmid (a gift from Wenbio Chen and available from Addgene: http://www.addgene.org/46757/) followed by an *in vitro* transcription reaction using mMESSAGE mMACHINE T3 kit (Ambion/Invitrogen AM1348) [324]. 150 ng/uL of nls-zCas9-nls and 34 ng/uL of each gRNA, 0.05% phenol red, 120 mM KCl, and 20mM Hepes (pH 7.0) were injected directly into the cell(s) of one to two cell stage developing zebrafish embryos of Tü background. Mutagenesis was initially screened using Melt Doctor High Resolution Melting Assay (Thermo Fisher Scientific 4409535) and subsequent screening of the -43 and +25 alleles was performed using 2% agarose sodium borate gel electrophoresis. Protein and DNA

sequences were visualized using CLC Sequence Viewer 7(CLC Bio) HNF4A protein phylogeny tree was generated in CLC Sequence Viewer 7. The majority of zebrafish experiments were performed using *hnf4a*$^{-43/-43}$ genotype. However, the gnotobiotic zebrafish experiment was performed using larvae from an intercross between *hnf4a*$^{-43/-43}$ and *hnf4a*$^{-43/+25}$ adults. We have not observed significant gene expression or morphological differences between *hnf4a*$^{-43/-43}$ and *hnf4a*$^{+25/+25}$ genotypes.

3.5.7 Zebrafish Immunohistochemistry:

6 dpf zebrafish larvae were fixed in 4% PFA overnight at 4 C. Fixed larvae were mounted in 4% low melting point agarose molds. 200 um axial cross sections of fixed larvae were generated using a Leica VT1000S. Vibratome slices were washed once in ice cold PBS followed by 4 times with PBS containing 0.1% tween 20 and then incubated in blocking solution (PBS with 10% heat inactivated calf serum, 0.1% Tween-20 and 0.5% Triton X-100) for 4 hours. Slices were incubated overnight with 4e8 antibody (Mouse anti-4e8, Abcam ab73643) diluted 1:200 in PBS with 5% heat inactivated calf serum, 0.1% Tween-20 and 0.5% Triton X-100 at 4°C with agitation. Samples were washed in PBST 3 times for 10 minutes per wash and incubated with secondary antibody (1:1000) (Goat Anti-Mouse Alexa Fluor 568 Invitrogen, A11004) and Alexa Fluor 647 phalloidin (1:300) (Invitrogen, A22287) in PBS with 5% heat inactivated calf serum, 0.1% Tween-20 and 0.5% Triton X-100 for 3 hours. Slices were washed in PBS 3 times for ten minutes per wash, mounted onto slides with DAPI mounting media (Vector Laboratories, Inc, H-1200) and imaged on a Leica SP8 confocal microscope. Images shown in Figure 3.1F are representative of two experiments with 3 larvae per experiment per genotype.

3.5.8 Mouse IEC Isolation:

Mice were euthanized under CO_2 and cervical dislocation and placed on a chilled wax dissection pad. The small intestine was removed from the mouse and the jejunum was excised from the duodenum and ileum. Duodenum was defined as the anterior 5 cm of the midgut and ileum was defined as posterior 6 cm of midgut as described (Camp, et al 2014). Adipose and vasculature were removed from the tissue. The jejunum was opened longitudinally along the length of the tissue, exposing the lumen and epithelial cell layer. Luminal debris was washed away from the epithelia with ice cold sterile PBS. The tissue was temporarily stored in 10 ml of ice cold sterile PBS with 1x Protease Inhibitor (Complete EDTA-Free, Roche 11873580001) and 10 uM Y-27632 (ROCK I inhibitor, Selleck Chemicals S1049) to inhibit spontaneous apoptosis. The jejunum was moved into a 15 ml conical tube containing 3 mM EDTA in PBS with 1x protease inhibitor and 10 uM Y-27632. The tissue was placed on a nutator in a cold room for 15 minutes. The jejunum was removed from the 3 mM EDTA and placed on an ice cold glass petri dish with PBS containing 1mM MgCl2 and 2 mM CaCl2 with protease inhibitors and 10 uM Y-27632. Villi were scraped off of the tissue using a sterile plastic micropipette and placed into a new 15 ml conical tube. The isolated IECs were then pelleted at 250 x g at 4°C for 5 minutes, resuspended in 15 ml of ice cold PBS containting 10 uM Y-27632 and 1x protease inhibitors and pelleted again at 250 x *g* at 4°C. The cell pellet was used for chromatin immunoprecipitation or for nuclear extractions.

3.5.9 Mouse Intestine Immunofluorescence and Western Blot:

Mid-jejunal tissue was dissected and cleaned as in the IEC villi isolation above. The whole, splayed open tissue was pinned to 3% agarose and fixed in 4% PFA overnight with gentle agitation at 4°C. The fixed tissue was washed 4 times with PBS for 15 minutes. The tissue was then permeabilized in PBS with 0.5% Tween 20 for 1.5 hours at room temperature. Following permeabilization, the tissue was blocked in 5% donkey serum in

PBS with 0.1% Tween 20 for 2 hours at room temperature. The tissue was moved into a 35 mm dish and incubated with the primary antibody (Mouse anti- HNF4A, Abcam 41898 or Goat anti-HNF4G, Santa Cruz sc-6558X) diluted 1:200 overnight at 4°C with gentle agitation. The tissue was washed 4 times in immunowash buffer (PBS, 0.1% Tween-20, 1% DMSO and 1% BSA) at room temperature and incubated in secondary antibody (Goat Anti-Mouse Alexa Fluor 488 Invitrogen, A11001 or Donkey Anti-Goat Alex Fluor 568 Invitrogen, A11057) diluted 1:100 and Alexa Fluor 488 phalloidin or Alexa Fluor 568 phalloidin diluted 1:250 (Invitrogen A12379 and A12380, respectively) in PBS with 0.1% Tween-20 for 4.5 hours at room temperature. The tissue was then washed 6x in Immunowash buffer, mounted on a microscope slide with DAPI mounting media (Vector Laboratories, Inc, H-1200) and imaged on a Leica SP8 confocal microscope.

Western blots were performed on non-crosslinked IEC lysates (see below) using standard chemoluminescence Western blot protocols, including ECL (Biorad 170-5061) and primary antibodies Goat anti-HNF4A (Santa Cruz sc-6556), Goat anti-HNF4G (Santa Cruz sc-6558X) and Rabbit anti-ACTB (Cell Signaling 13E5), Donkey anti-Goat-HRP conjugate (Santa Cruz sc-2020), and Goat anti-Rabbit-HRP conjugate (Jackson ImmunoResearch 111-035-003, a gift from Stacy Horner at Duke University). The western blot shown in Supplemental Figure 3.S6H is a representative of two experiments.

3.5.10 Cell Lysis and Chromatin Sonication for ChIP:

Washed and pelleted IECs were resuspended in 10 ml of 1% EM grade formaldehyde (Electron Microscopy sciences, 15710) in ice cold PBS containing 10 uM Y-27632 and 1x protease inhibitors. The cells were fixed for 25 minutes at room temperature with agitation. Formaldehyde fixation was quenched by adding glycine to a final concentration of 125 mM. The cells were pelleted at 250 x g for 5 minutes at 4°C and resuspended in ice cold PBS with 1x protease inhibitors and 10 uM y-27632. This wash

step was repeated twice. Upon the third wash, the cell pellet was aliquoted into 3 equal volumes in 3 microfuge tubes. The cells were pelleted in the microfuge tubes and resuspended in 300 uL of ChIP Lysis Buffer (1% IGEPAL, 0.5% sodium deoxycholate, 1% SDS in 1x PBS) containing 1x protease inhibitor (Roche). The cells were stored on ice for 2 hours and sonicated using a Bioruptor 4°C water bath sonicator. Chromatin was sheared to mean size of 250 – 300 bp (10 minutes of 30 seconds on High, 30 seconds off, repeated once for a total of 20 minutes – total sonication time is 10 minutes on High, 10 minutes off). Sonicated material was spun at 14,000 x *g* for 20 minutes and the supernatant was transferred to a new microfuge tube. ChIP was performed immediately on sonicated chromatin or it was snap frozen and stored at -80°C. To check chromatin shearing efficiency and to prepare ChIP input samples, 20 uL of each sonicated sample was removed and added to a new tube. 180 uL of ChIP elution buffer and 8 uL of 5 M NaCl was added to the 20 uL input samples. Chromatin shearing efficiency was visualized on a gel following reverse crosslinking by incubating the input sample at 65°C overnight.

3.5.11 Chromatin Immunoprecipitation, Library Preparation and Next-Generation Sequencing:

Frozen sonicated chromatin was thawed on ice. Thawed and fresh chromatin samples were diluted in 1 mL of ChIP dilution buffer (1% Triton X-100, 2 mM EDTA, 20 mM Tris-Cl (pH 8.1), and 150 mM NaCl) containing 1x protease inhibitor and precleared with washed protein G Dynabeads for 3 hours at 4°C on a nutator. Precleared chromatin was incubated with ChIP grade antibodies [4 ug H3K4me1 (Rabbit anti-H3K4me1, Abcam ab8895), 4 ug H3K27ac (Rabbit anti-H3K27ac, Abcam ab4729), 8 ug HNF4A (Mouse anti-HNF4A, Abcam 41898), 8 ug HNF4G (Goat anti-HNF4G, Santa Cruz sc-6558X), 8 ug CTCF (Rabbit anti-CTCF, Active Motif, 61311)] overnight at 4°C on a nutator. Antibody-chromatin complexes were pulled down with washed protein G dynabeads for 4 hours at 4°C on a

nutator. The beads were washed 5x for 3 minutes with ice cold LiCl wash buffer (100 mM Tris-Cl (pH 7.5), 500 mM LiCl, 1% IGEPAL, 1% sodium deoxycholate)) and 1x with ice cold TE buffer at 4°C on a nutator. Washed beads were resuspended in 100 uL of ChIP elution buffer (1% SDS and 0.1 M sodium bicarbonate)) and placed in a thermomixer heated to 65°C and programed to vortex at 2000 RPM for 15 seconds, rest for 2 minutes for a total of 30 minutes. The beads were pelleted, placed on a magnet, and the supernatant was moved to a new tube. This elution process was repeated once and corresponding elutions were combined for a total of 200 uL. To reverse crosslink immunoprecipitated chromatin, 8 uL of 5 M NaCl was added to each 200 uL ChIP elution and elutions were incubated at 65°C overnight. Immunoprecipitated chromatin was isolated using a QIAquick PCR quick preparation kit (Qiagen 28104), quantified using a Qubit 2.0 flourometer and stored at -80°C until library preparations and amplification. Libraries were always prepared within 3 days of the immunoprecipitation with the NEBNextUltra DNA Library Prep Kit for Illumina (New England Biolabs E7370S). Prepared libraries were quantified using a Qubit 2.0 fluorometer and submitted to Hudson Alpha Genomic Services Laboratory for 50 bp single end sequencing on an Illumina HiSeq 2500 with 4 samples per lane in the flow cell.

Germ free or conventionalized chromatin for input normalization was generated using the same protocol as above except no antibody was used during the overnight antibody incubation; instead, chromatin was incubated at 4°C with gentle agitation. Bead incubation, reverse-crosslinking and library preparations for these samples were performed using the same protocol as the ChIPs.

3.5.12 DNase Hypersensitivity on IECs:

DNase hypersensitivity was performed as described [53] with the following modifications: IECs were isolated as above from jejunum and subjected to endogenous DNase activity to digest chromatin. DNase-seq libraries were constructed as previously described, with “Oligo

1b" phosphorylated at the 5'-end to enhance ligation efficiency [325]. Libraries were sequenced by Illumina HiSeq 2000 with 50 bp single end reads with 3 samples per lane.

3.5.13 RNA Isolation, qRT-PCR, RNA-seq:

For RNA-seq, zebrafish digestive tracts (n = 13 – 20 per condition per genotype) were removed by microdissection and resuspended in 1 mL TRIzol (Ambion/Invitrogen/ThermoFisher Scientific 15596026). Larval digestive tracts were lysed by being passing through a 25 G needle followed by a 27.5 G needle 5 times each. Mouse jejunum intestinal epithelial cells were collected as mentioned above. Prior to crosslinking, 1/50 of the isolated IECs were suspended in 1 ml TRIzol and stored at -80°C. For both zebrafish and mouse RNA samples, 200 uL of chloroform was added to the TRIzol and the sample was vortexed on high for 30 seconds at room temperature. The samples were incubated at room temperature for 2 minutes and centrifuged at 12,000 x *g* for 15 minutes at 4°C. The top aqueous layer was removed and added to equal volume of isopropanol. The nucleic acids were isolated using a column-based RNA-isolation kit (Ambion Cat 12183018A) with an on column DNase I (RNase-free) treatment (New England Biolabs M0303L) to remove DNA contamination. RNA was eluted off the column in nuclease-free water, quantified using a Qubit 2.0 and stored at -80°C until submission to the Duke Sequencing and Genomic Technologies Core. RNA-seq libraries were prepared and sequenced by Duke Sequencing and Genomic Technologies Core on an Illumina HiSeq 2500 with 4 samples per lane in the flow cell.

For qRT-PCR, adult zebrafish IECs from one adult or 6 dpf larvae (n= 5 – 10) with the same genotype were suspended in 1 mL of TRIzol. RNA was extracted using the same protocol as above with the following exceptions: following sample resuspension in isopropanol, samples were frozen at -20°C O/N and spun at 15,000 x *g* for 30 minutes. Pellets were washed twice with RNase-free 70% ethanol and left to air dry for 10 minutes.

Nucleic acids were resuspended in RNase-free water containing DNase (DNA-free DNA Removal Kit, ThermoFisher Scientific AM1906) and incubated for 30 minutes at 37°C. DNase was inhibited and purified RNA was quantified using a Nanodrop spectrophotometer and stored at -80°C. cDNA was generated using iScript cDNA synthesis kit (Biorad 1708891) and qRT-PCR was performed using Quanta's PerfeCTa Sybr-green (Quanta 101414-154) in an Applied Biosystems StepOnePlus Real-Time PCR Systems machine (Supplemental Table 1).

For ChIP-PCR, Immunoprecipitated chromatin was isolated using a QIAquick PCR quick preparation kit (Qiagen 28104), and stored at -20°C. Immunoprecipitated chromatin was used as template in a qRT-PCR reaction using Quanta's PerfeCTa Sybr-green (Quanta 101414-154) in an Applied Biosystems StepOnePlus Real-Time PCR Systems machine (Supplemental Table 1).

3.5.14 RNA-seq Bioinformatics:

Zebrafish RNA-seq reads were aligned to the zebrafish genome (danRer7) using TopHat2 v0.6 using *de novo* splice junction mapping (default TopHat settings). FPKM expression values were obtained for transcripts via Cufflinks, and pairwise differential gene expression tests were carried out with Cuffdiff v0.0.6 (Trapnell et al. 2012) using a minimum alignment count of 100 and using multi-read correct and read group datasets. The default significance threshold of FDR < 5% was used for each comparison. Hierarchical clustering of replicates and gene expression heatmap of RNA-seq data were generated using complete linkage clustering and averaging the $\log_{10}$ (FPKM) of each gene with Cluster v3.0. Subsequent heatmaps in Supplemental Figure 3.S3 were generated using complete hierarchical gene clustering of the $\log_2$ (fold change) between compared conditions with Cluster v3.0. Principle components analysis was performed with a wide estimation method with JMP13. Discriminant analysis was performed with a wide linear method with JMP13.

GO enrichments were performed with DAVID 6.7 [326, 327]. The relatedness heatmap was generated with principle components in R (ggplot2 package). A 4-way Venn diagram was generated using the 4,007 genes that were differentially regulated in any comparisons with online software: http://bioinformatics.psb.ugent.be/webtools/Venn/. Disease associations were performed using DRSC Disease Gene Query Tool [328] (http://www.flyrnai.org/diopt-dist). The number of genes associated with various diseases were added together and ranked by total number of associated genes. Unique genes associated with "Inflammatory Bowel Disease", "Crohn's Disease", and "Ulcerative Colitis" were combined into one IBD list. If a zebrafish gene was orthologus to many human gene associated with the same gene, the zebrafish gene was counted only once for a given disease. For example, *fads2* has 3 human othologs: *FADS1*, *FADS2*, and *FADS3*. *FADS1* and *FADS2* are both associated with IBD but the gene family was only counted once.

Mouse RNA-seq reads were aligned to the mouse genome (mm9) using TopHat v2.1.0 using *de novo* splice junction mapping with default settings). Normalized fragments per kilobase of transcript per million mapped reads (FPKM) expression values were obtained for genes via Cufflinks, and pairwise differential gene expression tests were carried out with Cuffdiff v2.2.1.3 (Trapnell et al. 2012) using multi-read correct, bias correction and read group datasets. The default significance threshold of FDR < 5% was used for each comparison. To assess the association of differential DHS, H3K4me1 and H3K27ac and nearby gene expression differences in the presence and absence of a microbiota, we linked chromatin marks found within 10kb upstream or downstream of the nearest expressed gene (as defined by RNA-seq; minimum alignment count of 100 reads was used to determine detectable expression) transcription start site with that gene using GREAT version 3.0.0 (Single nearest gene definition) for putative regulation. The distributions of fold-change FPKM values in the presence and absence of a microbiota were compared to the distributions of all 9,173 expressed genes by a two-sided Kolmogorov-Smirnov test. GO

enrichments were performed using DAVID 6.7 [326, 327]. IBD and NEC gene expression summary tables were derived from the published studies [237, 238, 297].

3.5.15 ChIP-seq and DNase-seq Bioinformatics:

Mouse ChIP-seq were aligned to the mouse genome (mm9) using Bowtie2 v2.2.6 with default settings. ChIP peaks were called using the appropriate aligned input reads as the control file with MACS2 callpeak Galaxy version 2.1.0.20151222.0 with FDR <5% as the peak detection threshold. Sequencing depth normalization was performed in two ways: DESeq was used for sequencing depth normalization, variance fitting, and pairwise differential analysis (Anders and Huber 2010) and bamCoverage (Galaxy version 2.2.3.0) using RPKM normalization.

To identify differential H3K27ac and H3K4me1 sites, H3K27ac and H3K4me1 ChIP-seq peak calls were merged using the same parameters as in DNase-seq analysis except using a FDR < 1%. Raw counts were pulled from BAM files and used for calling differential peaks using R packaged DESeq2 v1.10.1 [329]. Motif enrichment in promoters near differential H3K27ac regions were generated using the single nearest gene definition in GREAT v3.0.0 and limiting the regulatory domain to 10kb. Heatmaps and average signal graphs were generated by aligning the average signals around MACS2 peaks from differential H3K27ac regions (Supplemental Figure 3.S4F) or median number of MACS2 called peaks of condition/ChIP: Supplemental Figure 3.S4G,H (CV DHS), Figure 3.4A,B (GF HNF4A-ChIP replicate), Supplemental Figure 3.S6A (CV HNF4A-ChIP replicate), Supplemental Figure 3.S6C (CV HNF4A-ChIP minus GF HNF4A-ChIP). Peaks generated from sequencing noise were omitted from these analyses by MACS2 or manually. Gene associations for GF and CV HNF4A peaks were generated using the replicate with the largest number of MACS2 called peaks using the single nearest gene definition in GREAT v3.0.0 and limiting the regulatory domain to 10kb. Repeat masked FASTA sequences

extracted from the noted regions were submitted for *de novo* transcription factor binding site analysis using MEME-suite (Figure 3.3E,F). Specific transcription factor binding site analysis (Figure 3.4D) was performed using HOMER2. The hierarchal clustering and heatmap representation of PWM enrichment was generated with Cluster v3.0. Deming linear regression analysis was used to determine significant increases in H3K27ac, H3K4me1, and DNase-seq signal around HNF4A binding sites. Pairwise comparison of ChIP-seq signal (Supplemental Figure 3.S5) from individual replicates was performed by generating read counts as determined by HTSeq Galaxy Version 0.6.1.galaxy1 [330] using intersection (nonempty) and non-stranded parameters with a minimum alignment quality of 10. Reads were counted at MACS2 peaks from the GF3 replicate (GFrep3) and GF4 (GFrep4) replicate for HNF4A and HNF4G respectively.

Mouse DNase-seq reads were aligned using Bowtie version 0.12.0, with 2 mismatches allowed and mapping up to 4 sites. The output BAM files were transformed to bed files. Blacklist regions and PCR artifacts were then filtered from bed files. DNase hypersensitivity sites narrow peak calls were generated from MACS2 (version 2.1.0.20140616, https://github.com/taoliu/MACS/), with FDR <1%. To identify differential DHS sites, DHS peak calls for each condition were merged and windowed as described in ChIP-seq analysis. Raw sum counts for each base-pair's DNase-seq signals within each 300 bp window from each replicates were input into R package DESeq2 v1.10.1[329] and differential peaks were identified using FDR <5% (no differential peak calls found) and p < 1%.

3.5.16 Bioinformatic and Statistical analysis:

Sample sizes for zebrafish experiments (noted in figure legends) were selected based on genotype availability and transgenesis efficiency. All sample collection was performed two or more times on independent days. For sequencing experiments, statistical

calls for differential gene expression were made by Cuffdiff2 using parameters stated above [331]. For the zebrafish RNA-seq experiment Next-Gen sequencing was performed once and at the same time to avoid batch effects: WTGF and WTCV (n = 3); MutGF and MutCV (n = 2). We originally collected n = 3 MutGF and MutCV biological replicates, however, using pre-established criteria and to avoid RNA contamination, we excluded one biological replicate from all analysis from these groups because of sequencing reads that mapped within the deleted *hnf4a* exon in the *hnf4a*$^{-/-}$ genotype.

GF mice were randomly chosen by gnotobiotic staff for microbiota colonization (CV) based on their availability and litter sizes. All sample collection was performed two or more times per condition on independent days. GF and CV mouse samples were collected on different days. For sequencing experiments, statistical calls for differential gene expression and differential peak calls were made by Cuffdiff, MACS2, and DSseq2 using parameters stated above [329, 331-333]. For the mouse RNA-seq experiment Next-Gen sequencing was performed once and at the same time to avoid batch effects: GF (n = 2) and CV (n = 2). Paired GF and CV ChIP and library amplification was performed simultaneously. Typically, biological ChIP replicates were sequenced on different days and were always paired with the other condition (i.e. CV and GF were always sequenced together). The number of biological ChIP replicates (noted in figure legends) was dependent on reproducibility between ChIP samples and/or our ability to determine statistical differential sites using DESeq2 (for H3K4me1 and H3K27ac).

All statistical metrics (except where otherwise noted) were performed in Graphpad Prism 7.01. Deming linear regression was used for Figure 3.5 because it is a stronger and more accurate assessment of correlation when both the x and y variables have experimental error. Details regarding the other statistical tests used in this study can be found in the figure legends or above.

3.6 Data Access

Transcription factor ChIP-seq, Histone ChIP-seq, DNase-seq and RNA-seq datasets have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE90462. More supplemental tables are accessible on the Genome Research website.

3.8 Supporting Information

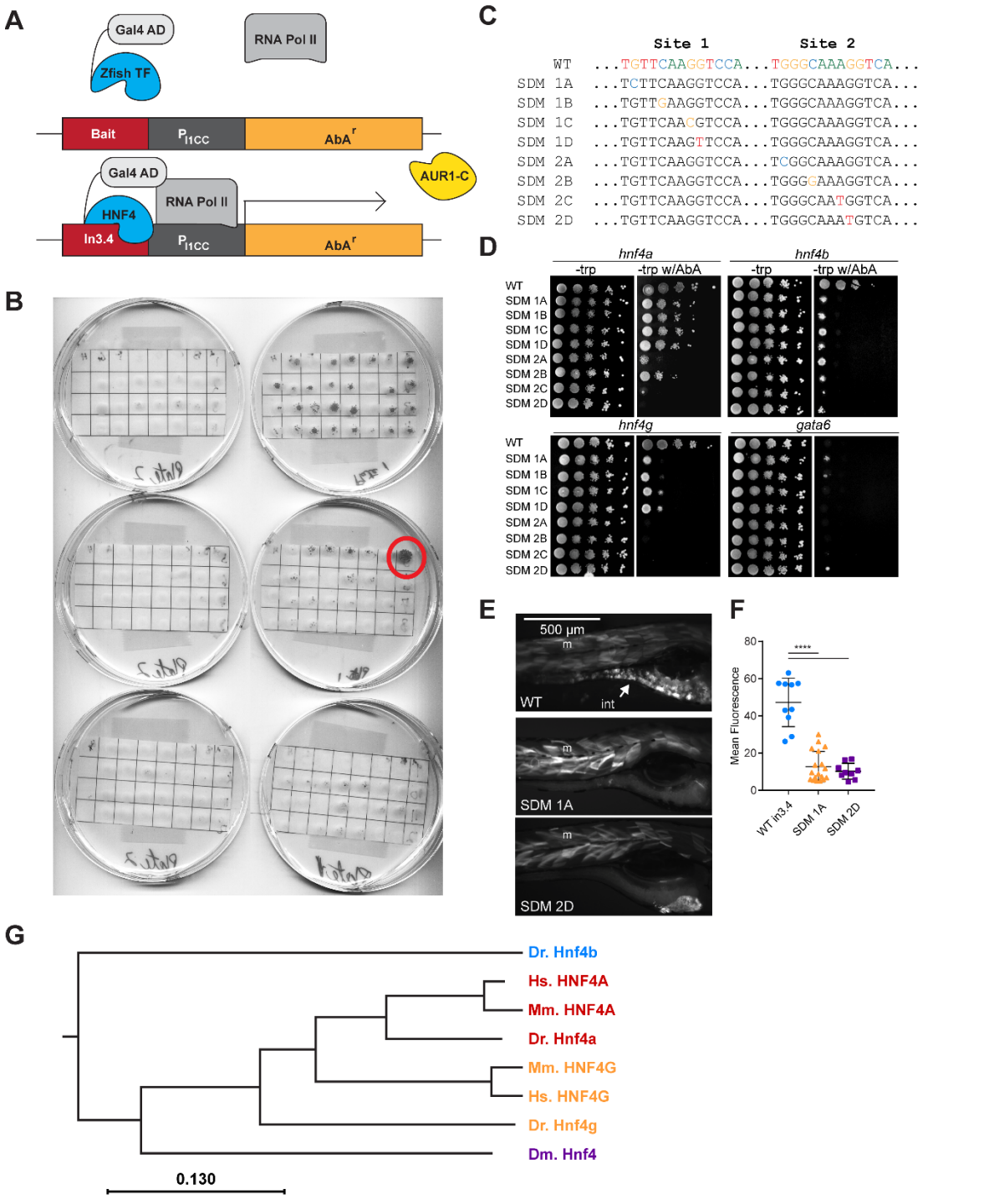


Figure 3.S1: The *hnf4* family of transcription factors bind specifically to a microbiota suppressed zebrafish enhancer. (A) Schematic showing the molecular mechanism of the yeast-1-hybrid assay. A library of 148 zebrafish transcription factors was transformed into a yeast strain that contained a reporter construct that contained the zebrafish CRR in3.4. If one of the zebrafish TFs bound to in3.4, the reporter gene will be transcribed and the yeast will grow on media containing the antibiotic Aureobasidin A (Aba). (B) Scanned plates from the primary yeast-1-hybrid (Y1H) screen

that identified hnf4b (red circle) as the only transcription factor that robustly rescued yeast growth on the selective media. hnf4a and hnf4g were not included in the initial 148 TF library. After cloning these TFs into the prey vectors, we found that all three Hnf4 family members bound to the in3.4 bait sequence and drove expression of the reporter gene (Figure 3.1B). Gata6 is used as a control since it is predicted to bind in3.4, but does not rescue yeast growth. However, this TF family failed to rescue yeast when in3.4 was replaced with a canonical p53 binding site control bait sequence, suggesting a sequence specific interaction (data not shown). (C) Single nucleotide site-directed mutations (SDM) within Site 1 and Site 2. The selected nucleotides were predicted to impact Hnf4 binding. The nucleotide mutation is highlighted in a color for each SDM. Using mammalian position weight matrices (PWMs), we found HNF4 is predicted to bind both regions previously shown to be essential for intestinal reporter expression (Figure 3.1A). To test if these putative binding sites are the location of Hnf4 binding and essential for reporter activity, we performed site-directed mutagenesis on individual nucleotides within the predicted Hnf4 binding sites (Figure 3.1C). (D) Images of plates from serial dilutions of a Y1H assay using WT and mutated in3.4 as bait and zebrafish hnf4 genes and gata6 as prey. Yeast were grown on media without the selective antibiotic to demonstrate equivalent CFUs were plated and on media containing AbA which inhibits yeast growth in the absence of an actively transcribed reporter gene. Mutations in the first predicted binding site (Site 1) resulted in severe growth attenuation of yeast transformed with the hnf4g and hnf4b prey vectors. However, yeast transformed with the hnf4a prey vector only had partial attenuated growth when harboring mutations in the Site 1 sequence. Strikingly, mutations in the second predicted binding site (Site 2) resulted in failed growth of yeast transformed with all three hnf4 prey vectors with the notable exception of Site 2.2, which only partially attenuated growth of yeast transformed with the hnf4a prey vector. (E) To test if the putative HNF4A site was essential for in3.4 enhancer activity, we generated new versions of the in3.4:cfos:gfp reporter which contained single nucleotide mutations in Site 1 or 2. These reporter constructs were injected into wild-type zebrafish to generate mosaic transgenics. Single nucleotide mutations in Site 1 or 2 of the in3.4:cfos:gfp reporter were sufficient to ablate in3.4 intestinal activity in zebrafish. Data shown in panels D and E establish that one or more Hnf4 family members bind in3.4 in a sequence dependent manner and that mutation of the predicted Hnf4 binding sites of this microbiota-suppressed CRR result in suppressed enhancer activity in the intestinal epithelium. (F) Chart showing the GFP fluorescence (mean ± sem) in 6dpf mosaic zebrafish injected with transposase and WT (n =10) and SDM in3.4:cfos:gfp Tol2 vectors (n =9 and n = 17). Mean fluorescence was measured within the intestine of mosaic animals using a constant region of interest (ROI) (Kruskal-Wallis, Kruskal Wallis statistic = 20.26 and **** $p < 0.0001$). (G) HNF4 protein phylogenetic tree showing the evolutionary relationship of the Hnf4s across species. Dm. – Drosophila melanogaster, Dr. – Danio rerio, Mm. – Mus musculus, Hs. – Homo sapiens.

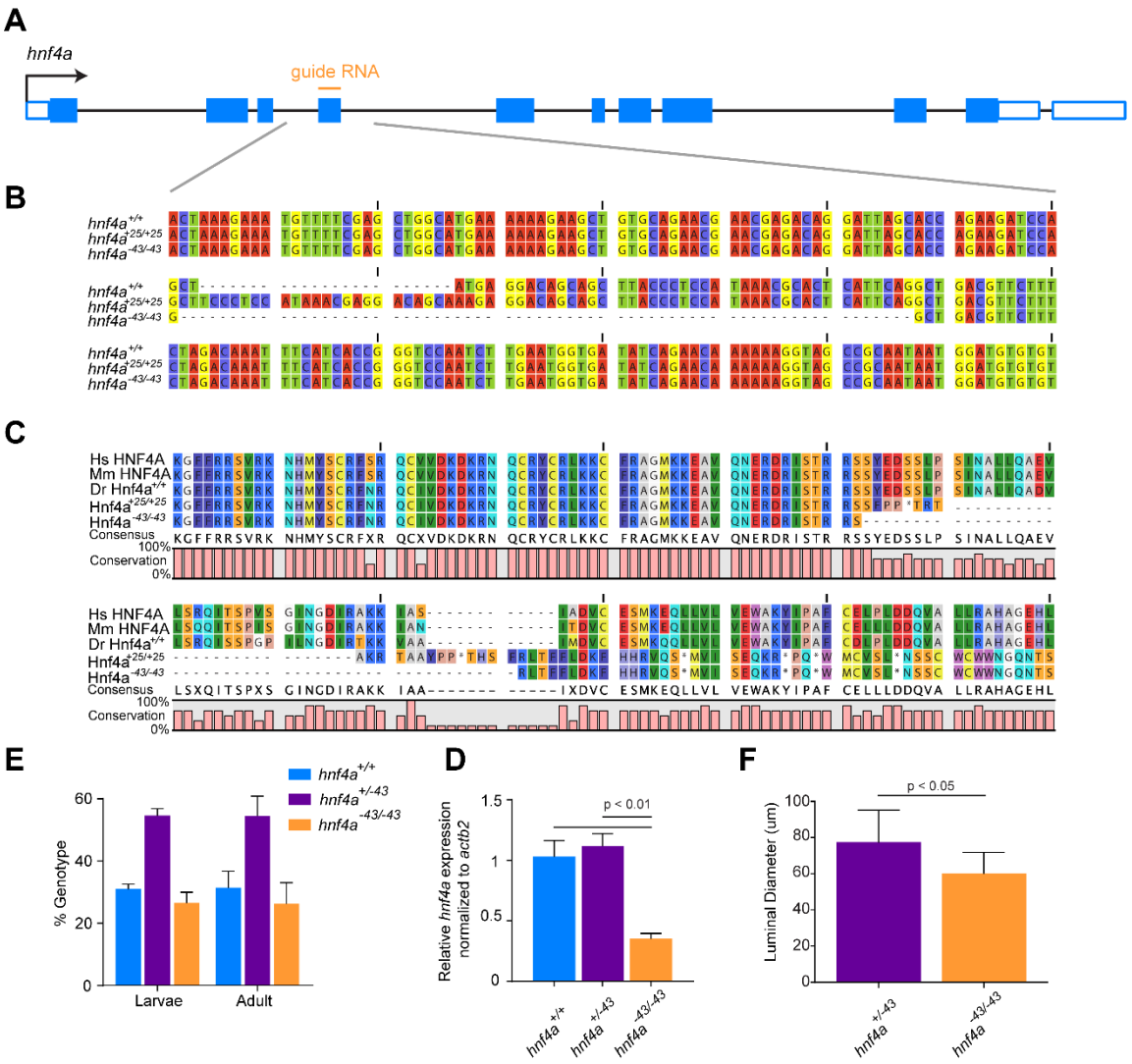


Figure 3.S2: *hnf4a*$^{-43/-43}$ mutants survive to adulthood and have reduced hnf4a transcript and reduced intestinal lumen size. (A) Schematic showing the zebrafish *hnf4a* gene locus (splice form hnf4a-201) and the region that was targeted by the guide RNAs. Exons are highlighted in solid blue blocks, untranslated regions are indicated by white blocks with blue outlines, and the CRISPR targeted region is indicated by the orange line. (B) DNA sequence showing the genomic region that is mutated in the *hnf4a*$^{+25}$ and *hnf4a*$^{-43}$ allele. (C) Amino acid sequence of human, mouse, and WT and mutant zebrafish Hnf4a proteins showing sequence conservation in the DNA binding domain and hinge domain. The *hnf4a*$^{+25/+25}$ and *hnf4a*$^{-43/-43}$ mutations are predicted to result in truncated proteins in this highly conserved domain. (D) Bar graph showing genotypes at the expected Mendelian ratios of progeny from an *hnf4a*$^{+/-43}$ heterozygous incross at both 6dpf and adult stages (mean ± sem). (E) Bar graph showing the *hnf4a* relative mRNA expression (mean ± sem) from whole *hnf4a*$^{+/+}$ (n = 4), *hnf4a*$^{+/-43}$ (n = 4) and *hnf4a*$^{-43/-43}$ (n = 4) 6dpf larvae (Two-tailed *t*-test, $t = 4.79, 6.734$, respectively and df = 6). (F) Bar graph showing the diameter of the intestinal lumen (mean ± sem) along the first segment of *hnf4a*$^{+/+}$ (n = 9) and *hnf4a*$^{-43/-43}$ (n = 8) 6dpf larvae (Two-tailed *t*-test, $t = 2.56$, df = 15 and p = 0.0219).

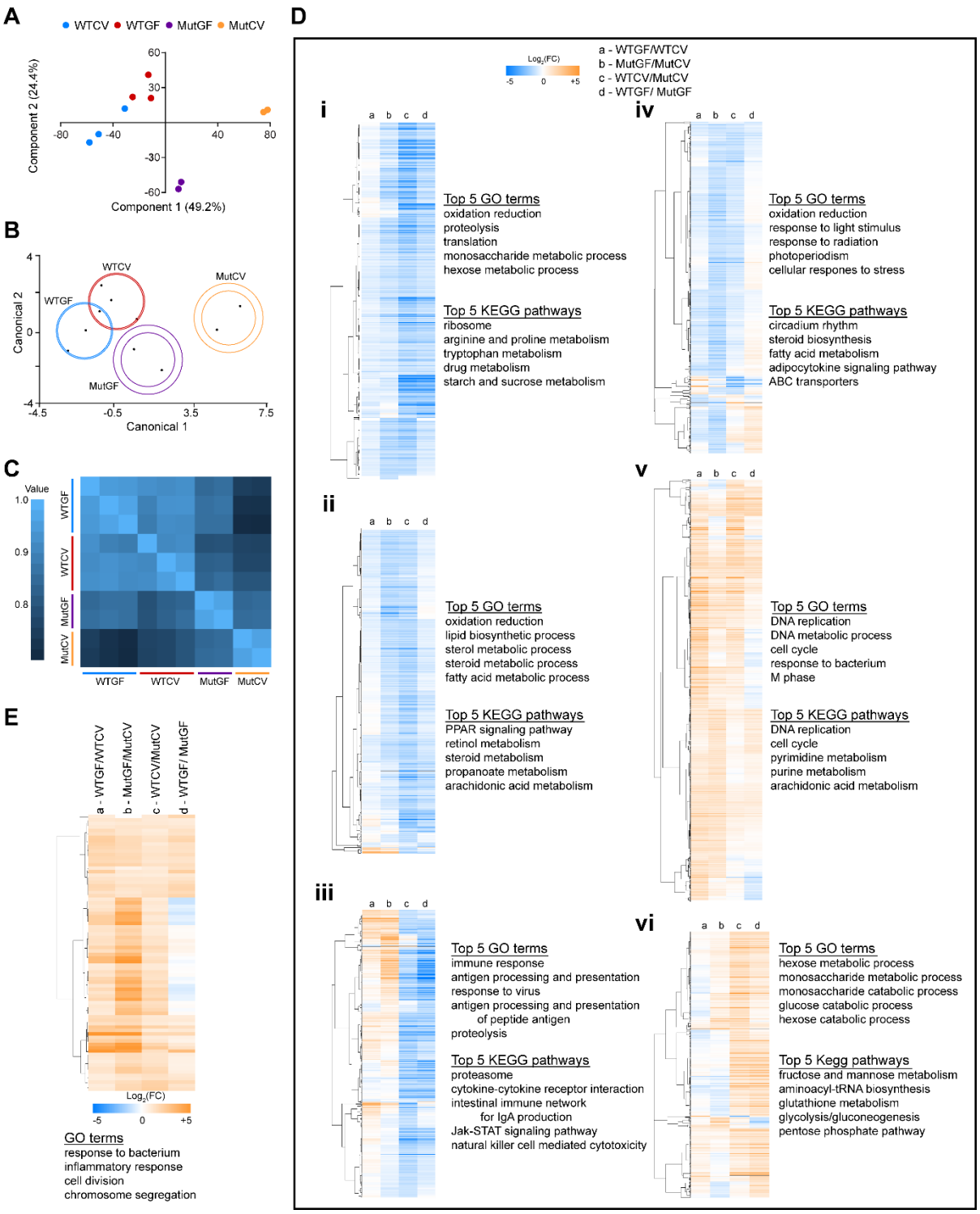


Figure 3.S3: Hnf4a maintains transcriptional homeostasis in the presence of a microbiota in zebrafish digestive tracts. (A) Principle components analysis (PCA) using a wide estimation method showing the relative similarities of mRNA-seq transcript abundances in digestive tracts from WTGF (Blue), WTCV (purple), MutGF (red), and MutCV (orange) 6dpf zebrafish. (B) Discriminant analysis using wide linear parameters plot showing the relative similarities and statistical groupings of mRNA-seq transcript abundances in digestive tracts from WTGF (Blue), WTCV (purple), MutGF (red), and MutCV (orange) 6dpf zebrafish. The inner ellipse of each group signifies the 95% confidence interval

to contain true mean of that group and the outer ellipse signifies 50% of the population within the group is contained within ellipse. (C) Heatmap showing the relative similarities of mRNA-seq transcript abundances in digestive tracts from WTGF (Blue), WTCV (purple), MutGF (red), and MutCV (orange) 6dpf zebrafish. (D) Heatmaps referred to in Figure 3.3B showing the log_2 (FC) relative expression of sublists of differentially regulated genes. The top 5 enriched GO terms and top 5 enriched KEGG pathways for each sublist of genes is included to the left of each heatmap. The 4 columns are: WTCV/WTGF, MutCV/MutGF, MutCV/WTCV, and MutGF/WTCV, respectively. (E) Heatmap showing the Log_2 (FC) relative expression of 86 genes that have an exacerbated microbiota induction in the *hnf4a*$^{-/-}$ digestive tracts. The GO term enrichment for this set of genes is included to the left of the heatmap.

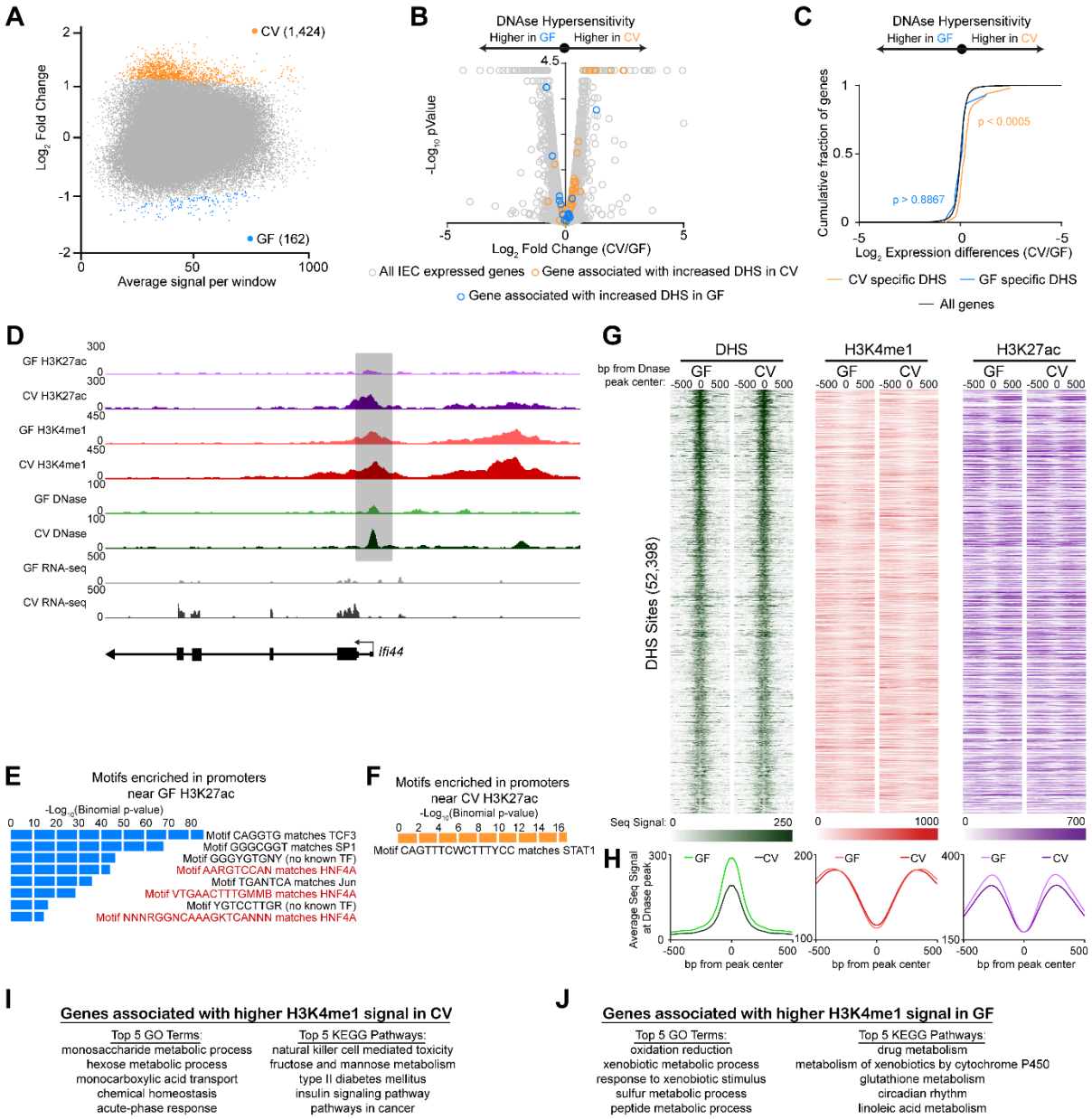


Figure 3.S4: HNF4A and STAT1 binding sites are enriched with within promoters of microbiota suppressed and induced genes, respectively. (A) MA plots from DESeq2 analysis of DNase hypersensitivity from GF and CV mouse jejunal IECs. Colored dots signify regions enriched for a histone mark in GF (blue) or CV (orange) jejunal IECs. We detect 162 unique DHSs in GF and 1,424 unique DHS in CV. These results were generated using a less stringent significance test ($p < 0.01$) rather than FDR < 0.01 utilized in our other comparisons in this study, therefore, we are less confident in these results. Using a FDR cutoff of < 0.01, none of these regions are significantly different. (B) Volcano plots showing pairwise comparison of RNA expression between GF and CV jejunal IECs. Blue and orange dots represent genes associated with a region enriched for DNase signal in GF or CV, respectively. (C) Two-sided Kolmogorov-Smirnov goodness-of-fit test shows a positive relationship between the presence of a DHS in a specific colonization state and increased transcript abundance in that same colonization state. The y-axis shows the cumulative fraction of genes linked to microbiota induced DHSs. Deviation from the null expectation that linked genes display a normal

distribution on a fold change of 1 (black line) suggests that microbiota induced DHSs are enriched near genes of higher expression upon microbiota colonization. Though the Kolmogrov-Smirnov test may not be sufficient for such a small number of sites, these results suggest that IEC gene expression responses to microbiota are not explained by changes in chromatin accessibility. (D) Representative signal track highlighting a microbiota induced gene associated with a CV-specific DHS and regions enriched for H3K4me1 and H3K27ac CV (*Ifi44*, *Interferon induced protein 44).* Bar graphs showing the enrichment for specific TF bindings sites within promoters for genes associated with increased H3K27ac regions in GF (E) or CV (F). (G) Heat maps of the replicate average DNase signal, H3K27ac signal, and H3K4me1 signal at individual CV DHS sites. Despite the MA plots indicating an increase in DNase-seq signal in CV conditions (Fig.3A and Supplemental Figure 3.S4A), we find these heat maps do not show this trend. (H) Line plots showing the average GF (light-colored line) and CV (dark-colored line) -seq signal for the indicated TF, histone mark or DHS at the 1000 bp flanking DHS sites found in CV. The average DNase-signal at all DHS sites is significantly increased in GF compared CV. Similarly, the average H3K27ac signal at these DHS sites shows a significant increase in H3K27ac signal in GF compared to CV by the Whitney-Mann U test. We do not see this trend in H3K4me1. The tentative discrepancies in results between the MA plots and average signal plots can be explained by inherent differences in DESeq2 and average signal analysis. DESeq2 performs a powerful statistical test to determine differential signal and therefore is a more ideal type of analysis to identify enrichment of DHS or histone marks based on variation and average signal of replicates within a set window. We performed the second type of analysis which shows the average signal at a given base pair relative to the center of a ChIP peak because DESeq2 would be an inappropriate analysis for downstream applications in which the GF conditions had an overwhelming signal compared to CV (i.e., HNF4A and HNF4G ChIP-seq, Figure 3.4). Instead, we find this second type of analysis is only useful when comparing the signals or relative signals from the same condition at two different sets of genomic locations (see Supplemental Figure 3.5F,G). Furthermore, the average number and median number of DHS sites, H3K27ac peaks and H3K4me1 peaks as determined by MACS2 were similar between GF and CV conditions; this was not true for HNF4A or HNF4G peaks (Supplemental Table 2). Based on our stringent DESeq2 analysis (FDR < 0.01) and because of the similar number of MACS2 peaks, we do not believe the average H3K27ac and DHS signal differences at DHS sites between CV and GF conditions is biologically or technically relevant to our conclusions. (I-J) GO terms and KEGG pathways enriched in genes associated with differential H3K4me1 sites shown in Figure 3.3I.

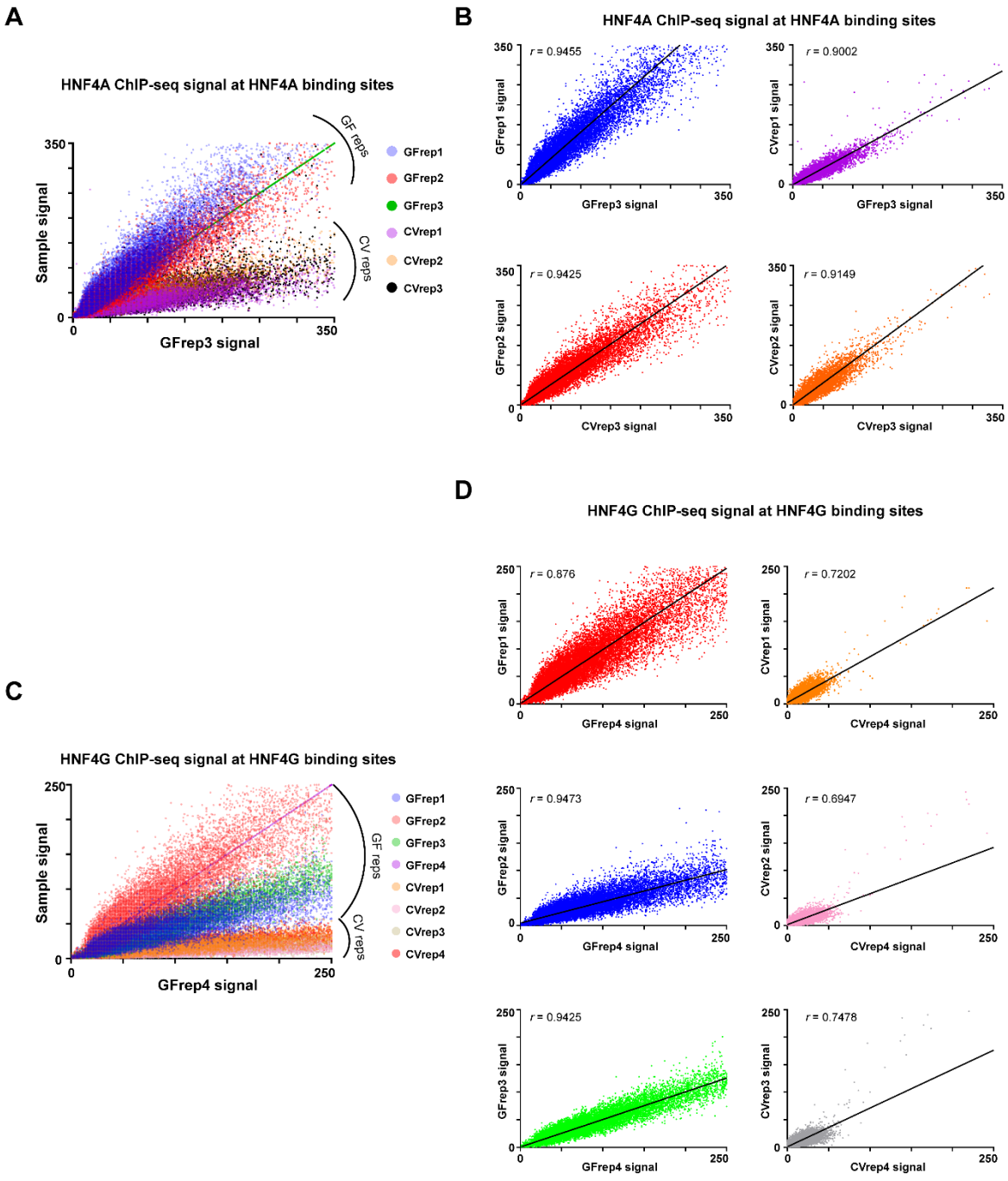


Figure 3.S5 HNF4 GF ChIP-Seq replicates have reproducibly higher signal than HNF4 CV Chip-seq replicates. (A) Grouped pairwise comparison of all HNF4A ChIP-seq signal at HNF4A binding sites compared to GFrep3 (third HNF4A GF replicate). (B) Individual pairwise comparison of HNF4A ChIP-seq signal at binding sites compared to GFrep3 (third HNF4A GF replicate) or CVrep3 (third HNF4A CV replicate). (C) Grouped pairwise comparison of all HNF4G ChIP-seq signal at HNF4G binding sites compared to GFrep4 (fourth HNF4G GF replicate). (D) Individual pairwise comparison of HNF4G ChIP-seq signal at HNF4G binding sites compared to GFrep4 (fourth HNF4G GF replicate) or CVrep4 (fourth HNF4G CV replicate). The correlation coefficient (*r*) is provided for each graph. We believe the HNF4G CV *r* value was substantially lower than the other reported HNF4A and HNF4G correlation coefficients because of the very low signal-to-noise ratio among these replicates.

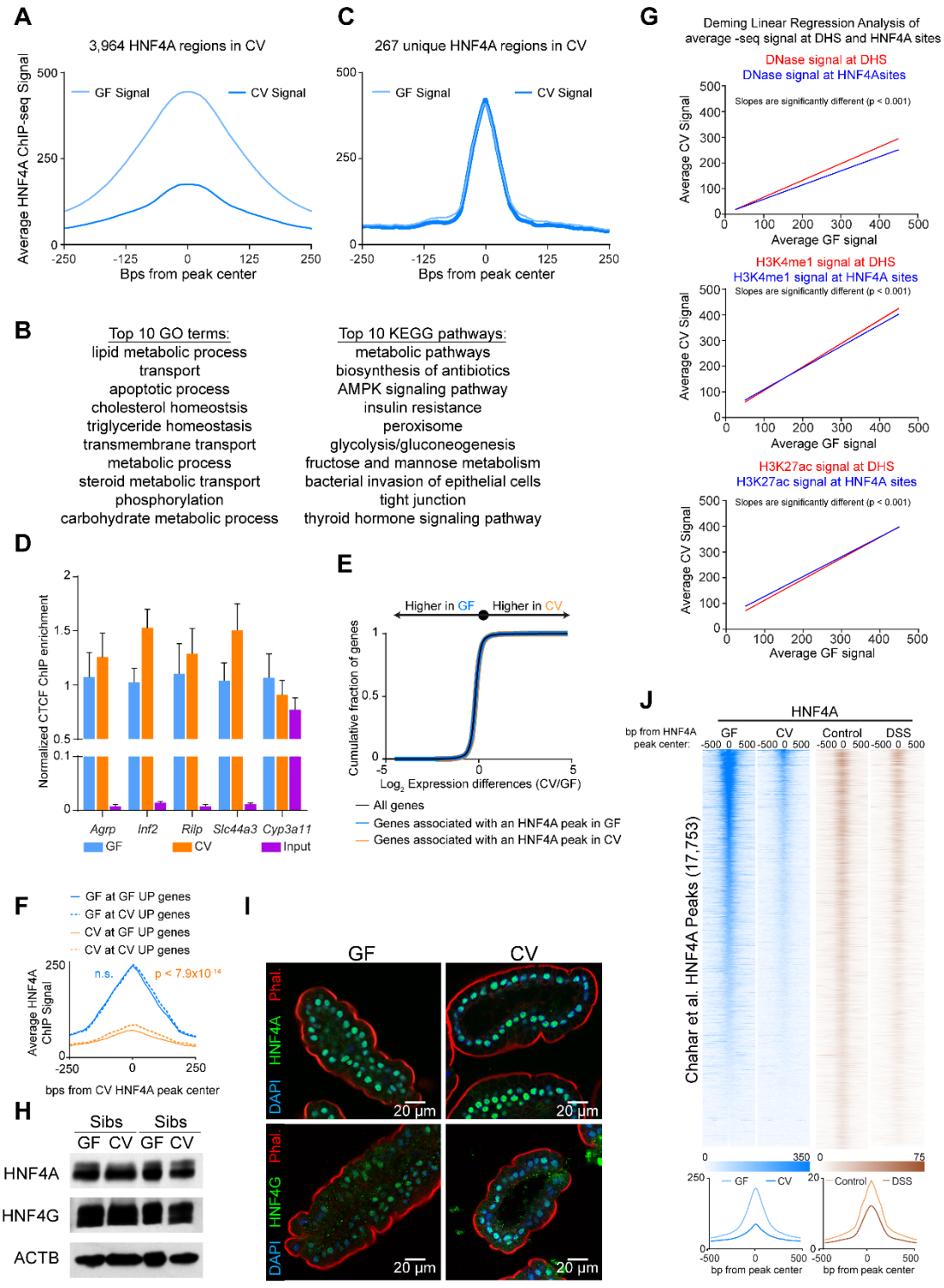


Figure 3.S6: Microbiota suppress HNF4A and HNF4G activity without overtly impacting protein levels or localization. (A) Line plots showing the average GF HNF4A (light-colored line) and CV

HNF4A (dark-colored line) -seq signal at the 500 bp flanking HNF4A sites found in CV. We chose to forgo DESeq2 analysis with the HNF4 ChIP samples because of the obvious signal bias toward GF conditions would skew the results and fail to provide useful statistical significances. We therefore performed all subsequent downstream analyses using average signal surrounding ChIP peaks. (B) Top 10 GO terms and top 10 KEGG pathways for genes associated with CV HNF4A sites. These enrichments indicate CV retains HNF4A binding near genes that are associated with canonical HNF4A and intestinal function. (C) Line plots showing the average GF HNF4A (light-colored line) and CV HNF4A (dark-colored line) -seq signal at the 500 bp flanking "Unique" HNF4A sites found in CV based on MACS2 peaks and peak coordinate intersections. (D) Bar graph of CTCF ChIP-PCR results at different loci (n = 2 per condition). Loci were chosen based on publically available mouse intestinal CTCF tracks on the UCSC genome browser. The relative CTCF enrichment at *Agrp*, *Inf2*, *Rilp*, *Slc44a3,* and *Cyp3a11* loci was normalized to CTCF signal at the *Neurog1* locus *(negative control*). Loci were chosen based on publically available intestinal CTCF ChIP-seq tracks on the UCSC genome browser These ChIP-seq tracks show CTCF binding at *Agrp*, *Inf2*, *Rilp*, and *Slc44a3* loci but not at *Cyp3a11* nor *Neurog1* loci. (E) Two-sided Kolmogorov-Smirnov goodness-of-fit test shows no relationship between the presence of an HNF4A site in GF (blue) or CV (orange) and increased transcript abundance in that same colonization state. Our zebrafish RNA-seq data predict that HNF4A directly or indirectly regulates both microbiota suppressed and induced genes. In accord, we did not find an overt association with HNF4A binding sites and microbiota suppressed or induced genes. (F) Line plots showing the average signal of GF (blue) and CV (orange) HNF4A ChIP-seq RPKM-normalized signal at the 500 bp flanking HNF4A peaks associated with microbiota-suppressed genes (solid) and microbiota-induced genes (dotted). Statistical measurements were performed using a two-tailed Mann-Whitney test. (G) Deming linear regression of the average GF and CV ChIP/DHS signals at HNF4A sites and DHS sites. To determine if histone marks correlate with the loss of HNF4A signal in CV conditions, we aligned the average histone ChIP signals and DHS signals to the 28,901 GF HNF4A sites. As expected, we found both GF and CV H3K27ac and H3K4me1 signals were enriched on the flanks of the HNF4A peaks while DHS signal was enriched near the center of the HNF4A peaks (Figure 3.4A,B). Interestingly, we observed that colonization resulted in a reduction in H3K4me1 signal at HNF4A sites, a trend we did not see when comparing signals at all DHS sites (Fig.4B and Supplemental Figure 3.S4H). Active CV enhancer signals and CV DHS signals were also reduced at HNF4A sites (Fig.4B). However, the average signals of these genomic marks were already reduced upon colonization at DHS sites (Supplemental Figure 3.S4H). Therefore, to determine if the presence of an HNF4A site corresponded to a reduction in enhancer activity or chromatin accessibility upon microbiota colonization, we performed Deming linear regression. If the slopes of the Deming linear regression are significantly different, we can conclude that the relative signal at the two different sets of genomic locations is significantly different. We found that HNF4A sites correspond with increased H3K27ac, H3K4me1 and DHS signal in GF compared to these same chromatin marks in CV. (H) Western blots of HNF4A and HNF4G from GF and CV mouse jejunal IECs. B-actin was used as a loading control. (I) Representative confocal immunofluorescence (n = 2 per condition) optical section of wholemount GF and CV mouse jejunal villi stained for phalloidin (red), HNF4A/HNF4G (green) and DAPI (blue). (J) Heat maps and line plots of showing the average GF and CV HNF4A ChIP-signal from primary mouse jejunal IECs (blue) and the single replicates of control and DSS-treated HNF4A ChIP-signal from primary colonocytes. When we assessed the average GF and CV jejunal HNF4A signals at the colonic HNF4A peaks from this previous study, we observed jejunal HNF4A signal at the majority of colonic peaks and reduced CV HNF4A averaged signal compared to GF. This finding reveals that the HNF4A cistrome in the small and large intestine is remarkably similar, and that HNF4A occupancy at many of these sites is similarly reduced by microbiota and inflammation.

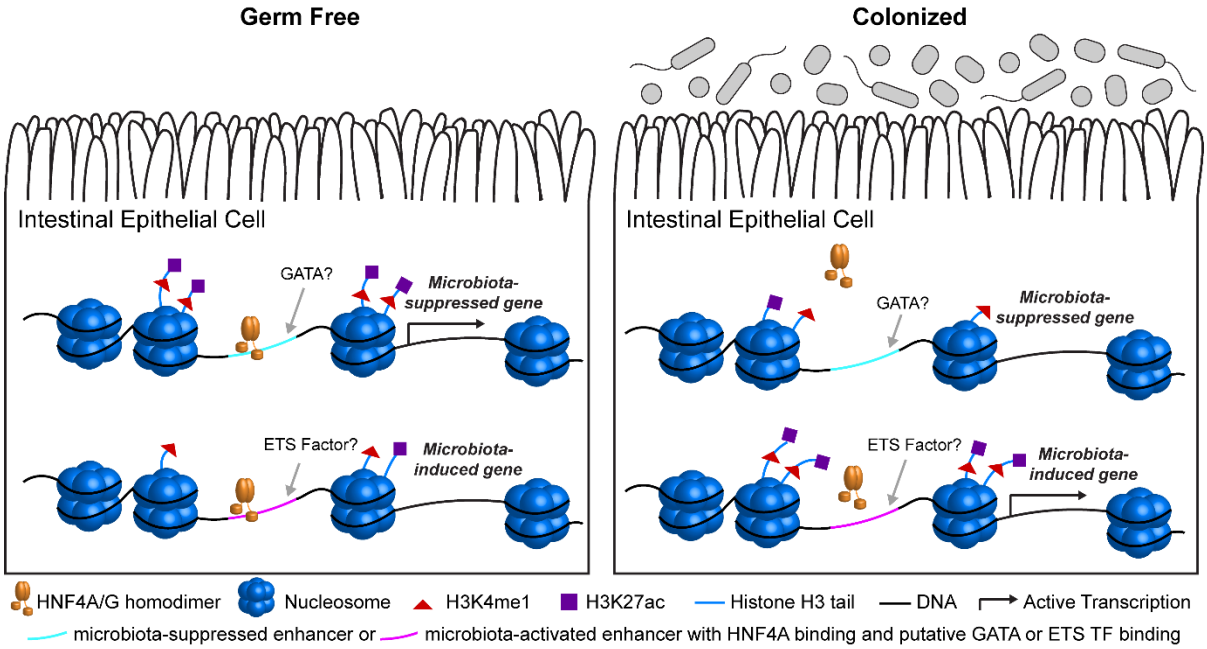


Figure 3.S7: Model of microbiota regulation of host gene transcription through modification of enhancer activity and suppression of HNF4A DNA binding. A model depicting the key findings from the manuscript. Microbiota-suppressed H3K27ac and H3K4me1 marks (microbiota-suppressed enhancers) are significantly enriched near genes that are downregulated by the microbiota. Microbiota-induced H3K27ac and H3K4me1 marks (microbiota-induced enhancers) are significantly enriched near genes that are upregulated upon microbiota colonization. Also, following microbiota colonization, HNF4A DNA binding is reduced across the genome. This reduced occupancy occurs near genes that are both microbiota-induced and microbiota-suppressed genes. GATA factor binding sites are located near HNF4A binding sites that associate with microbiota-suppressed genes and microbiota-suppressed enhancers. ETS factor binding sites are located near HNF4A binding sites that associate with microbiota-induced genes and microbiota-induced enhancers.

Supplemental Table 1: Primers and Oligos

Primer name	Sequence	
ChIP-PCR Primers		Genomic Location
Agrp_CTCF_F	CAAGGAGTACGCCGCAAGAAGTG	8:105622531-105622553
Agrp_CTCF_R	CCACAACCTAAAGTTGCTTCTTGAG	8:105622650-105622674
Slc44a3_CTCF_F	CTCAAAGGATCAAGGACTGCAG	3:121500329-121500350
Slc44a3_CTCF_R	GATCTTGCCTGCACCAACACAG	3:121500449-121500470
Rilp_CTCF_F	GAAGCCAAGAGACCGAAGCGGTGAG	11:75510719-75510743
Rilp_CTCF_R	GATCAGCAGGAGGCGCTGTAGCTG	11:75510831-75510854
Inf2_CTCF_F	GTAGCTACCATAGTCTTATCTAAG	12:112605998-112606021
Inf2_CTCF_R	GCTGTCTTCCCTTGACTTGG	12:112606087-112606106
Mm_Neurog1_F	GGCTACATTTGGTCTTTATCC	13:56247085-56247105
Mm_Neurog1_R	GTGGAGCCAAGCTAACAATTTGC	13:56247171-56247193
Mm_Apoa1_F	CTAGGGAGTTGGGGAGTTTCCT	9:46227267-46227288
Mm_Apoa1_R	TCTCTCAGCCTTAGAGGCAAGG	9:46227355-46227376
Mm_Angptl4_F	ATCTAATCTACAGTCCATATTCCAC	17:33788903-33788927
Mm_Angptl4_R	AGGGCATCAATGCAAAGTGCAGTG	17:33788988-33789011
Mm_Pck1_F	CCAGGTTGCAGAAAGGAGTGTC	2:173138399-173138420
Mm_Pck1_R	AGAATGTGGTAAACAGGACTCAAG	2:173138479-173138502
qRT-PCR Primers		
hnf4a2.1 F	CAGTGTCGGTACTGCAGACTAAAG	
hnf4a2.1 R	GTGAGCTCGCAGTAAAGCCACCTG	
hnf4b1.1 F	AGACCGAGCCACTGGAAAAC	
hnf4b1.1 R	CATGTGTAGGCATGGTTCTTG	
hnf4g2.1 F	ATGAAGTTTTCTCCAACTCCTCTCC	
hnf4g2.1 R	CTGCTGTGAAAGTGCTTCAGCGTGAGC	
saa F	CGCAGAGGCAATTCAGAT	
saa R	CAGGCCTTTAAGTCTGTATTTGTTG	
gfp F	GAAGAAGTCGTGCTGCTTCA	
gfp R	CCTGAAGTTCATCTGCACCA	
angptl4 F	CGAGCGCATCAAGCAACA	

angptl4 R	TCGCTCGTTTTTCATCGTAATCT
ifabp F	TGCCCATGACAACCTGAAGA
ifabp R	GTTAATTTCCAGTGTGCGGAAAG
18S F	CACTTGTCCCTCTAAGAAGTTGCA
18S R	GGTTGATTCCGATAACGAACGA
elf1a F	CTTCTCAGGCTGACTGTGC
elf1a R	CCGCTAGGATTACCCTCC
Primers for Yeast-1-Hybrid Screen	
in3.4 F	AAAAGAGCTCCCTTGTAGGCTGTTGGAAATAC
in3.4 R	AAAACTCGAGACTGAAAGACACAAACACA
hnf4a_pENTR F	CCGCCCCCTTCACCATGGAGATGGCAGACTATAGCGAG
hnf4a R	TCGGCGCGCCCACCCTTTCAGATGGCCTCTTGTTTAGTG
hnf4g F	CCGCCCCCTTCACCATGGATGTAGCCAATTACTGCGA
hnf4g R	TCGGCGCGCCCACCCTTTCATAGCGGGGGCTCCGGAGA
Oligos for guide RNAs	
hnf4a_gRNA F 1	TAGGGCACCAGAAGATCCAGCTATG
hnf4a_gRNA R 1	AAACCATAGCTGGATCTTCTGGTGC
hnf4a_gRNA F 3	TAGGGTAAGCTGCTGTCCTCATAGC
hnf4a_gRNA R 3	AAACGCTATGAGGACAGCAGCTTAC
hnf4a_gRNA F 4	TAGGGTCCTCATAGCTGGATCTTC
hnf4a_gRNA R 4	AAACGAAGATCCAGCTATGAGGAC
Primers for Hnf4a mutation screening	
hnf4a_Cris_checkF	TGATTCACACTACTTACTTGTCTAG
hnf4a_Cris_checkR	GATTAAAAGTAGTTATCTCATCCTCAG

Supplemental Table 2: Total number of MACS2 peak calls per ChIP per replicate

Number of peaks indicated are raw results generated from MACS2 and peaks generated by background signal/noise have not been filtered out. All peaks generated by background sequencing noise were removed manually for downstream analysis.

	DHS	H3K4me1	H3K27ac	HNF4A	HNF4G
GF rep 1	89,507	132,275	82,935	36,850	12,465
GF rep 2	61,355	131,876	83,336	27,016	35,302
GF rep 3		137,770		29,070	15,934
GF rep 4					27,816
CV rep 1	70,794	139,559	83,661	1,889	101
CV rep 2	52,431	144,137	81,997	8,473	18
CV rep 3	42,433	146,199		4,195	27
CV rep 4					106
GF Average	57,485	133,974	83,136	30,979	22,879
CV Average	55,219	143,298	82,829	4,852	63
GF Median	57,485	132,275	83,136	29,070	21,875
CV Median	52,431	144,137	82,829	4,195	64

CHAPTER 4: PROSPECTUS

4.1 Introduction

The microbiota contribute to human physiology by facilitating energy harvest, tuning metabolic programs, and promoting immune system development. In addition to these important roles in health, intestinal microbiota have been implicated in a growing number of human diseases associated with loss of intestinal epithelial identity, like cancers and inflammatory bowel diseases [221, 284]. However, these diseases represent extreme circumstances and do not reflect the normal symbiotic relationships that have been maintained between animals and their microorganisms for over 650 million years. From an evolutionary perspective, the presence of microbiota is a normal part of the animal's life cycle, and physiological states have evolved to assume the presence of microbes. Extensive research has identified microbial factors and aberrant host responses that impair intestinal epithelial function. Similarly, thought the impact of the microbiota on host IEC transcriptomes and their downstream consequences have been extensively documented, the upstream transcriptional regulatory mechanisms remain poorly understood. Recently, this gap in knowledge has started to be filled with both my work and others that have identified differences in histone modifications following microbiota colonization [141, 142, 334]. However, my work is the first to show the impact of microbiota colonization on the binding activities of a transcription factor. This finding has provided a novel genomic mechanism for understanding how the microbiota tune intestinal epithelial cell transcription programs and provides a potential model for how dysregulation of these same transcription programs may lead to human disease.

The identification of HNF4A as a novel transcription factor that mediates the epithelia's response to the microbiota provides new insight into nuclear receptor biology, host-microbiota interactions and intestinal pathophysiology. However, my work only provides a genomic mechanism for the host response. That is, I only show that genome wide, HNF4A activity is suppressed. We still do not understand the signaling mechanisms that mediate HNF4A activity. Since HNF4A is a nuclear receptor, does microbial colonization result in changes in ligand availability? Does PRR signaling regulate HNF4A DNA binding? These are questions we are now poised to address, but due to complexity of nuclear receptor regulation, they were beyond the scope of my initial effort in characterizing mechanisms of transcriptional control. We first needed to identify a candidate transcription factor that mediates epithelial response. The identification of HNF4A as a regulator of the host response is profound. Loss of HNF4A function and the microbiota are both linked to human diseases like metabolic syndrome and IBDs. Further, HNF4A is the most ancient of the nuclear receptor transcription factors, so framing this transcription factor in a new role could provide insight into the evolution of nuclear receptor biology and host-microbiota interactions.

In this last chapter of my dissertation, I discuss several possible mechanisms that may mediate HNF4A activity. I speculate how suppression of HNF4A is advantageous to the host and to some extent the microbiota. I speculate how human disease pathology may be determined by microbial suppression of HNF4A. Finally, I speculate on the evolution of the nuclear receptor superfamily as novel mediators in the host response to the microbiota.

4.2 Models of microbiota suppression of HNF4A activity:

Previous *in vivo* and cell culture studies indicate HNF4A serves as both an activator and repressor of target gene transcription. In the third chapter of this dissertation, we learned that HNF4A binds at loci of both microbiota-suppressed and microbiota-induced

genes. And we learned that microbiota colonization is associated with loss of HNF4A occupancy on the genome. Therefore, HNF4A may activate microbiota-suppressed genes and may repress microbiota-induced genes. Transcription factors commonly have bimodal functions, like HNF4A, where they both activate and repress transcription [335]. These different modes are sometimes governed by cofactor binding that designate the transcription factor as a repressor or activator at a locus [336]. However, the kinetics of transcription factor binding can also mediate repressor or activator assignment. For instance, RAP1, a well-studied yeast transcription factor, occupies several different loci on the yeast genome. However, the kinetics of RAP1 binding differ at these loci. These kinetic properties of RAP1 binding are associated with repressor or activator assignments. RAP1 may “treadmill” at genes it represses, meaning it has short residency time on the DNA [337]. However, RAP1 has long residency at genes it activates. By conventional ChIP, these two modes of function are associated with approximately the same occupancy signal. A technique called “competition-ChIP” reveals kinetics of DNA binding and the kinetic associations with repressor vs activator assignments [338].

Perhaps, HNF4A adopts a “treadmilling” mode of function at microbiota-induced genes and maintains a long DNA residency at microbiota-suppressed genes (further discussed below under 4.2.2). Whatever mode of function HNF4A initiates across the genome, microbiota colonization is associated with loss of HNF4A genome occupancy and thus *trans*-activity (either repressor or activator). However, the mechanism that mediates this loss of HNF4A occupancy remains unknown. Do microbes affect HNF4A activity through a microbial-derived metabolite/molecule that directly binds and blocks HNF4A function? Or does the upregulation of TLR signaling following microbiota colonization result in loss of HNF4A function? Perhaps HNF4A activity is suppressed by the microbiota indirectly through a reshaping of the metabolic programs. Below I explore several models that describe how the microbiota may suppress HNF4A.

4.2.1 Ligand binding:

The HNF4A ligand binding domain contains a single fatty acid (although other papers have demonstrated it may also bind flavonoids [339]). Originally labeled as an orphan receptor, in non-physiological conditions, researchers have found HNF4A can bind several different species of long chain fatty acids [308]. Although these putative ligands may be valid HNF4A ligands, their role in controlling and mediating HNF4A activity in physiological conditions remains unknown. In the mouse liver, HNF4A binds linoleic acid (C18:2) almost exclusively [316]. Some papers have suggested linoleic acid may suppresses HNF4A activity in cell culture lines [316, 340]. However, linoleic acid does not suppress transcription of HNF4A target genes, like *fabp2*, in the Germ free (GF) zebrafish intestine (data not shown), suggesting dietary linoleic acid does not suppress HNF4A activity. However, similar to how cofactors regulate transcriptional activity, ligand binding may provide a mechanism for controlling HNF4A. Perhaps different fatty acid ligands tune HNF4A activity so that small conformational adjustments on the protein modify the affinity for the same binding site. Perhaps conformational changes impact HNF4A's ability to bind DNA altogether or inhibit its ability to form homodimers, a necessary step for *trans*-activity. Specific taxa within the microbiota express proteins called linoleic acid isomerases. This family of enzymes edits the location of the double bond in linoleic acid, generating a new isomer of LA (18:2; 9-cis, 12-cis) called conjugated linoleic acid (CLA – 18:2; 9-cis, 11-trans and 18:2,10-tras, 12-cis). Importantly, even though this fatty acid has the same atomic makeup, the location of the double bonds and the double bond isomer (i.e. cis vs trans) greatly impact the structure of the fatty acid. It remains to be determined if HNF4A binds CLA. However, if HNF4A can bind both CLA and LA independently, these fatty acids will likely fill the ligand binding pocket differently due to their differences in chemical structure. Thus, HNF4A may need to change confirmation to bind CLA versus LA which may modify its activity (Figure 4.1A). It is also certainly possible that the microbiota regulate HNF4A activity through the modification of

other putative ligands like flavonoids which are known to modulate microbiota composition [341].

For an unbiased approach to identify if the microbiota impact the FA species bound by HNF4A, I recommend performing GC/MS on organic compounds isolated from immunoprecipitated HNF4A from gnotobiotic mouse IECs. My initial attempts at performing these experiments failed but similar experiments from mouse liver provide a blueprint for success [316]. The major limiting factor for performing these experiments is the starting material. To detect the ligand, the fatty acid should ideally be at a concentration of 50 ng/mL (personal correspondence with George Dubay at Duke University). Assuming a 1:1 stoichiometry between HNF4A and the fatty acid (meaning HNF4A binds only one ligand at a time), pulled down HNF4A concentration must be 9.4 ug/mL (based on the molecular weight of HNF4A and LA: 52,785 g/mol and 280 g/mol respectively). Previous published experiments used 3 mouse livers for each pull down to concentrate HNF4A to a sufficient level [316]. Therefore, based on the number of cells isolated during an IEC extraction, I predict we would need approximately 3 – 4 gnotobiotic mice for a single replicate to perform the equivalent experiment from isolated jejunal IECs. Pilots of this experiment could be performed with thin-layer chromatography (TLC). If HNF4A binds different ligands in germ free and colonized conditions, these fatty acids may migrate at different speeds with the correct organic solutions. However, due to the similar chemical structures, TLC is not an appropriate assay to address the LA and CLA hypothesis). Another key aspect for success with these experiments is the efficacy of the antibody used for the immunoprecipitation. Antibody choices will be discussed in a future section under post-translation modifications.

I also propose three experimental models that test if different fatty acids have the capacity to suppress or modify HNF4A activity: 1) Perform qPCR from zebrafish that have been fed a single fatty acid species, like oleic acid (18:1), LA (18:2) or CLA (18:2). This experiment may identify if HNF4A target genes are differentially transcribed during feeding

with these various fatty acids. 2) Perform ChIP-PCR or ChIP-seq from gnotobiotic mice that have received an oral gavage of CLA or another fatty acid species. This experiment may identify if HNF4A binding at specific locations on the genome is impacted by specific fatty acids. 3) Test the HNF4A activity in the presence of different fatty acids in a quantitative Yeast-1-Hybrid (Y1H) assay. I have piloted this experiment in the original Y1H system that identified HNF4A as a potential microbiota-regulated host transcription factor. However, incubating yeast with LA, OA, or CLA conferred resistance to the antibiotic and increased the noise of the assay. Instead, I recommend an assay that utilizes a luciferase reporter instead of antibiotic resistance. The luciferase reporter will provide more quantitative measurement of HNF4A activity in the presence of different fatty acids. All three of these experiments may provide new insight into how HNF4A responds to specific putative fatty acid ligands. Unlike the proposed GC/MS experiment, these fatty acid feeding experiments are biased and will not inform us if the microbiota impact HNF4A ligand binding. Rather, these experiments provide a potential model for how the microbiota mediate HNF4A activity.

4.2.2 Splice form abundances:

In humans, the HNF4A gene encodes 9 different splice forms [342]. Expression of these different splice forms is driven by two promoters (P1 and P2). HNF4A splice form variation and the mutations within the possible splice forms contributes to human disease pathology including Crohn's disease and diabetic phenotypes [290, 343]. The different splice forms differ in transcriptional activity and modes of coactivation. For instance, splice forms transcribed from the P1 promoter include the "AF-1 domain", a conserved n-terminal domain found in several members of the nuclear receptor super family. This domain regulates cofactor binding and increases *trans*-activation potential [344]. Therefore, HNF4A splice forms transcribed from the P1 promoter may interact with different coactivators than splice forms transcribed from the P2 promoter [345].

The antibody that I used in my studies for ChIP only detects HNF4A splice forms transcribed from the P1 promoter. However, the antibody I used in the western and immunofluorescence data detects all possible splice forms. Therefore, the ChIP data only represent a subpopulation of HNF4A and the western and immunofluorescence data includes all possible HNF4A splice forms. Perhaps the microbiota suppress transcription of *Hnf4A* at P1 by activating a known repressor of the P1 promoter, SREBP2 [346]. Or, perhaps the microbiota induce HNF1A activity, a known activator of transcription at the P2 promoter [347] (Figure 4.1F). However, my current data do not support these hypotheses. DEX-seq analysis of RNA-seq data from GF and CV mouse IECs indicate there in no difference in exon bias between GF and CV mice at the *HNF4A* gene locus. Perhaps the downregulation happens at a protein level. However, the western and immunofluorescence I performed in chapter 3 do not indicate reduced protein levels upon colonization. Perhaps different transcriptional cofactors become available upon colonization and preferentially activate P2 isoforms. To test if P2 isoforms preferentially occupy HNF4A binding sites on the genome upon microbiota colonization, I recommend performing ChIP-seq using an antibody that recognizes all possible isoforms. The initial polyclonal antibody I tested for the ChIP-seq experiments detects all isoforms, however this antibody failed to immunoprecipitate HNFA. I therefore chose a monoclonal antibody that only detects 6 HNF4A isoforms, but had been previously shown to successfully immunoprecipitate HNF4A from intestinal epithelia.

Finally, microbiota colonization may promote “treadmilling” activity of HNF4A splice forms. If HNF4As transcribed from promoters 1 and 2 begin to compete for the same binding site, the conventional ChIP I performed would only detect P1-splice-form DNA-binding and therefore HNF4A occupancy would appear reduced. It would be interesting to perform true competition ChIP by using a ChIP antibody for P1 splice forms and a second antibody for P2

splice forms. These data would provide a new level of insight on HNF4A DNA binding kinetics that have never been attempted, let alone in a gnotobiotic setting.

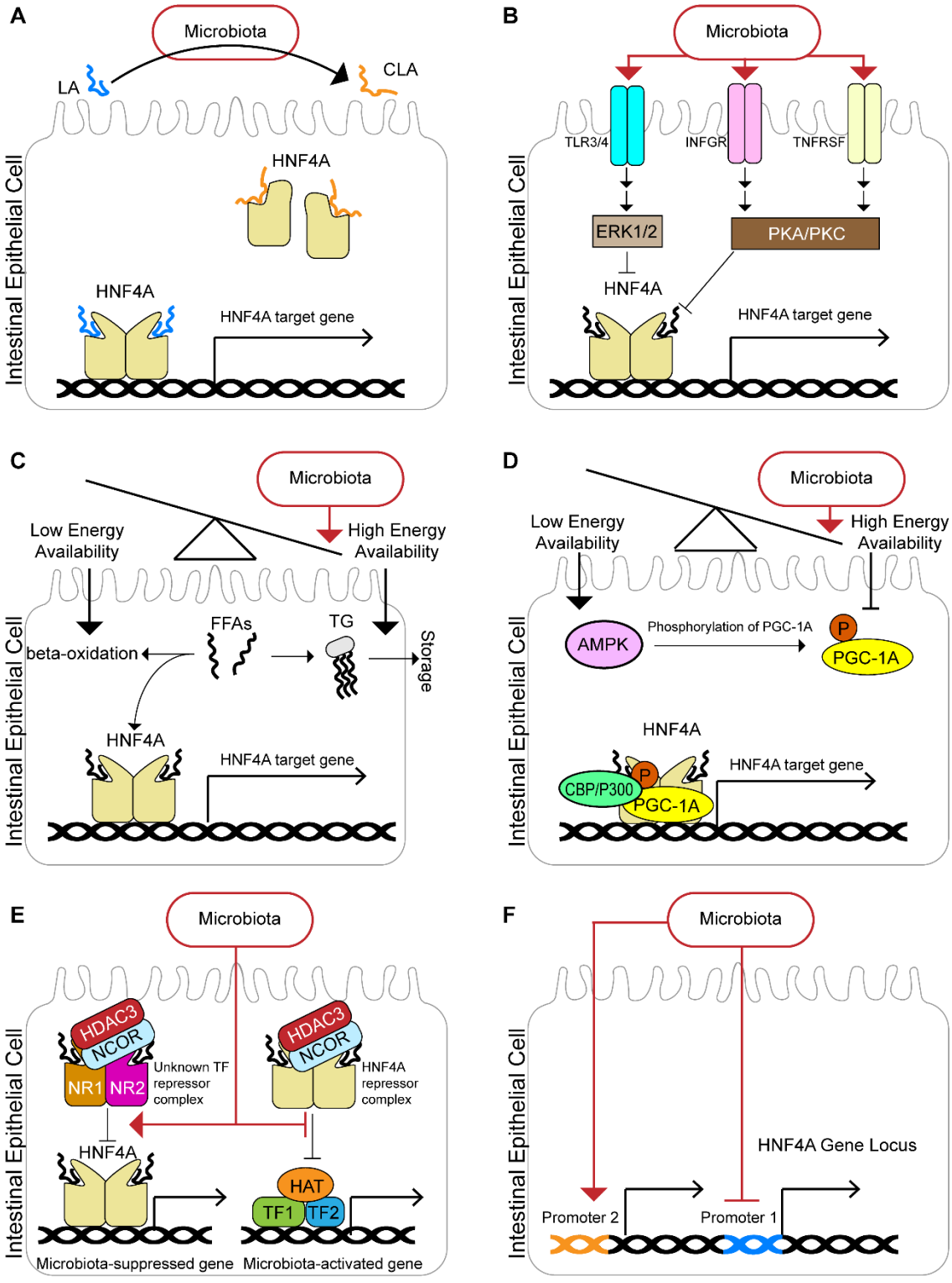


Figure 4.1: Six possible mechanisms that regulate suppression of HNF4A activity associated with microbiota colonization. (A) The first model predicts that the microbiota modify the

endogenous ligand of HNF4A from linoleic acid (LA) to conjugated linoleic acid (CLA). Because CLA is structurally unique from LA, HNF4A may need to change protein confirmation to bind it and as a result be incapable of binding DNA. (B) The second model predicts HNF4A might be phosphorylated by PKA, PKC or the MAPK pathway following activation of transmembrane receptors that activate these kinases. For instance, TLR4 signaling may initiated a MAPK signal transduction, or TNFa or IFNg might activate their corresponding receptors to which may activate these kinases. Phosphorylation of HNF4A reduces its potential to activate transcription. (C) The third model predicts microbiota colonization results in a rewiring of fatty acid handling within the intestinal epithelial. Instead of being consumed by the cell for energy via beta-oxidation, the cell may preferentially package dietary fatty acids in triglycerides and store them. Therefore, the fatty acids are unavailable to serve as ligands for HNF4A. Active beta-oxidation may increase the availability of fatty acids for HNF4A to bind and thus increase its potential for transcriptional activation. (D) The liver and skeletal muscle of GF mice have increased activated-AMPK compared to CV mice. The fourth model assumes that IECs also have increased activated-AMPK levels in GF mice. This increase in AMPK activity may result in increased PGC-1A activity, a known target of AMPK. PGC-1A is a known coactivator of HNF4A. Therefore, upon microbiota colonization, HNF4A-mediated transcriptional activation might be reduced due to reduced AMPK and thus PGC-1A activity. (E) The fifth model predicts that HNF4A is replaced by other transcription factors on the genome in complex with known corepressors, like NCORs. It similarly suggests that HNF4A complexes with NCORs at microbiota induced-genes, which might indicate why these genes are suppressed in GF conditions. (F) The last model predicts that HNF4A activity is reduced because of a change in splice form/exon usage. Perhaps microbiota colonization results in increased usage of the HNF4A promoter 2. My ChIP antibody did not detect splice forms generated from this promoter. Therefore, my ChIP HFN4A occupancy may be reduced because of less frequent binding of HNF4A from promoter P1.

4.2.3 Energy Balance: Fatty acid availability:

Microbiota colonization induces significant shifts in metabolic programs in several tissues in the mice. These metabolic shifts occur at both the transcriptional level and protein level. Following colonization, intestinal tissue and liver transcriptomes shifts from gluconeogenesis and fatty acid lipolysis programs to glycolysis, lipogenic, amino acid and nucleotide metabolic program [119, 120, 169]. This shift in metabolic programs occurs concurrently with increased serum glucose levels and glycogen production following colonization [169]. Compared to GF mice, Livers and skeletal muscle of colonized (CV) mice have reduced active-AMPK levels, and reduced *Cpt1a* expression and activity, an enzyme involved in the rate limiting step of beta-oxidation. These data indicate that beta-oxidation may be the primary metabolic program in GF mice. Because of increased beta-oxidation in GF mice, fatty acids will preferentially be consumed for energy production rather than packaged and stored. In support, previous work in zebrafish and mice shows microbiota promote lipid droplet accumulation in enterocytes and in extra intestinal tissue [168] [and

data not shown]. Perhaps GF animals accumulate fewer lipid droplets upon lipid feeding because the lipid is preferentially directed toward beta-oxidation and not toward storage. Given all of the data and previous studies that show the shift in metabolic activities following colonization, this is the model I favor for HNF4A regulation.

Since lipids enter the beta-oxidation cycle as individual fatty acids, more free-fatty acids may be available within cells of GF animals compared to CV animals. As a results, HNF4A may bind one of these free fatty acids as a ligand and become active. Therefore, HNF4A activity may be higher in GF animals because of increased free-fatty acids (Figure 4.1C). This hypothesis is in accord with previous models of HNF4 activity [308]. To test this hypothesis, we can try to capture and detect the total free-fatty acids within IECs from GF and CV mice. I also recommend performing GC/MS and TLC experiments (discussed in 4.2.1) which may elucidate if HNF4A protein from GF mice binds more fatty acid compared to CV mice. Finally, chemical screens of agonists or antagonists for enzymes involved in lipolysis, gluconeogenesis and lipogenesis in GF zebrafish may provide some insight into the mechanism controlling HNF4A.

4.2.4 Energy Balance: AMPK activity and PGC-1A:

As mentioned in the previous section, phosphorylated AMPK (activated-AMPK) levels are reduced in skeletal muscle and liver in CV mice compared to GF mice. AMPK functions as a metabolic rheostat which regulates several metabolic pathways based on AMP levels. AMPK activity suppresses cellular proliferation and lipolysis and activates mitochondrial biogenesis and fatty acid oxidation. It performs these functions through phosphorylation and inhibition of Acetyl-CoA Carboxylase (ACC), which directs Acetyle-CoAs to lipogenesis, and phosphorylation and activation of PGC-1A, which functions as a transcriptional coactivator [348]. PGC-1A binds the HNF4A homodimer at multiple locations and promotes stability on DNA and induces transcriptional activity [312]. I recommend

performing phospho-PGC-1A westerns to confirm microbiota suppress PGC-1A in this tissue as well. If PGC-1A is more active in GF IECs, like it is in the liver, perhaps microbiota colonization suppresses HNF4A DNA binding by reducing activity of PGC-1A (Figure 4.1D).

When bound to a transcription factor, the cofactor PGC-1A can also recruit CREB Binding protein (CBP/P300) to genomic locations. CBP/P300 is a histone acetyltransferase enzyme that modifies lysine 27 of Histone 3 to generate H3K27ac. CBP/P300 also acetylates HNF4A at lysines 97 and/or 99 [310]. Therefore, a model in which the microbiota suppress PGC-1A activity provides mechanisms for two genomic observations from Chapter 3: 1) Perhaps microbial colonization inactivates PGC-1A in IECs and thus reduces HNF4A stability on DNA. 2) Locations bound by HNF4A had statistically higher H3K27ac signal compared to open chromatin regions that were not bound by HNF4A. Perhaps PGC-1A recruited CBP/P300 to regions bound by HNF4A. Once recruited to these regions, CBP/P300 could acetylate nearby H3K27 and thus increase H3K27ac signal around HNF4A bound enhancers. Other cofactors and other transcription factors are perhaps more "permitted" to bind and coactivate these enhancer elements following H3K27ac modification.

To test if reduced PGC-1A activity mediates microbial suppression of HNF4A, I recommend performing co-immunoprecipitations from IEC nuclear extracts from GF and CV mice. Perhaps PGC-1A only co-immunoprecipitates with HNF4A in GF conditions. This experiment could also be scaled up to perform mass spectrometry which would provide an unbiased approach for testing cofactor binding. Mass spectrometry also permits detection of post-translation modifications therefore allows testing of the impact of microbiota colonization on HNF4A PTMs.

Co-immunopreciptation experiments, particularly from in vivo tissue, can be technically challenging, especially if the starting material is limiting. Therefore, these experiments could also be tested *in vivo* using the zebrafish model. Transgenesis of zebrafish coupled with genetic mutation provides an opportunity to quickly test the function

of multiple genes and mutant proteins. For these experiments, I recommend generating a transgenesis construct that will encode and induce expression of wild-type *hnf4a* and rescues the *hnf4a*$^{-/-}$ zebrafish line (Figure 4.2). With this construct and with the power of site-directed mutagenesis, we can test if Hnf4a with a single amino acid mutations maintain the capacity to rescue the *hnf4a*$^{-/-}$ phenotype. The simplest assay to test for function of the transgenic and mutant Hnf4a is qRT-PCR, specifically assaying genes that are known Hnf4a targets. Mutations at amino acids that interact with PGC-1A recognizes will inform us of the importance of PGC-1A in regulating HNF4A activity [312]. Phospho-mimetic mutations or another amino acid substitution at locations of post-translational modifications will provide insight into which kinases or other protein modifying enzymes are regulating Hnf4a activity (see 4.2.6). Finally, these experiments can be moved into a gnotobiotic system to test if the microbiota impact the activity of these transgenic and mutant hnf4a zebrafish.

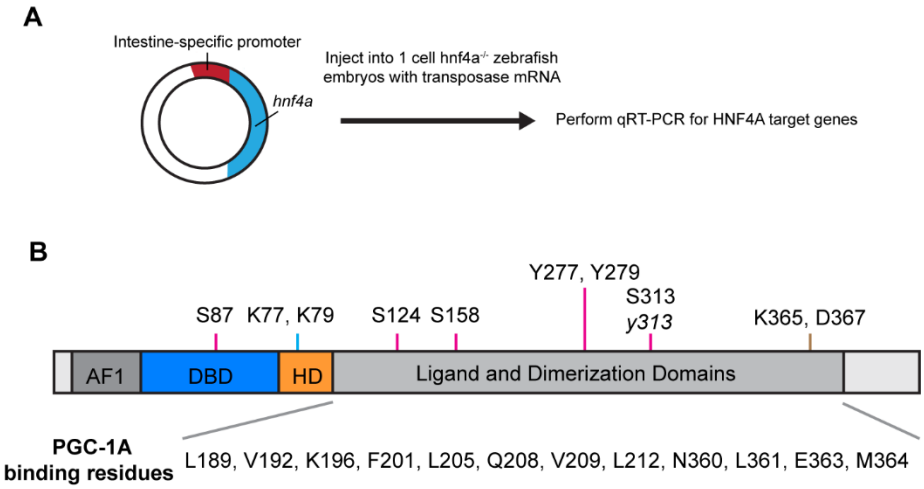


Figure 4.2:Transgenic screening strategy to test for the influence of PTMs and PGC-1A binding on Hnf4a activity in zebrafish. (A) A simple plasmid that includes an intestine specific promotor driving expression of hnf4a. This plasmid can be injected into *hnf4a*$^{-/-}$ zebrafish embryos w/ transposase to generate transgenic animals and to test if intestinal expression of hnf4a is sufficient to rescue *hnf4a*$^{-/-}$ mutant zebrafish from the IBD-like gene expression signatures. The hnf4a gene within the plasmid can be modified via SDM to introduce amino acids that mimic or inhibit PTMs or inhibit PGC-1A from binding. (B) An Hnf4a protein schematic with functional domains annotated. Amino acids with pink lines have been shown to be phosphorylated [349-352]; all annotated residues are conserved between mammals and zebrafish except for S313, which is Y313 in zebrafish (annotated as y313 in the diagram). The K77 and K79 (blue line) have been shown to be acetylated by CBP [310] and K365 and D367 (brown line) have been shown to be SUMOylated [353], which promotes

protein degradation via a ubiquitination pathway. The 12 amino acids below the protein schematic have been shown to directly interaction with PGC-1A via protein crystallography [312].

4.2.5 Nuclear receptor repressors and HDACs:

Nuclear receptors function as both transcriptional activators and repressors [354, 355]. Repression of genes is facilitated by a class of proteins called corepressors which bind to the transcription factors and form a repressor complex. This repressor complex occupies *cis*-regulatory regions, and blocks the transcription of a target gene. A family of transcriptional repressors called Nuclear receptor Corepressors (NCOR1 and NCOR2) bind transcription factors (not just nuclear receptors, as suggested by their name) and repress transcription of a target gene. NCORs function together or individually with several different transcription factors at many loci with varying degrees of repressor activities [356]. NCOR2 interacts directly with HNF4A and recruits histone deacetylases to repress transcription of genes [357]. No studies show a direct interaction between NCOR1 and HNF4A, but NCOR1 does repress several genes in the HNF4A regulon through VDR-mediated repression [358].

Microbiota colonization results in significant induction of *NCOR1* and *VDR* expression in the small intestine [119]. Perhaps these factors form a repressor complex and replace HNF4A on the genome, thereby suppressing HNF4A activity and repressing HNF4A target genes. This model may be unlikely since VDR, like HNF4A, has been shown to promote transcription of fatty acid oxidation genes, which are suppressed upon microbiota colonization. However, members of the PPAR and RAR family of nuclear receptors also interact with NCORs [359]. Like HNF4A, PPARs and RARs recognize “DR2” (see figure 3.3E) DNA sequence motifs, indicating these transcription factors have the capacity to replace HNF4A at the same genomic location and perhaps recruit the NCOR corepressor. In support, the microbiota stimulate PPARG-mediated transcription in the colon [360]. Perhaps microbes similarly promote PPARG-repressor activity which could then replace HNF4A on the genome. Regardless of the specific transcription factors, I propose a model in which a

repressor complex replaces HNF4A on the genome and represses transcription of HNF4A target genes (Figure 4.1E). In this model and others, it is important to note that HNF4A target genes are not completely repressed; and similarly, HNF4A activity is not completely inhibited. Instead, these models provide a mechanism that may attenuate HNF4A activity and its target gene transcription.

NCORs also recruit histone deacetylases (HDACs) to chromatin which then deacetylate H3K27ac, making the chromatin less “permissive” for transcription factor binding and thus reducing enhancer activity. Therefore, if the NCOR models above are correct, following replacement of HNF4A by a repressor complex, HNF4A bound regions will have reduced H3K27ac signal. Indeed, microbiota-suppressed enhancers are significantly enriched for HNF4A binding sites, indicating these may be sites where HNF4A was actively promoting transcription prior to colonization (see Figure 3.3F). However, following transcription, a repressor complex replaced HNF4A and reduced enhancer activity. Similarly, HNF4A also functions as a transcription repressor with NCOR2; perhaps microbiota-induced genes are repressed by HNF4A [357]. Following microbiota colonization, perhaps the HNF4A-NCOR2 repressor complex is replaced by transcription factors that activate the gene and recruit histone acetyltransferases. This model explains why microbiota-activated enhancers contain HNF4A sites (see Figure 3 in Chapter 2).

4.2.6 Post-translational Modifications:

Post-translational modifications on HNF4A can both activate and suppress its activity by either impacting its ability to bind DNA or by destabilizing homodimer formation. CREB binding protein (discussed in 4.2.4) acetylates HNF4A and promotes HNF4A binding affinity for DNA [310]. In a human embryonic stem cell based model of hepatocyte differentiation, HNF4a becomes modified by SUMOylation on the c-terminus. Following SUMOylation, HNF4A becomes targeted for degradation by RNF4-mediated ubiquitination [353]. Four

kinases have been shown to phosphorylate HNF4A in either *in vitro* and cell culture conditions. AMP-activated kinases (AMPK) phosphorylates HNF4A and destabilize homodimerization and DNA binding affinity [311]. Phosphorylation of HNF4A by Protein Kinase A (PKA) inhibits recruitment to its target genes and blocks nuclear localization [349]. Protein Kinase C (PKC) phosphorylation similarly blocks HNF4A nuclear localization and targets it for degradation via the proteasome pathway [350]. The Mitogen Activated Protein Kinase (MAPK) pathway also inhibits HNF4A activity [352]. A recent paper suggests the ERK1/2 may be the kinases that suppress HNF4A activity following MAPK signaling [351]. Excitingly, the microbiota stimulate the transcriptional activity of the nuclear receptor PPARG through phosphorylation by ERK1/2 [360]. Perhaps the microbiota suppress HNF4A through the same signaling pathway.

Microbiota colonization may activate the MAPK pathway through TLR4, which binds microbe associated molecular patterns like LPS. Similarly, cytokine receptors, like TNFSFR and IFNGR, and growth factor receptors may also activate MAPK signaling; however, these receptors may also activate PKC and PKA. The activation of these signaling cascades may result in phosphorylation of HNF4A and thus reduce HNF4A activity (Figure 4.1B). However, I predict neither PKA nor PKC are involved in HNF4A suppression upon colonization since phosphorylation by these kinases causes HNF4A to localize to the cytoplasm. Immunofluorescence from jejunal villi indicate HNF4A remains in the nucleus of IECs following colonization (See Supplemental Information in Chapter 3). The mode of inhibition by MAPK activity on HNF4A remains unknown, but perhaps ERK1/2-mediated inhibition does not alter nuclear localization and only impacts HNF4A DNA binding affinity. Excitingly, LPS treatment does suppress transcription of HNF4A target genes [361]. LPS activates TLR4 which signals through both a MYD88 independent and MYD88 dependent pathways (See 2.4.2). MAPK activation by TLR4 is dependent on MYD88 function. Loss of MYD88 in mice does not impact the suppression of most HNF4A target genes following microbiota

colonization [119], indicating TLR4 signaling through the MAPK pathway is not necessary for the microbiota to suppress HNF4A. However, MAPK pathways may become active through other receptors following microbiota colonization.

Microbiota colonization also induces activation of other receptor signaling cascades, like TNFSFR. TNFA, which binds TNFSFR, signaling also activates MAPK pathways, along with PKC [362]. Perhaps activation of these membrane receptors results in suppression of HNF4A by a post-translational modification. To test if the microbiota induce post-translational modifications on HNF4A, I recommend proteomic analysis both from mouse tissue and using transgenic zebrafish. Please see Section 4.2.4 for details regarding experimental design. We similarly do not know if these kinases are more active in a colonized state. So I recommended performing western blots for phosphorylated (activated) forms PKA, PKC, ERK1/2 from IECs from gnotobiotic mice. If one of these kinases are more active following colonization, I recommend treating GF zebrafish with kinase agonists like 8-bromo-cAMP [363], which selectively activates PKA. GF fish treated with this chemical may have reduced expression of HNF4A target genes compared to untreated germ free controls if PKA suppresses HNF4A activity. Similar experiments can be performed for the other kinases using both agonists and antagonists. These experiments could also be coupled with the transgenic zebrafish experiments detailed in Section 4.2.4 to provide specific information about where the PTMs are located on HNF4A.

4.2.7 Combination of several of the models

Gene expression is a nuanced biological process with several layers of regulation that include nucleosome location, transcription factor binding, cofactor binding, RNA polymerase binding, and microRNA silencing. Several competing and compensatory mechanisms orchestrate these layers of regulation. Maintaining the appropriate level of regulation remains a vital process for all cellular life. The models above describe six

strategies cells may use to regulate HNF4A. Each model was described independently of the others; but these models may blend together to regulate HNF4A activity. For instance, HNF4A may bind PGC-1A only in the presence of microbiota-generated CLA. Furthermore, although microbiota colonization suppresses HNF4A activity, these models similarly provide competing pathways that function to maintain or increase HNF4A activity.

The paradigm of science is that all answers leads to new questions. My work has answered a few important questions about the nature of host-microbiota interactions, however it has led to several new questions that the field is poised to address. I foresee future progress requiring more biochemical approaches to identify the molecular processes that control intestinal epithelial transcription. Transcriptional assays, like those I have discussed extensively, do not necessarily translate to function and therefore more functional/biochemical assays must be performed to gain a deeper understanding of this intimate relationship.

4.3 Why suppress HNF4A activity?

I have now established that microbiota colonization is associated with a suppression of HNF4A activity, and I have proposed 6 possible methods that may coordinate this suppression. However, what are the downstream effects of microbial suppression of HNF4A? What are the possible consequences and the advantages to the microbiota for the suppression of HNF4A activity? What are the possible consequences and the advantages to the intestinal epithelia and host for this interaction? In the following section, I discuss the possible roles HNF4A may play in regulating metabolic, proinflammatory, and innate immune responses and how these processes are advantageous to the intestinal epithelia and the host. I then explore the concept of commensalism. Does a true mutualistic relationship exists between the microbiota and the host? Or, are host-microbiota interactions a constant tug-of-war for the available resources within the intestinal lumen?

4.3.1 Suppressing HNF4A may provide a significant advantage to the host

Both proinflammatory and metabolic genes are directly and indirectly regulated by HNF4A [248, 249, 308, 309]. In the mutant *hnf4a* zebrafish model, Hnf4a activity appears to repress expression of *tnfa*, *duox2*, and *il-1b* (proinflammatory and innate immune genes) and activate expression of *fabp2*, *elovl2,* and *apoa1a* (metabolism or mobilization of fatty acid genes). My data indicate the microbiota regulate these same genes in gnotobiotic zebrafish. Similarly, in mice, HNF4A binding sites are located at loci of innate immune or redox genes and fatty acid metabolism and mobilization genes. The exact role of HNF4A activity in regulating some of these genes, particularly the proinflammatory genes, remains unknown. However, the coincidences between the expression patterns in mutant zebrafish and the occurrence of HNF4A bindings sites at the same gene in mice suggest a conserved role of HN4A in regulating proinflammatory and metabolic transcription networks.

An obvious disadvantage to my data is that the mouse ChIP-seq, RNA-seq and DNAseq were performed from all villi-epithelial cells types. Therefore, most of my data is derived from enterocytes since this is the most prevalent cell type on villi; however, goblet cells and EECs may have some impact on my datasets since they make up a small percentage of cells on villi. My -omics datasets likely exclude Paneth cells, progenitor cells and stem cells, however they may make up a very small percentage of cells in the extracts. Therefore, the induction of *Reg3b* and *Reg3g*, classically thought be markers of Paneth cell which reside in the intestinal crypts, following microbiota colonization is likely be derived from villus cells.

Based on my datasets, HNF4A binds to the loci of *Muc13* and *Muc4* genes. Upon microbiota colonization, HNF4A occupancy at these loci is reduced and these genes become highly expressed. These data suggest that HNF4A may be repressing their transcription (or HNF4A binding near these loci is coincidence). Upon microbial stimuli, it is advantageous for the epithelium to upregulate its mucus production to maintain a barrier

between itself and the microbiota. A similar logic may exist for the gene *Reg3b* which functions as an antimicrobial peptide. Upon microbial colonization, HNF4A occupancy at the *Reg3b* loci is lost and its expression increases. The protein product of the *Duox2* gene generates reactive oxygen species that can kill bacteria. HNF4A also binds at the *Duox2* locus in the intestinal epithelium and this binding is reduced following colonization. Like the mucins and *Reg3b*, microbiota colonization induces *Duox2* expression, again suggesting HNF4A suppresses these epithelial defense mechanisms. Therefore, if the binding of HNF4A near these downregulated genes is not just coincidence, reduced HNF4A-repressive activity following microbiota colonization offers significant advantages to the intestinal epithelium because it provides a mechanism for the host to adapt to life with a microbiota by initiating anti-microbial defenses.

Activation of defense mechanisms upon microbiota colonization; is a logical response; however, what are the possible advantages to downregulating metabolic genes, particularly those involved in lipolysis, lipid mobilization, gluconeogenesis and beta-oxidation? As mentioned several times in my dissertation, microbiota colonization results in a rapid and prolonged shift in metabolic programs that results in downregulation of ketogenetic, beta-oxidative and gluconeogenic programs and an increase in glycolysis, amino acid metabolism, nucleic acid metabolism, and lipogenesis [364, 365]. This reorientation in metabolic programs may be a result of increased energy availability since the microbiota facilitate nutrient absorption and storage by the epithelium [156, 167, 168, 366]. I speculate the response to microbiota colonization is the host attempting to store all available energy. The epithelium shuts down its lipolysis and beta-oxidation programs so those fatty acids can be stored and later consumed during times of fasting. Meanwhile the host uses other sources of energy that may have just become available. Since time between meals could be an indefinite wait over the course of animal evolution, storage of energy would have been critical for survival. Since HNF4A activates the transcription of several

genes involved in the beta-oxidation pathway, suppression of HNF4A in the intestinal epithelium upon microbiota colonization may provide the host an advantage by enabling storage of energy in preparation for fasting. Furthermore, HNF4A appears to play a central role in the epithelium's adaptation to life with a microbiota, since loss of HNF4A activity results in severe metabolic derangements in response to microbiota colonization.

4.3.2 Mutualism or antagonism?

The microbiota field often refers to the relationship between the microbiota and the host as commensal. These paradigms suggest that the microbiota provide a benefit to the host while the host similarly provides a benefit to the microbial communities. For example, the host provides the microbiota a niche within the intestinal epithelium. Within this niche, the microbiota collect the resources and energy they need to survive and in return facilitate energy harvest for the host. However, even if majority of these interactions appear symbiotic, they may stem from a cellular struggle for control of the environment. An analogy of this relationship may be the relationship between two parties in a bipartisan political system. Both parties require each other to maintain political homeostasis and suppress new forms of government. That is, both parties will protect the bipartisan system to the mutual benefit of both parties. However, both parties similarly adopt strategies to undermined and attack the opposing party to gain independent power within the system. They fight over the same resources (voters) to gain and maintain political power. The microbiota and the intestinal epithelium appear to function similarly. Both the host and microbiota support homeostasis by killing invading pathogens; however, both the host and the microbiota fight over the same resources (dietary nutrients) and both have adopted strategies to impede the other's progress (Figure 4.2B).

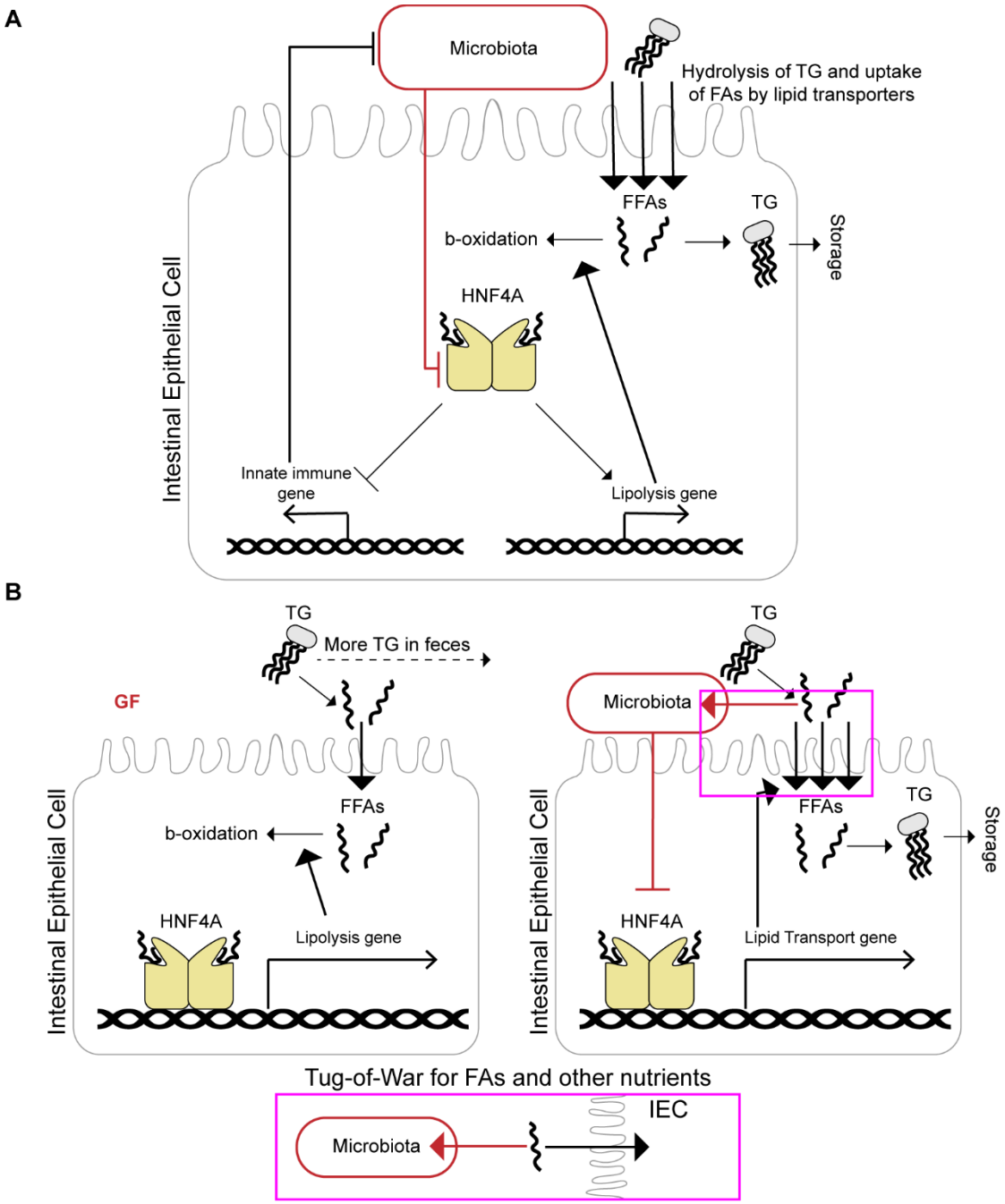


Figure 4.3: The advantages to the host and the microbiota following suppression of HNF4A activity. (A) Suppression of HN4A activity associated with microbiota colonization provides two potential advantages to the host. 1) Suppression of HNF4A relieves a possible repressive activity by HNF4A on proinflammator9y, innate immune, and redox genes. Upregulation of these genes protects the intestinal epithelia from the microbiota. 2) Suppression of HFN4A reduces the transcription of genes involved in beta-oxidation, lipolysis, and lipid transport, enabling FAs to be stored for consumption during fasting. (B) Microbiota colonization increases the energy availability to the host. Indeed, fecal samples from GF mice fed a high-fat diet have increased TG levels compared to fecal samples from CV mice [167]. The intestinal epithelial may begin rapid transport of FAs because of a new competition for the energy sources, i.e. a “Tug-of-War” for FAs and other nutrients.

Both the microbiota and the host utilize fatty acids as energy. Microbiota colonization results in repressed expression of fatty acid transporter genes in the intestinal epithelium; strategy used by both the epithelia and the microbiota to control energy availability. This suppression provides each an advantage in this tug-of-war for resources. Upon colonization, the host begins to utilize other energy sources and preferentially stores dietary fatty acids and retains them from the colonizing microbiota. Some microbial taxa similarly consume dietary fatty acids for metabolic processes [367]. Microbiota-suppression of HNF4A reduces transcription of fatty acid transport genes. This may result in a retention of fatty acids in the intestinal lumen where the microbiota can maintain access to these dietary nutrients. Similarly, the downregulation of fatty acid absorption genes may be a secondary effect following an initial burst in fatty acid absorption and storage following colonization. Therefore, suppression of HNF4A activity provides advantages to both the microbiota and the host, which may sound mutualistic, but the mechanisms that drive this process may be antagonistic.

4.4 The overlap of HNF4A activity and the microbiota in human diseases:

Both the microbiota and HNF4A are implicated in a variety of intestinal and metabolic diseases. Loss of HNF4A activity and HNF4A variants are associated with Crohn's disease, ulcerative colitis, maturity onset diabetes of the young (MODY1), and metabolic syndrome [233, 256, 303, 368]. Similarly, the microbiota are associated with obesity, insulin resistance, type 2 diabetes, Crohn's disease and ulcerative colitis [221, 364, 369]. Given that microbiota colonization results in reduced HNF4A activity, what role does this suppression play in these diseases? Is HNF4A a viable therapeutic target for treating these human diseases? As discussed in the previous section, how does this antagonistic relationship, which typically maintains homeostasis, become detrimental to the host? What role does HNF4A play in maintaining the balance in this relationship and what are the consequences of improper

HNF4A regulation? In this next section, I will discuss how both HNF4A and the microbiota contribute to metabolic syndrome and inflammatory bowel diseases. I will discuss models how suppression of HNF4A may mediate the pathology of these diseases.

4.4.1 Metabolic syndrome

Metabolic Syndrome is defined as a cluster of risk factors that indicate increased risk of diabetes, cardiovascular disease and premature mortality [370]. These risk factors include but are not limited to insulin resistance, increased ratio of visceral to subcutaneous adiposity, dyslipidemia and arterial hypertension [370-372]. The composition of the gut microbiota is linked to several of these risk factors including obesity and insulin resistance [373, 374]. Based on the microbial alterations of host metabolic programs, it is not surprising that the microbiota contribute to these risk factors. Indeed, microbiota colonization in mice is associated with increased serum glucose levels, increased adipose tissue size, and decreased activity [156, 169]. Furthermore, the transcriptional changes associated with microbiota colonization indicate a shift from a “starved” state to a “fed” state. This transcriptional shift is logical because of enhanced nutrient availability including fatty acids [168]. However, impaired adipose tissue handling of dietary fats may result in increased circulating triglycerides [375]. If the body fails to compensate for these circulating triglycerides by increasing fatty acid oxidation within cells, these triglycerides may be stored in non-adipose tissue, such as skeletal muscle, or continue to circulate. Both the increase of fat stores in non-adipose tissue and increased circulating triglycerides can contribute to the development of insulin resistance [376]. Therefore, maintaining a sensitivity to circulating triglycerides is essential for proper clearance of circulating triglycerides and maintaining insulin sensitivity.

HNF4A serves as a master regulator of metabolism. Aberrant suppression of its activity impairs cellular fatty acid-oxidation [308] and increases risk of diabetes [309].

Furthermore, gene variants have been associated with both type 2 diabetes and metabolic syndrome [303, 304]. Genetic factors, diet, activity level, and microbiota collectively contribute to metabolic syndrome. These factors control the function of several tissues and cellular processes that become impaired during the progression of insulin resistance and obesity. Therefore, I speculate that microbial suppression of HNF4A may only play a small contribution to insulin resistance but is not the only element mediating the risk factor.

The *Hnf4a* intestine-specific knockout mouse has impaired fatty acid uptake and does not develop insulin resistance. These data suggest aberrant suppression of HNF4A activity in the gut may not be sufficient to drive insulin resistance. Instead, these results address an important aspect of HNF4A activity and microbiota impact on epithelial function: fatty acid uptake. Aside from beta-oxidation genes, the microbiota also suppress several fatty acid transport and mobilization genes, including *Slc27a2, ApoA1,* and *Fabp2*. Based on my zebrafish RNA-seq and my mouse RNA-seq and ChIP-seq datasets, these genes also appear to be regulated by HNF4A. So, aberrant suppression of HNF4A activity in the intestine should, in part, be protective of obesity and insulin resistance due to an impairment of fatty acid uptake. So microbial suppression of HNF4A activity in the intestine may not contribute to metabolic syndrome. However, loss of HNF4A activity in the liver may contribute to insulin resistance. Indeed, the liver-specific HNF4A knockout mouse has increased fat deposition in the liver likely due to lipid transport failure [377]. Liver fat is highly correlative to metabolic syndrome and insulin resistance [378, 379]. Therefore, I speculate that microbial suppression of intestinal HNF4A does not contribute to metabolic syndrome. Instead, since the microbiota are known to mediate transcription and metabolic responses in the liver [169, 380], perhaps microbial regulation of liver HNF4A contributes to metabolic syndrome. This regulation may occur through any of the mechanisms detailed in section 4.2.

4.4.2 Inflammatory Bowel Disease

Both the microbiota and HNF4A have been independently implicated in inflammatory bowel diseases (IBD). Indeed, HNF4A variants are associated with both CD and UC [232, 233]. Similarly, HNF4A expression has been shown to be decreased in UC as well as in mouse models of colitis [256, 260]. Finally, transient knockdown of Hnf4a in liver cell culture induces an inflammatory response that continues to repress HNF4A expression [250]. Microbiota composition is also associated with onset of inflammation in CD and UC [221, 381]. Antibiotics have been shown to provide relief to some CD and UC patients, indicating episodes of inflammation are associated to microbiota composition [382]. And many/most mouse models of IBD are asymptomatic when raised GF, underscoring the central importance of microbial stimulation in IBD pathogenesis. The multi-species meta-analysis performed in chapter 3 (Figure 3.5) is the first to draw a link between the microbiota and HNF4A in the context of IBD. My data indicate that HNF4A may protect from intestinal inflammation and disease pathology. Deletion of *hnf4a* activity in the presence of a microbiota in gnotobiotic zebrafish results in a transcription program that resembles the transcriptome of CD and UC patients. This IBD transcription program generates high expression of inflammatory genes and low expression of metabolic genes. Together, these studies indicate HNF4A may play a significant role in UC and CD pathologies. However, several questions remain to be answered: Does microbial suppression of HNF4A contribute to IBD progression? Or, does loss of HNF4A activity through other mechanisms contribute to IBD pathologies? How does HNFA protect against IBD? Is HNF4A a viable target for therapies in IBD research?

We do not know if microbial suppression of HNF4A contributes to disease pathology. The data collected only indicate the microbiota have the capacity to suppress HNF4A activity, but we do not know if this suppression causes IBDs. Perhaps the microbiota only induce IBD after a microbiota-independent loss of HNF4A activity. Perhaps, genetic variants

of HNF4A are more susceptible to suppression by the microbiota and thus predispose patients to IBDs. We also do not know if the mechanisms that control HNF4A activity upon microbiota colonization are the same mechanisms that suppress it in IBD.

How does the suppression of HNF4A, a master regulator of metabolic gene expression suppress inflammation and epithelial function? And similarly, why does HNF4A appear to activate metabolic genes and repress proinflammatory and innate immune genes (as discussed in the previous section)? A common theme in inflammation studies is that the activation of innate immune genes coincides with a suppression of metabolic gene programs [132]. However, we do not have a strong understanding as to why the activation of these programs are mutually exclusive.

Poly-unsaturated fatty acids have also been shown to reduce inflammation in patients; however, the mechanisms for this relief are largely unknown. Some studies have shown that poly unsaturated fatty acid reduces NF-κB activity thereby suppressing the immune response [319, 383, 384]. The putative endogen ligand of HNF4A, linoleic acid, protects zebrafish from chemical-induced inflammation [385]. Perhaps this fatty acid, and other poly unsaturated fatty acids promote HNF4A activity and suppresses the inflammation.

HNF4A activity may protect the intestine from chronic inflammation through several mechanisms: 1) HNF4A promotes general IEC homeostasis, including barrier function. 2) HNF4A activates transcription of specific anti-inflammatory genes like *APOA1* and *miR-124*, a gene that is commonly suppressed in both UC and CD [2, 237]. Injection of APOA1 mimetic peptide rescues experimental colitis [262]. 2) HNF4a activates the transcription of miR-124 in liver cells [250]. STAT3 expression is silenced by miR-124 [261]. Since *miR-124* has reduced expression in biopsies from pediatric CD patients compared to healthy patients [261], perhaps HNF4A regulates miR-124 in the intestine as well. 3) My ChIP-seq and RNA-seq data suggest HNF4A may repress several proinflammatory and innate immune genes, including *DUOX2*, which is commonly upregulated in UC and CD [237]. Increased *DUOX2*

expression is only correlative with the onset of disease and inflammation and has yet to be shown as a causative factor in IBD; however, its activity does increase the potential for DNA damage and epithelial cell death [386]. 4) Four HNF4A binding sites are found at *Il10rb* gene locus, an important anti-inflammatory cytokine receptor. Loss of IL10 signaling can similarly result in spontaneous colitis [387], so perhaps HNF4A activates transcription of this receptor. Loss of HNF4A respective activities at these genes may lead to aberrant inflammation.

4.5 Concluding Remarks: HNF4 and the expansion of the nuclear receptor superfamily

The genome of the demosponge, the most ancient member of the metazoan, encodes a single nuclear receptor that resembles HNF4A [148]. Because HNF4A resembles the most ancient of the nuclear receptors, the discovery that HNF4A mediates microbial suppression of host genes is exciting because it suggests a possible model that an HNF4-like nuclear receptor has been interpreting and protecting the host from microorganisms since the dawn of the metazoa. It also permits a possible theory for the evolution of other metabolically regulated nuclear receptors: as animal evolution progressed and as animal tissues and cell-types became more diverse and their microbial communities that interface with these host cells became more diverse, the need for molecules and signaling pathways that can interpret and respond to these new microbiota-derived cues arose. Since nuclear receptors provide a simple mechanism that can fulfill this need, it makes sense that these evolutionary pressures resulted in radial expansion of the nuclear receptor superfamily.

Over the course of metazoan evolution, molecular tinkering [388] drove the expansion of the nuclear receptors superfamily to over 50 members family found in humans [147]. Even If HNF4A resembles the most ancient of the nuclear receptors, it certainly is not nature’s most malleable nuclear receptor. That role belongs to the RXR family of nuclear

receptors. The RXR family of nuclear receptors arose shortly after HNF4-like nuclear receptors evolved [148]. Evolution has tinkered with RXR activity repeatedly as new metabolite-binding nuclear receptors evolved. Indeed, almost every non-steroid binding nuclear receptor, with the notable exception of the HNF4A, heterodimerizes with RXR [147]. Examples of these nuclear receptors that heterodimerize with RXR are: PPARs, FXRs, RARs and VDRs, all of which are activated by metabolites that are either generated or modified by the microbiota [147, 152, 155, 389, 390]. So even if HNF4A represents the first of the putative-microbiota-sensing nuclear receptors, the evolution of RXRs, along with the evolution of their dimerization partners, has vastly increased the sensitivities of host cells, enabling new responses to the environment which may include responses to the microbiota.

An obvious question regarding nuclear receptor evolution is: what drove the expansion of the family? Did the complexities of host-microbiota relationships and the environment drive the expansion of the nuclear receptors? Or have the microbiota and the environment only tapped into developmental programs that were established before exogenous interventions? Several nuclear receptors are required for mouse embryonic development, including HNF4A [391, 392], RXR [393], and PPARG [394, 395]. These developmental failures suggest, at the very least, that these transcription factors became essential for development after they evolved. Unfortunately, we do not know if demosponges require their HNF4-like nuclear receptors for development, which may provide insight into what drove HNF4 evolution. However, the *hnf4a* mutant zebrafish is viable. *Hnf4a* knockout mouse embryos do not initiate the gastrula stage and stall during visceral endoderm differentiation [391], indicating *Hnf4a* plays a critical role in the development of this essential extra-embryonic tissue. We believe the *hnf4a* mutant zebrafish is viable because zebrafish and other fishes do not have a visceral endoderm and therefore *hnf4a* is less critical during embryogenesis. The visceral endoderm, which develops from the hypoblast, appears to be a mammalian-specific extra-embryonic tissue, suggesting that

Hnf4a may have only become essential for embryonic development since the evolution of mammals. These studies permit the theory that HNF4A, and perhaps other nuclear receptors required for development of mammalian-specific tissues (PPARG [394]), may have originally evolved to establish cellular sensitivities to the environment and the microorganisms within it.

My work represents one of the first studies to demonstrate how the microbiota regulate transcription factor binding genome wide. My data place a focus on HNF4A a as key determinant in the host response to the microbiota. However, recent work has indicated that other nuclear receptors play important roles in the host response along the entire length of the intestine as well as along the villus-crypt axis [152, 360, 389]. Together with my work, these studies suggest that nuclear receptors maintain homeostasis between the intestinal epithelia and the microbiota. Because nuclear receptors bind directly to exogenous molecules that impact their activities, this superfamily of transcription factors may provide sensitivity to the environment that tunes intestinal epithelial function and identity.

APPENDIX 1: A NOTE ON MODEL SYSTEMS

The epithelia must respond to fluctuations of two primary components within the luminal environment: microbiota and dietary nutrients. However, distinguishing the individual effects of these components on epithelial response and gene regulation is confounded by dietary impact on microbial community compositions and the microbial modifications of dietary molecules (Figure 2.2. Indeed, some bacterial taxa including *Rosburia intestinalis* encode genes that generate non-native isomers of poly-unsaturated fatty acids which have been shown to affect host physiology [129] and the large cohort of anaerobic bacteria in the colon have the capacity to catabolize dietary fibers that generate short-chain fatty acids [130]. Furthermore, diets with high lipid composition promote the growth of specific bacterial taxa, which can induce proinflammatory transcriptional responses from the intestinal epithelial cells, and diets with high-protein and low-carbohydrate intake promote growth of other bacterial taxa [124, 396, 397]. These studies indicate there is an intimate relationship between microbiota composition and the availability of specific dietary molecules and vice versa. Therefore, because the addition of one component to the luminal environment can have a rippling effect that impacts whole microbial communities and subsequent dietary molecules, establishing controlled and high throughput *in vivo* model systems remains paramount to understanding clear mechanisms that mediate the epithelial response to the microbiota and diet. To study these complex relationships, the microbiota field has taken advantage of several model systems, each with their own advantageous tools that permit the dissection of these multidimensional interactions.

The zebrafish and *Drosophila* model systems have provided important understandings to intestinal epithelial development and transcriptional regulatory programs. Non-mammalian host systems are equipped with the powerful genetic and amenable tools and screening platforms for determining the molecular mechanisms that mediate

transcriptional responses. A forward genetic screen in zebrafish demonstrated that the loss of function of a DNA-methylation protein (*uhrf1*) results in hypomethylation of the *tnfa* locus. This hypomethylation results in high expression of the proinflammatory marker and an intestinal epithelial phenotype that resembles IBD [285]. Knocking down *tnfa* expression by morpholino injection rescues this IBD-like phenotype. Although Human GWAS studies have suggested similar mechanisms, this forward genetic screen provided mechanistic understanding to how loss of epigenetic repression of a proinflammatory marker can lead to IBD onset. Furthermore, compared to the mammalian model systems, the relatively simple microbiome of the *Drosophila* makes the fruitfly an ideal model organism to study host-microbiota interactions [398]. Taking advantage of these smaller microbial communities, researchers could identify that microbiota regulate insulin signaling and are required for pupal survival [399]. Further studies indicated that this interaction was diet dependent and that supplementing a glucose only diet with a vitamin B source rescued the requirement for a microbiota [400]. These drosophila studies demonstrate the power of the fast, low-cost and high throughput system; this system enables researchers to perform gnotobiotic experiments with multiple diet manipulations to determine nutritional requirements for development and how the microbiota can and cannot satisfy those requirements.

The ability for high-throughput transgene screening of an *in vivo* system is a unique attribute of non-mammalian models. The high throughput transgenic tools available in *Drosophila* facilitated a study addressing how each epithelial cell type along the intestinal tract rewires its transcriptional program in response to infection [401]. Furthermore, transgenic tools available in these non-mammalian systems permits functional testing of non-coding genomic regulatory elements. Indeed, zebrafish have provided the first example of a microbiota controlled *cis*-regulatory region [145], indicating that the microbiota may also regulate the transcription factors that bind within the region. Studies such as these provide

the necessary context to determine which transcription factors are regulated by the microbiota.

Most of what we know from host-microbiota interactions comes from data collected from mammalian model organisms. An obvious reason why mammalian systems have an advantage over non-mammalian systems is their homology to human physiology and health. However, they also provide other unique opportunities that are simply not available or are limited in zebrafish or fruitfly. For instance, functional genomic studies, particularly Chromatin Immunoprecipitation, are more easily performed in mammalian systems because these techniques require a lot of starting material and good antibodies. Furthermore, the large tissue sizes enable pairwise comparisons for different functional genomic datasets from the same animal, strengthening statistical power from these datasets. Although becoming more common in zebrafish systems [402], tissue specific knockout mutations in murine models represents a commonly used technique in mammalian studies and have been used to show how the transcription factor CDX2 or HNF4A both maintain homeostasis in the intestinal epithelia [29, 89] as well as showing how HDAC3 is necessary for the host response to microbiota colonization [236]. The mammalian system also provides the primary platform longterm gnotobiotic studies. Gnotobiot zebrafish studies end during early larval stages, providing insight into the how the microbiota control developmental programs, such as the edification of the immune system [146, 194]. However, murine gnotobiotic studies maintain GF and conventionalized status for several weeks, providing important information regarding the length of time it takes to reach mucosal and transcriptional homeostasis in the intestinal epithelia following microbiota colonization. Only recently, have there been advances toward longterm zebrafish gnotobiotic zebrafish husbandry [403].

Enteroid culture has also provided unique opportunities to study intestinal epithelium biology. Enteroids are small cultured epithelial colonies that are commonly generated from a single small intestinal epithelial crypt. The stem cells in these crypts asymmetrically divide

as they would in the intestine and generate a small luminal organ differentiated daughter cells [404]. These cultured organoids have provided key understandings in:1) how the enteric nervous system participates in microbiota stimulated inflammatory responses [405]; 2) cytokine maintenance of the stem cell niche [406]; 3) signaling pathways that mediate cellular differentiation [407]. However, the obvious limitations to cultured organoids are the lack of a true basolateral membrane as well as a lack of extra-intestinal systems, like an enteric nervous system or a liver which are both known to communicate with the intestine and regulate digestive processes. A current cell culture technology that would be fascinating to apply to any model system would line scan photo-activatable Fluorescence Correlation Spectroscopy, in which the activity of individual transcription factors is monitored within a single nucleus [408]. Establishing this microscopy system in enteroid culture or in zebrafish model would provide novel insight into how transcription factors are responding within a given epithelial cell that is exposed to different luminal environments.

Higher throughput transgenic technologies in mammalian model systems have recently become to rival those in non-mammalian systems. Although incapable of the same offspring sizes as zebrafish and fruitfly, and therefore limited by its throughput, shotgun delivery of transgenic reporters for functional genomic studies [409] as well as CRISPR targeting for genetic mutation [410] and transcriptional trans-activation/repression [411] have become powerful techniques in the mammalian and cell culture toolsheds. Although these techniques have not been applied to understanding intestinal biology or host-microbiota interactions, they hold great potential to unlock new knowledge in intestinal homeostasis. For instance, previous studies have identified differentially regulated enhancers in IBD patients and some of these enhancers also harbor known SNPs in IBD. Using a viral transgenic shotgun approach, the activity of these enhancers can be tested in the intestinal epithelia in a gnotobiotic setting, providing insight into the function of these mutations and if the microbiota, known contributors to IBD, mediate the activity of these mutant enhancers.

APPENDIX 2: MICROBIOTA COLONIZATION METHODS AND "THE WINDOW OF OPPURTUNITY"

Various methods of colonization or depletion may elicit different host responses Microbiota colonization of germ free mice elicits a strong transcriptional response in the intestinal epithelia that is most robust in the first 1 – 2 weeks following colonization and persists for several weeks before reaching equilibrium [120, 121]. This sustained transcriptional response may be driven in part by the fact that the mucus layer within the small and large intestine does not achieve conventionally-reared (animals that were colonized by the microbiota from birth) viscosity and thickness until 8 weeks post colonization [412]. Surprisingly, the mucus layer of GF animals is highly viscous and difficult to remove from the epithelial layer because the mucin is still attached to the goblet cells [69]. Following microbial colonization, bacteria begin to modify the mucus layer, which begins to expand and develop a gradient of viscosity, where the most difficult to penetrate mucus resides closest to the epithelia. Following colonization, bacterial communities residing in the intestine show dramatic fluctuations and the most abundant phyla switches between Firmicutes and Bacteroidetes [412]. Perhaps, induction of antimicrobial peptides, derived from the host epithelium, sustains this long-term battle for most abundant bacteria.

These mucosal and transcriptome data provide the basis for a debate about when to assay the host response. Although not necessarily physiologically relevant in healthy animals, assaying during the most robust response (2 weeks post colonization) provides important information that indicate which cellular processes the microbiota do mediate, particularly in non-homeostatic settings like in human disease. One final frequently used method to investigate microbiota control of host physiology is to evaluate the impact of treating conventionally-reared animals with antibiotics. However, it has been established that the majority of host transcriptional changes induced by antibiotic treatment can be

explained by direct effects of the antibiotic on host cells or by the effects of remaining antibiotic-resistant microbes [287].

Both the method of colonization and the age of the host impact the epithelial responses. Some colonization methods fail to preserve the anaerobic bacteria, which are known to generate short chain fatty acids in the colon which can prevent human diseases [130]. Other colonization methods fail to control for differences between inoculums, indicating the host responses may differ between experiments due to differences in microbial communities. Studies involving cesarean-born mouse neonates vs vaginally-born neonates show that microbiota colonization from vaginal births mediates down-regulation of TLR signaling and downstream transcriptional programs in IECs. This suppression reduces sensitivities to gram negative bacteria during the first few weeks of life and protects against epithelial damage and epithelial tolerance later in life [413, 414].

Regardless of the method of colonization or the age of the animal at the time of colonization, studies have shown the microbiota have profound impacts on host gene expression. Therefore, we can continue to learn something regarding how the microbiota mediate these transcriptional changes. The debate of how to colonize and when to colonize will likely continue for as long as the microbiome field remains an intense area of study.

REFERENCES

1. Dalerba, P., et al., *CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer.* New England Journal of Medicine, 2016. **374**(3): p. 211-222.

2. Weiser, M., et al., *Molecular classification of Crohn's disease reveals two clinically relevant subtypes.* Gut, 2016.

3. Perrigoue, J., A. Das, and J.R. Mora, *Interplay of nutrients and microbial metabolites in intestinal immune homeostasis: distinct and common mechanisms of immune regulation in the small bowel and colon.* Nestle Nutr Inst Workshop Ser, 2014. **79**: p. 57-71.

4. Perrone, E.E., et al., *Dietary bile acid supplementation improves intestinal integrity and survival in a murine model.* Journal of Pediatric Surgery. **45**(6): p. 1256-1265.

5. Nigro, G. and P.J. Sansonetti, *Microbiota and Gut Stem Cells Cross-Talks: A New View of Epithelial Homeostasis.* Current Stem Cell Reports, 2015. **1**(1): p. 48-52.

6. Wang, Z., et al., *Morphological and molecular evidence for functional organization along the rostrocaudal axis of the adult zebrafish intestine.* BMC Genomics, 2010. **11**(1): p. 392.

7. Ng, A.N., et al., *Formation of the digestive system in zebrafish: III. Intestinal epithelium morphogenesis.* Dev Biol, 2005. **286**(1): p. 114-35.

8. Rodríguez-Fraticelli, A.E., et al., *Developmental regulation of apical endocytosis controls epithelial patterning in vertebrate tubular organs.* Nat Cell Biol, 2015. **17**(3): p. 241-250.

9. Bates, J.M., et al., *Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation.* Dev. Biol., 2006. **297**(2): p. 374-86.

10. Gilmore, A.P., *Anoikis.* Cell Death Differ, 2005. **12 Suppl 2**: p. 1473-7.

11. Aghaallaei, N., et al., *Identification, visualization and clonal analysis of intestinal stem cells in fish.* Development, 2016. **143**(19): p. 3470-3480.

12. Wallace, K.N., et al., *Intestinal growth and differentiation in zebrafish.* Mechanisms of Development, 2005. **122**(2): p. 157-173.

13. Small, D.M., R.H. Dowling, and R.N. Redinger, *The enterohepatic circulation of bile salts.* Archives of Internal Medicine, 1972. **130**(4): p. 552-573.

14. Sandle, G.I., *Salt and water absorption in the human colon: a modern appraisal.* Gut, 1998. **43**(2): p. 294-9.

15. Coskun, A.F., U. Eser, and S. Islam, *Cellular identity at the single-cell level.* Mol Biosyst, 2016. **12**(10): p. 2965-79.

16. Dasika, G.K., et al., *DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis.* Oncogene, 1999. **18**(55): p. 7883-99.

17. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.

18. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.

19. Bulger, M. and M. Groudine, *Functional and mechanistic diversity of distal transcription enhancers.* Cell, 2011. **144**(3): p. 327-39.

20. Boyle, A.P., et al., *High-resolution mapping and characterization of open chromatin across the genome.* Cell, 2008. **132**(2): p. 311-22.

21. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

22. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-5.

23. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.

24. Lee, Tong I. and Richard A. Young, *Transcriptional Regulation and Its Misregulation in Disease.* Cell, 2013. **152**(6): p. 1237-1251.

25. Iwafuchi-Doi, M. and K.S. Zaret, *Pioneer transcription factors in cell reprogramming.* Genes Dev, 2014. **28**(24): p. 2679-92.

26. Erceg, J., et al., *Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements.* Genes Dev, 2017.

27. Gao, N., P. White, and K.H. Kaestner, *Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2.* Dev Cell, 2009. **16**(4): p. 588-99.

28. Verzi, M.P., et al., *Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2.* Dev Cell, 2010. **19**(5): p. 713-26.

29. Verzi, M.P., et al., *Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding.* Mol Cell Biol, 2013. **33**(2): p. 281-92.

30. San Roman, A.K., et al., *Distinct Processes and Transcriptional Targets Underlie CDX2 Requirements in Intestinal Stem Cells and Differentiated Villus Cells.* Stem Cell Reports, 2015. **5**(5): p. 673-81.

31. Kim, T.H., et al., *Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity.* Nature, 2014. **506**(7489): p. 511-5.

32. Groisman, G.M., M. Amar, and A. Meir, *Expression of the intestinal marker Cdx2 in the columnar-lined esophagus with and without intestinal (Barrett's) metaplasia.* Mod Pathol, 2004. **17**(10): p. 1282-8.

33. Tamagawa, Y., et al., *Notch signaling pathway and Cdx2 expression in the development of Barrett's esophagus.* Lab Invest, 2012. **92**(6): p. 896-909.

34. Sakamoto, N., et al., *BRAFV600E cooperates with CDX2 inactivation to promote serrated colorectal tumorigenesis.* Elife, 2017. **6**.

35. Kazumori, H., et al., *Bile acids directly augment caudal related homeobox gene Cdx2 expression in oesophageal keratinocytes in Barrett's epithelium.* Gut, 2006. **55**(1): p. 16-25.

36. Domon-Dell, C., et al., *Stimulation of the intestinal Cdx2 homeobox gene by butyrate in colon cancer cells.* Gut, 2002. **50**(4): p. 525-529.

37. Tan, D.W. and N. Barker, *Intestinal stem cells and their defining niche.* Curr Top Dev Biol, 2014. **107**: p. 77-107.

38. Peck, B.C., et al., *Functional Transcriptomics in Diverse Intestinal Epithelial Cell Types Reveals Robust MicroRNA Sensitivity in Intestinal Stem Cells to Microbial Status.* J Biol Chem, 2017. **292**(7): p. 2586-2600.

39. Thomas, H., *Gut microbiota: Host faecal miRNA regulates gut microbiota.* Nat Rev Gastroenterol Hepatol, 2016. **13**(3): p. 122-3.

40. Lee, J., E.J. Park, and H. Kiyono, *MicroRNA-orchestrated pathophysiologic control in gut homeostasis and inflammation.* BMB Rep, 2016. **49**(5): p. 263-9.

41. Liu, S., et al., *The Host Shapes the Gut Microbiota via Fecal MicroRNA.* Cell Host Microbe, 2016. **19**(1): p. 32-43.

42. Gerbe, F., C. Legraverend, and P. Jay, *The intestinal epithelium tuft cells: specification and function.* Cell Mol Life Sci, 2012. **69**(17): p. 2907-17.

43. McHale, P.T. and A.D. Lander, *The protective role of symmetric stem cell division on the accumulation of heritable damage.* PLoS Comput Biol, 2014. **10**(8): p. e1003802.

44. Snippert, H.J., et al., *Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells.* Cell, 2010. **143**(1): p. 134-44.

45. Shirazi, T., et al., *Mucins and inflammatory bowel disease.* Postgrad Med J, 2000. **76**(898): p. 473-8.

46. Sato, T., et al., *Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts.* Nature, 2011. **469**(7330): p. 415-8.

47. Zhang, Z. and Z. Liu, *Paneth cells: the hub for sensing and regulating intestinal flora.* Sci China Life Sci, 2016. **59**(5): p. 463-7.

48. Birchenough, G.M., et al., *A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion.* Science, 2016. **352**(6293): p. 1535-42.

49. Medema, J.P. and L. Vermeulen, *Microenvironmental regulation of stem cells in intestinal homeostasis and cancer.* Nature, 2011. **474**(7351): p. 318-26.

50. Umar, S., *Intestinal stem cells.* Curr Gastroenterol Rep, 2010. **12**(5): p. 340-8.

51. Korinek, V., et al., *Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4.* Nat Genet, 1998. **19**(4): p. 379-383.

52. Jadhav, U., et al., *Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells.* Cell, 2016. **165**(6): p. 1389-400.

53. Camp, J.G., et al., *Microbiota modulate transcription in the intestinal epithelium without remodeling the accessible chromatin landscape.* Genome Res, 2014. **24**(9): p. 1504-16.

54. Chen, T. and S.Y.R. Dent, *Chromatin modifiers and remodellers: regulators of cellular differentiation.* Nat Rev Genet, 2014. **15**(2): p. 93-106.

55. Yang, Q., et al., *Requirement of Math1 for secretory cell lineage commitment in the mouse intestine.* Science, 2001. **294**(5549): p. 2155-8.

56. Shroyer, N.F., et al., *Intestine-specific ablation of mouse atonal homolog 1 (Math1) reveals a role in cellular homeostasis.* Gastroenterology, 2007. **132**(7): p. 2478-88.

57. Windham, T.C., et al., *Src activation regulates anoikis in human colon tumor cell lines.* Oncogene, 2002. **21**(51): p. 7797-807.

58. Singh, A.B., et al., *Claudin-1 up-regulates the repressor ZEB-1 to inhibit E-cadherin expression in colon cancer cells.* Gastroenterology, 2011. **141**(6): p. 2140-53.

59. Lo, Y.H., et al., *Transcriptional Regulation by ATOH1 and its Target SPDEF in the Intestine.* Cell Mol Gastroenterol Hepatol, 2017. **3**(1): p. 51-71.

60. Gribble, F.M. and F. Reimann, *Enteroendocrine Cells: Chemosensors in the Intestinal Epithelium.* Annu Rev Physiol, 2016. **78**: p. 277-99.

61. Jenny, M., et al., *Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium.* EMBO J, 2002. **21**(23): p. 6338-47.

62. Gerbe, F., et al., *Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium.* J Cell Biol, 2011. **192**(5): p. 767-80.

63. Jepeal, L.I., et al., *Cell-Specific Expression of Glucose-Dependent-Insulinotropic Polypeptide Is Regulated by the Transcription Factor PDX-1.* Endocrinology, 2005. **146**(1): p. 383-391.

64. Cataland, S., et al., *Gastric inhibitory polypeptide (GIP) stimulation by oral glucose in man.* J Clin Endocrinol Metab, 1974. **39**(2): p. 223-8.

65. Cheung, A.T., et al., *Glucose-dependent insulin release from genetically engineered K cells.* Science, 2000. **290**(5498): p. 1959-62.

66. Servitja, J.M. and J. Ferrer, *Transcriptional networks controlling pancreatic development and beta cell function.* Diabetologia, 2004. **47**(4): p. 597-613.

67. Baraille, F., et al., *Glucose Tolerance Is Improved in Mice Invalidated for the Nuclear Receptor HNF-4gamma: A Critical Role for Enteroendocrine Cell Lineage.* Diabetes, 2015. **64**(8): p. 2744-56.

68. Ye, D.Z. and K.H. Kaestner, *Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice.* Gastroenterology, 2009. **137**(6): p. 2052-62.

69. Birchenough, G.M., et al., *New developments in goblet cell mucus secretion and function.* Mucosal Immunol, 2015. **8**(4): p. 712-9.

70. Ridley, C., et al., *Assembly of the respiratory mucin MUC5B: a new model for a gel-forming mucin.* J Biol Chem, 2014. **289**(23): p. 16409-20.

71. Johansson, M.E., J.M. Larsson, and G.C. Hansson, *The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4659-65.

72. Shroyer, N.F., et al., *Gfi1 functions downstream of Math1 to control intestinal secretory cell subtype allocation and differentiation.* Genes & Development, 2005. **19**(20): p. 2412-2417.

73. Gregorieff, A., et al., *The Ets-Domain Transcription Factor Spdef Promotes Maturation of Goblet and Paneth Cells in the Intestinal Epithelium.* Gastroenterology. **137**(4): p. 1333-1345.e3.

74. van der Sluis, M., et al., *Forkhead box transcription factors Foxa1 and Foxa2 are important regulators of Muc2 mucin expression in intestinal epithelial cells.* Biochem Biophys Res Commun, 2008. **369**(4): p. 1108-13.

75. Yamamoto, H., Y.-Q. Bai, and Y. Yuasa, *Homeodomain protein CDX2 regulates goblet-specific MUC2 gene expression.* Biochemical and Biophysical Research Communications, 2003. **300**(4): p. 813-818.

76. Tompkins, D.H., et al., *Sox2 activates cell proliferation and differentiation in the respiratory epithelium.* Am J Respir Cell Mol Biol, 2011. **45**(1): p. 101-10.

77. Goldsmith, S., R. Lovell-Badge, and K. Rizzoti, *SOX2 is sequentially required for progenitor proliferation and lineage specification in the developing pituitary.* Development, 2016. **143**(13): p. 2376-88.

78. Raghoebir, L., et al., *SOX2 redirects the developmental fate of the intestinal epithelium toward a premature gastric phenotype.* Journal of Molecular Cell Biology, 2012. **4**(6): p. 377-385.

79. Bastide, P., et al., *Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium.* J Cell Biol, 2007. **178**(4): p. 635-48.

80. Mori-Akiyama, Y., et al., *SOX9 is required for the differentiation of paneth cells in the intestinal epithelium.* Gastroenterology, 2007. **133**(2): p. 539-46.

81. Akiyama, H., et al., *Interactions between Sox9 and beta-catenin control chondrocyte differentiation.* Genes Dev, 2004. **18**(9): p. 1072-87.

82. Formeister, E.J., et al., *Distinct SOX9 levels differentially mark stem/progenitor populations and enteroendocrine cells of the small intestine epithelium.* American Journal of Physiology - Gastrointestinal and Liver Physiology, 2009. **296**(5): p. G1108-G1118.

83. Cheng, H., *Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. IV. Paneth cells.* Am J Anat, 1974. **141**(4): p. 521-35.

84. Gracz, A.D., S. Ramalingam, and S.T. Magness, <*em*>*Sox9*</*em*> *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, 2010. **298**(5): p. G590-G600.

85. Ayabe, T., et al., *Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria.* Nat Immunol, 2000. **1**(2): p. 113-8.

86. Ouellette, A.J., et al., *Mouse Paneth cell defensins: primary structures and antibacterial activities of numerous cryptdin isoforms.* Infect Immun, 1994. **62**(11): p. 5040-7.

87. Ouellette, A.J., et al., *Purification and primary structure of murine cryptdin-1, a Paneth cell defensin.* FEBS Lett, 1992. **304**(2-3): p. 146-8.

88. Farin, H.F., et al., *Visualization of a short-range Wnt gradient in the intestinal stem-cell niche.* Nature, 2016. **530**(7590): p. 340-3.

89. Babeu, J.P., et al., *Hepatocyte nuclear factor 4alpha contributes to an intestinal epithelial phenotype in vitro and plays a partial role in mouse intestinal epithelium differentiation.* Am J Physiol Gastrointest Liver Physiol, 2009. **297**(1): p. G124-34.

90. von Moltke, J., et al., *Tuft-cell-derived IL-25 regulates an intestinal ILC2-epithelial response circuit.* Nature, 2016. **529**(7585): p. 221-5.

91. Gerbe, F., et al., *Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites.* Nature, 2016. **529**(7585): p. 226-30.

92. Howitt, M.R., et al., *Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut.* Science, 2016. **351**(6279): p. 1329-33.

93. Ueo, T., et al., *The role of Hes genes in intestinal development, homeostasis and tumor formation.* Development, 2012. **139**(6): p. 1071-1082.

94. Zecchini, V., et al., *Notch signaling regulates the differentiation of post-mitotic intestinal epithelial cells.* Genes & Development, 2005. **19**(14): p. 1686-1691.

95. Fre, S., et al., *Notch signals control the fate of immature progenitor cells in the intestine.* Nature, 2005. **435**(7044): p. 964-8.

96. Gao, X., et al., *Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation.* Mol Cell Biol, 1998. **18**(5): p. 2901-11.

97. Beuling, E., et al., *GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice.* Gastroenterology, 2011. **140**(4): p. 1219-1229 e1-2.

98. Walker, E.M., C.A. Thompson, and M.A. Battle, *GATA4 and GATA6 regulate intestinal epithelial cytodifferentiation during development.* Dev Biol, 2014. **392**(2): p. 283-94.

99. Barker, N., M. van de Wetering, and H. Clevers, *The intestinal stem cell.* Genes & Development, 2008. **22**(14): p. 1856-1864.

100. Daniel, H. and T. Zietek, *Taste and move: glucose and peptide transporters in the gastrointestinal tract.* Exp Physiol, 2015. **100**(12): p. 1441-50.

101. SANDLE, G.I., *Salt and water absorption in the human colon: a modern appraisal.* Gut, 1998. **43**(2): p. 294-299.

102. Iqbal, J. and M.M. Hussain, *Intestinal lipid absorption.* American Journal of Physiology - Endocrinology And Metabolism, 2009. **296**(6): p. E1183-E1194.

103. Jong, M.C., M.H. Hofker, and L.M. Havekes, *Role of ApoCs in Lipoprotein Metabolism.* Functional Differences Between ApoC1, ApoC2, and ApoC3, 1999. **19**(3): p. 472-484.

104. Xu, X., et al., *Transcriptional regulation of apolipoprotein A-IV by the transcription factor CREBH.* Journal of Lipid Research, 2014. **55**(5): p. 850-859.

105. Kikuchi, T., et al., *Intestinal CREBH overexpression prevents high-cholesterol diet-induced hypercholesterolemia by reducing Npc1l1 expression.* Mol Metab, 2016. **5**(11): p. 1092-1102.

106. Alrefai, W.A., et al., *Modulation of human Niemann-Pick C1-like 1 gene expression by sterol: Role of sterol regulatory element binding protein 2.* American Journal of Physiology - Gastrointestinal and Liver Physiology, 2007. **292**(1): p. G369-G376.

107. Shimano, H., et al., *Sterol Regulatory Element-binding Protein-1 as a Key Transcription Factor for Nutritional Induction of Lipogenic Enzyme Genes.* Journal of Biological Chemistry, 1999. **274**(50): p. 35832-35839.

108. Cruz-Garcia, L. and A. Schlegel, *Lxr-driven enterocyte lipid droplet formation delays transport of ingested lipids.* J Lipid Res, 2014. **55**(9): p. 1944-58.

109. Bosse, T., et al., *Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine.* Mol Cell Biol, 2006. **26**(23): p. 9060-70.

110. Kim, I., et al., *Differential regulation of bile acid homeostasis by the farnesoid X receptor in liver and intestine.* J Lipid Res, 2007. **48**(12): p. 2664-72.

111. Modica, S., et al., *Selective Activation of Nuclear Bile Acid Receptor FXR in the Intestine Protects Mice Against Cholestasis.* Gastroenterology, 2012. **142**(2): p. 355-365.e4.

112. Schaap, F.G., M. Trauner, and P.L.M. Jansen, *Bile acid receptors as targets for drug development.* Nat Rev Gastroenterol Hepatol, 2014. **11**(1): p. 55-67.

113. Sayin, Sama I., et al., *Gut Microbiota Regulates Bile Acid Metabolism by Reducing the Levels of Tauro-beta-muricholic Acid, a Naturally Occurring FXR Antagonist.* Cell Metabolism, 2013. **17**(2): p. 225-235.

114. Mowat, A.M. and W.W. Agace, *Regional specialization within the intestinal immune system.* Nat Rev Immunol, 2014. **14**(10): p. 667-685.

115. Brennan, P.C., J.S. McCullough, and K.E. Carr, *Variations in Cell and Structure Populations along the Length of Murine Small Intestine.* Cells Tissues Organs, 1999. **164**(4): p. 221-226.

116. Bansemir, A.D. and M.V. Sukhdeo, *Villus length influences habitat selection by Heligmosomoides polygyrus.* Parasitology, 1996. **113 (Pt 3)**: p. 311-6.

117. Garrison, W.D., et al., *Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon.* Gastroenterology, 2006. **130**(4): p. 1207-20.

118. Frochot, V., et al., *The transcription factor HNF-4alpha: a key factor of the intestinal uptake of fatty acids in mouse.* Am J Physiol Gastrointest Liver Physiol, 2012. **302**(11): p. G1253-63.

119. Larsson, E., et al., *Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88.* Gut, 2012. **61**(8): p. 1124-31.

120. El Aidy, S., et al., *The gut microbiota elicits a profound metabolic reorientation in the mouse jejunal mucosa during conventionalisation.* Gut, 2013. **62**(9): p. 1306-14.

121. El Aidy, S., et al., *Temporal and spatial interplay of microbiota and intestinal mucosa drive establishment of immune homeostasis in conventionalized mice.* Mucosal Immunol, 2012. **5**(5): p. 567-79.

122. Beyaz, S., et al., *High-fat diet enhances stemness and tumorigenicity of intestinal progenitors.* Nature, 2016. **531**(7592): p. 53-58.

123. Russell, W.R., et al., *High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health.* The American Journal of Clinical Nutrition, 2011. **93**(5): p. 1062-1072.

124. Duncan, S.H., et al., *Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces.* Appl Environ Microbiol, 2007. **73**(4): p. 1073-8.

125. Walker, A.W., et al., *Dominant and diet-responsive groups of bacteria within the human colonic microbiota.* ISME J, 2011. **5**(2): p. 220-230.

126. Sayin, S.I., et al., *Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist.* Cell Metab, 2013. **17**(2): p. 225-35.

127. David, L.A., et al., *Diet rapidly and reproducibly alters the human gut microbiome.* Nature, 2014. **505**(7484): p. 559-63.

128. David, L.A., et al., *Host lifestyle affects human microbiota on daily timescales.* Genome Biol, 2014. **15**(7): p. R89.

129. Devillard, E., et al., *Metabolism of linoleic acid by human gut bacteria: different routes for biosynthesis of conjugated linoleic acid.* J Bacteriol, 2007. **189**(6): p. 2566-70.

130. Albenberg, L., et al., *Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota.* Gastroenterology, 2014. **147**(5): p. 1055-1063.e8.

131. Mukherji, A., et al., *Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs.* Cell, 2013. **153**(4): p. 812-27.

132. Shulzhenko, N., et al., *Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut.* Nat Med, 2011. **17**(12): p. 1585-93.

133. Longo, V.D. and S. Panda, *Fasting, Circadian Rhythms, and Time-Restricted Feeding in Healthy Lifespan.* Cell Metab, 2016. **23**(6): p. 1048-59.

134. Vijay-Kumar, M., et al., *Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5.* Science, 2010. **328**(5975): p. 228-31.

135. Stroeve, J.H.M., et al., *Intestinal FXR-mediated FGF15 production contributes to diurnal control of hepatic bile acid synthesis in mice.* Lab Invest, 2010. **90**(10): p. 1457-1467.

136. Cheesman, S.E., et al., *Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4570-4577.

137. Rawls, J.F., B.S. Samuel, and J.I. Gordon, *Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota.* Proc. Natl. Acad. Sci. U. S. A., 2004. **101**(13): p. 4596-601.

138. Rawls, J.F., et al., *In vivo imaging and genetic analysis link bacterial motility and symbiosis in the zebrafish gut.* Proc. Natl. Acad. Sci. U. S. A., 2007. **104**(18): p. 7622-7.

139. Keilbaugh, S.A., et al., *Activation of RegIIIβ/γ and interferon γ expression in the intestinal tract of SCID mice: an innate response to bacterial colonisation of the gut.* Gut, 2005. **54**(5): p. 623-629.

140. Rawls, J.F., et al., *Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection.* Cell, 2006. **127**(2): p. 423-33.

141. Thaiss, Christoph A., et al., *Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations.* Cell, 2016. **167**(6): p. 1495-1510.e12.

142. Krautkramer, K.A., et al., *Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues.* Mol Cell, 2016. **64**(5): p. 982-992.

143. Takahashi, K., et al., *Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells.* J Biol Chem, 2011. **286**(41): p. 35755-62.

144. Yu, D.H., et al., *Postnatal epigenetic regulation of intestinal stem cells requires DNA methylation and is guided by the microbiome.* Genome Biol, 2015. **16**: p. 211.

145. Camp, J.G., et al., *Intronic cis-regulatory modules mediate tissue-specific and microbial control of angptl4/fiaf transcription.* PLoS Genet, 2012. **8**(3): p. e1002585.

146. Kanther, M., et al., *Microbial colonization induces dynamic temporal and spatial patterns of NF-kappaB activation in the zebrafish digestive tract.* Gastroenterology, 2011. **141**(1): p. 197-207.

147. Evans, R.M. and D.J. Mangelsdorf, *Nuclear Receptors, RXR, and the Big Bang.* Cell, 2014. **157**(1): p. 255-66.

148. Bridgham, J.T., et al., *Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor.* PLoS Biol, 2010. **8**(10).

149. Makishima, M., et al., *Identification of a Nuclear Receptor for Bile Acids.* Science, 1999. **284**(5418): p. 1362-1365.

150. Wahlstrom, A., et al., *Induction of farnesoid X receptor signaling in germ-free mice colonized with a human microbiota.* J Lipid Res, 2017. **58**(2): p. 412-419.

151. Li, F., et al., *Microbiome remodelling leads to inhibition of intestinal farnesoid X receptor signalling and decreased obesity.* Nat Commun, 2013. **4**: p. 2384.

152. Degirolamo, C., et al., *Microbiota modification with probiotics induces hepatic bile acid synthesis via downregulation of the Fxr-Fgf15 axis in mice.* Cell Rep, 2014. **7**(1): p. 12-8.

153. Parséus, A., et al., *Microbiota-induced obesity requires farnesoid X receptor.* Gut, 2016.

154. Devlin, A.S. and M.A. Fischbach, *A biosynthetic pathway for a prominent class of microbiota-derived bile acids.* Nat Chem Biol, 2015. **11**(9): p. 685-690.

155. Alex, S., et al., *Short-chain fatty acids stimulate angiopoietin-like 4 synthesis in human colon adenocarcinoma cells by activating peroxisome proliferator-activated receptor gamma.* Mol Cell Biol, 2013. **33**(7): p. 1303-16.

156. Bäckhed, F., et al., *The gut microbiota as an environmental factor that regulates fat storage.* Proc Natl Acad Sci U S A, 2004. **101**(44): p. 15718-23.

157. Aronsson, L., et al., *Decreased Fat Storage by Lactobacillus Paracasei Is Associated with Increased Levels of Angiopoietin-Like 4 Protein (ANGPTL4).* PLOS ONE, 2010. **5**(9): p. e13087.

158. Korecka, A., et al., *ANGPTL4 expression induced by butyrate and rosiglitazone in human intestinal epithelial cells utilizes independent pathways.* Am J Physiol Gastrointest Liver Physiol, 2013. **304**(11): p. G1025-37.

159. Nakamoto, M., et al., *The Glucocorticoid Receptor Regulates the ANGPTL4 Gene in a CTCF-Mediated Chromatin Context in Human Hepatic Cells.* PLoS One, 2017. **12**(1): p. e0169225.

160. Koliwad, S.K., et al., *Angiopoietin-like 4 (ANGPTL4, Fasting-induced Adipose Factor) Is a Direct Glucocorticoid Receptor Target and Participates in Glucocorticoid-regulated Triglyceride Metabolism.* Journal of Biological Chemistry, 2009. **284**(38): p. 25593-25601.

161. Albenberg, L., et al., *Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota.* Gastroenterology, 2014. **147**(5): p. 1055-63 e8.

162. Kelly, C.J., et al., *Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function.* Cell Host Microbe, 2015. **17**(5): p. 662-71.

163. Louis, N.A., et al., *Selective induction of mucin-3 by hypoxia in intestinal epithelia.* J Cell Biochem, 2006. **99**(6): p. 1616-27.

164. Kelly, C.J., et al., *Fundamental role for HIF-1alpha in constitutive expression of human beta defensin-1.* Mucosal Immunol, 2013. **6**(6): p. 1110-8.

165. Furuta, G.T., et al., *Hypoxia-inducible factor 1-dependent induction of intestinal trefoil factor protects barrier function during hypoxia.* J Exp Med, 2001. **193**(9): p. 1027-34.

166. Fan, D., et al., *Activation of HIF-1alpha and LL-37 by commensal bacteria inhibits Candida albicans colonization.* Nat Med, 2015. **21**(7): p. 808-14.

167. Rabot, S., et al., *Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism.* FASEB J, 2010. **24**(12): p. 4948-59.

168. Semova, I., et al., *Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish.* Cell Host Microbe, 2012. **12**(3): p. 277-88.

169. Backhed, F., et al., *Mechanisms underlying the resistance to diet-induced obesity in germ-free mice.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 979-84.

170. Cox, L.M., et al., *Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences.* Cell, 2014. **158**(4): p. 705-21.

171. Cho, I., et al., *Antibiotics in early life alter the murine colonic microbiome and adiposity.* Nature, 2012. **488**(7413): p. 621-6.

172. Erkosar, B., et al., *Drosophila microbiota modulates host metabolic gene expression via IMD/NF-kappaB signaling.* PLoS One, 2014. **9**(4): p. e94729.

173. Takeuchi, O. and S. Akira, *Pattern Recognition Receptors and Inflammation.* Cell. **140**(6): p. 805-820.

174. Rakoff-Nahoum, S., et al., *Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis.* Cell, 2004. **118**(2): p. 229-41.

175. Girardelli, M., et al., *Novel missense mutation in the NOD2 gene in a patient with early onset ulcerative colitis: causal or chance association?* Int J Mol Sci, 2014. **15**(3): p. 3834-41.

176. Andriulli, A., et al., *The frame-shift mutation of the NOD2/CARD15 gene is significantly increased in ulcerative colitis: an *IG-IBD study.* Gastroenterology, 2004. **126**(2): p. 625-7.

177. Corridoni, D., et al., *Genetic deletion of the bacterial sensor NOD2 improves murine Crohn/'s disease-like ileitis independent of functional dysbiosis.* Mucosal Immunol, 2016.

178. Petnicki-Ocwieja, T., et al., *Nod2 is required for the regulation of commensal microbiota in the intestine.* Proceedings of the National Academy of Sciences, 2009. **106**(37): p. 15813-15818.

179. Maeda, S., et al., *Nod2 Mutation in Crohn's Disease Potentiates NF-κB Activity and IL-1ß Processing.* Science, 2005. **307**(5710): p. 734-738.

180. Rakoff-Nahoum, S., et al., *Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis.* Cell. **118**(2): p. 229-241.

181. Vaishnava, S., et al., *Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface.* Proceedings of the National Academy of Sciences, 2008. **105**(52): p. 20858-20863.

182. Sipos, F., et al., *Contribution of TLR signaling to the pathogenesis of colitis-associated cancer in inflammatory bowel disease.* World J Gastroenterol, 2014. **20**(36): p. 12713-21.

183. Abreu, M.T., *Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function.* Nat Rev Immunol, 2010. **10**(2): p. 131-144.

184. Sun, Y., et al., *TLR4 and TLR5 on corneal macrophages regulate Pseudomonas aeruginosa keratitis by signaling through MyD88-dependent and -independent pathways.* J Immunol, 2010. **185**(7): p. 4272-83.

185. Sachdev, U., et al., *TLR2 and TLR4 mediate differential responses to limb ischemia through MyD88-dependent and independent pathways.* PLoS One, 2012. **7**(11): p. e50654.

186. Yamamoto, M., et al., *Role of Adaptor TRIF in the MyD88-Independent Toll-Like Receptor Signaling Pathway.* Science, 2003. **301**(5633): p. 640-643.

187. Janssens, S. and R. Beyaert, *A universal role for MyD88 in TLR/IL-1R-mediated signaling.* Trends in Biochemical Sciences, 2002. **27**(9): p. 474-482.

188. Lebeis, S.L., et al., *TLR signaling mediated by MyD88 is required for a protective innate immune response by neutrophils to Citrobacter rodentium.* J Immunol, 2007. **179**(1): p. 566-77.

189. Pott, J. and M. Hornef, *Innate immune signalling at the intestinal epithelium in homeostasis and disease.* EMBO Rep, 2012. **13**(8): p. 684-98.

190. Mogensen, T.H., *Pathogen recognition and inflammatory signaling in innate immune defenses.* Clin Microbiol Rev, 2009. **22**(2): p. 240-73, Table of Contents.

191. Lavelle, E.C., et al., *The role of TLRs, NLRs, and RLRs in mucosal innate immunity and homeostasis.* Mucosal Immunol, 2010. **3**(1): p. 17-28.

192. Yamamoto, M., et al., *TRAM is specifically involved in the Toll-like receptor 4-mediated MyD88-independent signaling pathway.* Nat Immunol, 2003. **4**(11): p. 1144-1150.

193. Ben-Neriah, Y. and M. Karin, *Inflammation meets cancer, with NF-[kappa]B as the matchmaker.* Nat Immunol, 2011. **12**(8): p. 715-723.

194. Kanther, M., et al., *Commensal microbiota stimulate systemic neutrophil migration through induction of serum amyloid A.* Cell Microbiol, 2014. **16**(7): p. 1053-67.

195. Kaci, G., et al., *Inhibition of the NF-kappaB pathway in human intestinal epithelial cells by commensal Streptococcus salivarius.* Appl Environ Microbiol, 2011. **77**(13): p. 4681-4.

196. Ryu, J.H., et al., *Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in Drosophila.* Science, 2008. **319**(5864): p. 777-82.

197. Nenci, A., et al., *Epithelial NEMO links innate immunity to chronic intestinal inflammation.* Nature, 2007. **446**(7135): p. 557-561.

198. Vereecke, L., R. Beyaert, and G. van Loo, *The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology.* Trends in Immunology. **30**(8): p. 383-391.

199. Panne, D., T. Maniatis, and S.C. Harrison, *Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-β enhancer.* The EMBO Journal, 2004. **23**(22): p. 4384-4393.

200. Ha, E.M., et al., *Coordination of multiple dual oxidase-regulatory pathways in responses to commensal and infectious microbes in drosophila gut.* Nat Immunol, 2009. **10**(9): p. 949-57.

201. Rada, B. and T.L. Leto, *Oxidative innate immune defenses by Nox/Duox family NADPH oxidases.* Contrib Microbiol, 2008. **15**: p. 164-87.

202. Geiszt, M., et al., *Dual oxidases represent novel hydrogen peroxide sources supporting mucosal surface host defense.* FASEB J, 2003. **17**(11): p. 1502-4.

203. Lo, H.W., et al., *Nuclear interaction of EGFR and STAT3 in the activation of the iNOS/NO pathway.* Cancer Cell, 2005. **7**(6): p. 575-89.

204. Choi, S.M., et al., *Innate Stat3-mediated induction of the antimicrobial protein Reg3gamma is required for host defense against MRSA pneumonia.* J Exp Med, 2013. **210**(3): p. 551-61.

205. Cash, H.L., et al., *Symbiotic bacteria direct expression of an intestinal bactericidal lectin.* Science, 2006. **313**(5790): p. 1126-30.

206. Lee, K.S., et al., *Helicobacter pylori CagA triggers expression of the bactericidal lectin REG3gamma via gastric STAT3 activation.* PLoS One, 2012. **7**(2): p. e30786.

207. van Es, J.H., et al., *Wnt signalling induces maturation of Paneth cells in intestinal crypts.* Nat Cell Biol, 2005. **7**(4): p. 381-6.

208. Sharma, R., et al., *Rat intestinal mucosal responses to a microbial flora and different diets.* Gut, 1995. **36**(2): p. 209-214.

209. Bates, J.M., et al., *Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation.* Developmental Biology, 2006. **297**(2): p. 374-386.

210. Bäckhed, F., et al., *Postnatal lymphatic partitioning from the blood vasculature in the small intestine requires fasting-induced adipose factor.* Proc Natl Acad Sci U S A, 2007.

211. Zhou, Y., et al., *Nuclear factor of activated T-cells 5 increases intestinal goblet cell differentiation through an mTOR/Notch signaling pathway.* Mol Biol Cell, 2014. **25**(18): p. 2882-90.

212. Zhou, Y., et al., *TSC2/mTORC1 signaling controls Paneth and goblet cell differentiation in the intestinal epithelium.* Cell Death Dis, 2015. **6**: p. e1631.

213. Buchon, N., et al., *Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in Drosophila.* Genes Dev, 2009. **23**(19): p. 2333-44.

214. Fukada, T., et al., *STAT3 orchestrates contradictory signals in cytokine-induced G₁ to S cell-cycle transition.* The EMBO Journal, 1998. **17**(22): p. 6670-6677.

215. Reikvam, D.H., et al., *Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression.* PLoS One, 2011. **6**(3): p. e17996.

216. Bullen, T.F., et al., *Characterization of epithelial cell shedding from human small intestine.* Lab Invest, 2006. **86**(10): p. 1052-1063.

217. Demehri, F.R., et al., *Intestinal epithelial cell apoptosis and loss of barrier function in the setting of altered microbiota with enteral nutrient deprivation.* Front Cell Infect Microbiol, 2013. **3**: p. 105.

218. Kramer, B., K. Wiegmann, and M. Kronke, *Regulation of the human TNF promoter by the transcription factor Ets.* J Biol Chem, 1995. **270**(12): p. 6577-83.

219. Hussain, M.M. and X. Pan, *Clock genes, intestinal transport and plasma lipid homeostasis.* Trends Endocrinol Metab, 2009. **20**(4): p. 177-85.

220. Thaiss, C.A., et al., *Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis.* Cell, 2014. **159**(3): p. 514-29.

221. Sartor, R.B. and G.D. Wu, *Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches.* Gastroenterology, 2016.

222. Ukhanova, M., et al., *Gut microbiota correlates with energy gain from dietary fibre and appears to be associated with acute and chronic intestinal diseases.* Clin Microbiol Infect, 2012. **18 Suppl 4**: p. 62-6.

223. Kamada, N., et al., *Role of the gut microbiota in immunity and inflammatory disease.* Nat Rev Immunol, 2013. **13**(5): p. 321-35.

224. Costea, I., et al., *Interactions between the dietary polyunsaturated fatty acid ratio and genetic factors determine susceptibility to pediatric Crohn's disease.* Gastroenterology, 2014. **146**(4): p. 929-31.

225. Kinger, M., S. Kumar, and V. Kumar, *Some Important Dietary Polyphenolic Compounds: An Anti-inflammatory and Immunoregulatory Perspective.* Mini Rev Med Chem, 2017.

226. Ankersen, D.V., et al., *Using eHealth strategies in delivering dietary and other therapies in patients with irritable bowel syndrome and inflammatory bowel disease.* J Gastroenterol Hepatol, 2017. **32 Suppl 1**: p. 27-31.

227. Gibson, P.R., *Use of the low-FODMAP diet in inflammatory bowel disease.* J Gastroenterol Hepatol, 2017. **32 Suppl 1**: p. 40-42.

228. Wang, A.Y., J. Popov, and N. Pai, *Fecal microbial transplant for the treatment of pediatric inflammatory bowel disease.* World J Gastroenterol, 2016. **22**(47): p. 10304-10315.

229. Gordon, H. and M. Harbord, *A patient with severe Crohn's colitis responds to Faecal Microbiota Transplantation.* J Crohns Colitis, 2014. **8**(3): p. 256-7.

230. Mokry, M., et al., *Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium.* Gastroenterology, 2014. **146**(4): p. 1040-7.

231. Meddens, C.A., et al., *Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease.* Genome Biol, 2016. **17**(1): p. 247.

232. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-24.

233. Barrett, J.C., et al., *Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region.* Nat Genet, 2009. **41**(12): p. 1330-4.

234. Kraiczy, J., et al., *Assessing DNA methylation in the developing human intestinal epithelium: potential link to inflammatory bowel disease.* Mucosal Immunol, 2016. **9**(3): p. 647-58.

235. Turgeon, N., et al., *The acetylome regulators Hdac1 and Hdac2 differently modulate intestinal epithelial cell dependent homeostatic responses in experimental colitis.* Am J Physiol Gastrointest Liver Physiol, 2014. **306**(7): p. G594-605.

236. Alenghat, T., et al., *Histone deacetylase 3 coordinates commensal-bacteria-dependent intestinal homeostasis.* Nature, 2013. **504**(7478): p. 153-7.

237. Haberman, Y., et al., *Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature.* J Clin Invest, 2014. **124**(8): p. 3617-33.

238. Arijs, I., et al., *Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment.* PLoS One, 2009. **4**(11): p. e7984.

239. Coskun, M., et al., *Involvement of JAK/STAT signaling in the pathogenesis of inflammatory bowel disease.* Pharmacological Research, 2013. **76**: p. 1-8.

240. Wullaert, A., *Role of NF-kappaB activation in intestinal immune homeostasis.* Int J Med Microbiol, 2010. **300**(1): p. 49-56.

241. Atreya, I., R. Atreya, and M.F. Neurath, *NF-kappaB in inflammatory bowel disease.* J Intern Med, 2008. **263**(6): p. 591-6.

242. Andersen, V., et al., *Polymorphisms in NF-kappaB, PXR, LXR, PPARgamma and risk of inflammatory bowel disease.* World J Gastroenterol, 2011. **17**(2): p. 197-206.

243. Yh, Y., et al., *Endogenous estrogen regulation of inflammatory arthritis and cytokine expression is predominantly mediated via estrogen receptor alpha*, in *Arthritis Rheum*. 2010.

244. Wang, Q., et al., *Vitamin D inhibits COX-2 expression and inflammatory response by targeting thioesterase superfamily member 4.* Journal of Biological Chemistry, 2014.

245. Martin, H., *Role of PPAR-gamma in inflammation. Prospects for therapeutic intervention by food components.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2010. **690**(1–2): p. 57-63.

246. Barish, G.D., et al., *PPARdelta regulates multiple proinflammatory pathways to suppress atherosclerosis.* Proc Natl Acad Sci U S A, 2008. **105**(11): p. 4271-6.

247. Yoon, C.-H., et al., *Activation of Liver X Receptors Suppresses Inflammatory Gene Expressions and Transcriptional Corepressor Clearance in Rheumatoid Arthritis Fibroblast Like Synoviocytes.* Journal of Clinical Immunology, 2013. **33**(1): p. 190-199.

248. Wang, Z., E.P. Bishop, and P.A. Burke, *Expression profile analysis of the inflammatory response regulated by hepatocyte nuclear factor 4alpha.* BMC Genomics, 2011. **12**: p. 128.

249. Babeu, J.P. and F. Boudreau, *Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks.* World J Gastroenterol, 2014. **20**(1): p. 22-30.

250. Hatziapostolou, M., et al., *An HNF4alpha-miRNA inflammatory feedback circuit regulates hepatocellular oncogenesis.* Cell, 2011. **147**(6): p. 1233-47.

251. Darsigny, M., et al., *Loss of hepatocyte-nuclear-factor-4alpha affects colonic ion transport and causes chronic inflammation resembling inflammatory bowel disease in mice.* PLoS One, 2009. **4**(10): p. e7609.

252. Ogawa, S., et al., *Molecular determinants of crosstalk between nuclear receptors and toll-like receptors.* Cell, 2005. **122**(5): p. 707-21.

253. Gadaleta, R.M., et al., *Farnesoid X receptor activation inhibits inflammation and preserves the intestinal barrier in inflammatory bowel disease.* Gut, 2011. **60**(4): p. 463-72.

254. Vavassori, P., et al., *The bile acid receptor FXR is a modulator of intestinal innate immunity.* J Immunol, 2009. **183**(10): p. 6251-61.

255. Jakobsson, T., et al., *The oxysterol receptor LXRbeta protects against DSS- and TNBS-induced colitis in mice.* Mucosal Immunol, 2014. **7**(6): p. 1416-28.

256. Marcil, V., et al., *Association between genetic variants in the HNF4A gene and childhood-onset Crohn's disease.* Genes Immun, 2012. **13**(7): p. 556-65.

257. Franke, A., et al., *Systematic association mapping identifies NELL1 as a novel IBD disease gene.* PLoS One, 2007. **2**(8): p. e691.

258. Ahn, S.H., et al., *Hepatocyte nuclear factor 4alpha in the intestinal epithelial cells protects against inflammatory bowel disease.* Inflamm Bowel Dis, 2008. **14**(7): p. 908-20.

259. Darsigny, M., et al., *Hepatocyte nuclear factor-4alpha promotes gut neoplasia in mice and protects against the production of reactive oxygen species.* Cancer Res, 2010. **70**(22): p. 9423-33.

260. Chahar, S., et al., *Chromatin profiling reveals regulatory network shifts and a protective role for hepatocyte nuclear factor 4alpha during colitis.* Mol Cell Biol, 2014. **34**(17): p. 3291-304.

261. Koukos, G., et al., *MicroRNA-124 regulates STAT3 expression and is down-regulated in colon tissues of pediatric patients with ulcerative colitis.* Gastroenterology, 2013. **145**(4): p. 842-52 e2.

262. Gkouskou, K.K., et al., *Apolipoprotein A-I inhibits experimental colitis and colitis-propelled carcinogenesis.* Oncogene, 2016. **35**(19): p. 2496-505.

263. Liu, D., et al., *The apolipoprotein A-I mimetic peptide, D-4F, alleviates ox-LDL-induced oxidative stress and promotes endothelial repair through the eNOS/HO-1 pathway.* J Mol Cell Cardiol, 2017.

264. Glas, J., et al., *Role of PPARG gene variants in inflammatory bowel disease.* Inflamm Bowel Dis, 2011. **17**(4): p. 1057-8.

265. Bank, S., et al., *Polymorphisms in the inflammatory pathway genes TLR2, TLR4, TLR9, LY96, NFKBIA, NFKB1, TNFA, TNFRSF1A, IL6R, IL10, IL23R, PTPN22, and PPARG are associated with susceptibility of inflammatory bowel disease in a Danish cohort.* PLoS One, 2014. **9**(6): p. e98815.

266. Dou, X., et al., *Peroxisome proliferator-activated receptor-gamma is downregulated in ulcerative colitis and is involved in experimental colitis-associated neoplasia.* Oncol Lett, 2015. **10**(3): p. 1259-1266.

267. Kelly, D., et al., *Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA.* Nat Immunol, 2004. **5**(1): p. 104-12.

268. Su, C.G., et al., *A novel therapy for colitis utilizing PPAR-gamma ligands to inhibit the epithelial inflammatory response.* J Clin Invest, 1999. **104**(4): p. 383-9.

269. Celinski, K., et al., *Comparison of the anti-inflammatory and therapeutic actions of PPAR-gamma agonists rosiglitazone and troglitazone in experimental colitis.* J Physiol Pharmacol, 2012. **63**(6): p. 631-40.

270. Wong, C., P.J. Harris, and L.R. Ferguson, *Potential Benefits of Dietary Fibre Intervention in Inflammatory Bowel Disease.* Int J Mol Sci, 2016. **17**(6).

271. Bassaganya-Riera, J., et al., *Conjugated linoleic acid modulates immune responses in patients with mild to moderately active Crohn's disease.* Clin Nutr, 2012. **31**(5): p. 721-7.

272. Wu, J.H., et al., *Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.* Circ Cardiovasc Genet., 2013. **6**.

273. Smoak, K.A. and J.A. Cidlowski, *Mechanisms of glucocorticoid receptor signaling during inflammation.* Mechanisms of Ageing and Development, 2004. **125**(10–11): p. 697-706.

274. Matsumoto, T., et al., *Regulation of osteoblast differentiation by interleukin-11 via AP-1 and Smad signaling.* Endocr J, 2012. **59**(2): p. 91-101.

275. Sakurai, H., et al., *Suppression of NF-kappa B and AP-1 activation by glucocorticoids in experimental glomerulonephritis in rats: molecular mechanisms of anti-nephritic action.* Biochim Biophys Acta, 1997. **1362**(2-3): p. 252-62.

276. Mandrekar, P., G. Bellerose, and G. Szabo, *Inhibition of NF-kappa B binding correlates with increased nuclear glucocorticoid receptor levels in acute alcohol-treated human monocytes.* Alcohol Clin Exp Res, 2002. **26**(12): p. 1872-9.

277. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nat Genet, 2010. **42**(12): p. 1118-25.

278. Barnes, P.J. and I.M. Adcock, *Glucocorticoid resistance in inflammatory diseases.* Lancet, 2009. **373**(9678): p. 1905-17.

279. Mwinyi, J., et al., *Glucocorticoid receptor gene haplotype structure and steroid therapy outcome in IBD patients.* World J Gastroenterol, 2010. **16**(31): p. 3888-96.

280. Krupoves, A., et al., *Variation in the glucocorticoid receptor gene (NR3C1) may be associated with corticosteroid dependency and resistance in children with Crohn's disease.* Pharmacogenet Genomics, 2011. **21**(8): p. 454-60.

281. Orii, F., et al., *Quantitative analysis for human glucocorticoid receptor alpha/beta mRNA in IBD.* Biochem Biophys Res Commun, 2002. **296**(5): p. 1286-94.

282. Bastl, C.P., et al., *Glucocorticoid stimulation of sodium absorption in colon epithelia is mediated by corticosteroid IB receptor.* J Biol Chem, 1984. **259**(2): p. 1186-95.

283. Zemanova, Z. and J. Pacha, *Corticosteroid induction of renal and intestinal K(+)-dependent p-nitrophenylphosphatase in young and adult rats.* Histochem J, 1996. **28**(9): p. 625-34.

284. Sanapareddy, N., et al., *Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans.* ISME J, 2012. **6**(10): p. 1858-68.

285. Marjoram, L., et al., *Epigenetic control of intestinal barrier function and inflammation in zebrafish.* Proc Natl Acad Sci U S A, 2015. **112**(9): p. 2770-5.

286. Bates, J.M., et al., *Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota.* Cell Host Microbe, 2007. **2**(6): p. 371-82.

287. Morgun, A., et al., *Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks.* Gut, 2015. **64**(11): p. 1732-43.

288. Evans, R.M. and D.J. Mangelsdorf, *Nuclear Receptors, RXR, and the Big Bang.* Cell, 2014. **157**(1): p. 255-266.

289. Stegmann, A., et al., *Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation.* Physiol Genomics, 2006. **27**(2): p. 141-55.

290. Chellappa, K., et al., *Opposing roles of nuclear receptor HNF4alpha isoforms in colitis and colitis-associated colon cancer.* Elife, 2016. **5**.

291. Berndt, S.I., et al., *Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture.* Nat Genet, 2013. **45**(5): p. 501-12.

292. San Roman, A.K., et al., *Transcription factors GATA4 and HNF4A control distinct aspects of intestinal homeostasis in conjunction with transcription factor CDX2.* J Biol Chem, 2015. **290**(3): p. 1850-60.

293. Duncan, S.A., A. Nagy, and W. Chan, *Murine gastrulation requires HNF-4 regulated gene expression in the visceral endoderm: tetraploid rescue of Hnf-4(-/-) embryos.* Development, 1997. **124**(2): p. 279-87.

294. Klapper, M., et al., *The human intestinal fatty acid binding protein (hFABP2) gene is regulated by HNF-4alpha.* Biochem Biophys Res Commun, 2007. **356**(1): p. 147-52.

295. Plevy, S., et al., *Combined serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients.* Inflamm Bowel Dis, 2013. **19**(6): p. 1139-48.

296. Fang, B., et al., *Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors.* Nucleic Acids Res, 2012. **40**(12): p. 5343-56.

297. Tremblay, E., et al., *Gene expression profiling in necrotizing enterocolitis reveals pathways common to those reported in Crohn's disease.* BMC Med Genomics, 2016. **9**: p. 6.

298. Veilleux, A., et al., *Altered intestinal functions and increased local inflammation in insulin-resistant obese subjects: a gene-expression profile analysis.* BMC Gastroenterol, 2015. **15**: p. 119.

299. Staiger, H., et al., *Muscle-derived angiopoietin-like protein 4 is induced by fatty acids via peroxisome proliferator-activated receptor (PPAR)-delta and is of metabolic relevance in humans.* Diabetes, 2009. **58**(3): p. 579-89.

300. Hrabovsky, V., et al., *Cholesterol metabolism in active Crohn's disease.* Wien Klin Wochenschr, 2009. **121**(7-8): p. 270-5.

301. Agouridis, A.P., M. Elisaf, and H.J. Milionis, *An overview of lipid abnormalities in patients with inflammatory bowel disease.* Ann Gastroenterol, 2011. **24**(3): p. 181-187.

302. Nijmeijer, R.M., et al., *Farnesoid X receptor (FXR) activation and FXR genetic variation in inflammatory bowel disease.* PLoS One, 2011. **6**(8): p. e23745.

303. Weissglas-Volkov, D., et al., *Common hepatic nuclear factor-4alpha variants are associated with high serum lipid levels and the metabolic syndrome.* Diabetes, 2006. **55**(7): p. 1970-7.

304. Ma, R., et al., *Association of HNF4alpha gene polymorphisms with susceptibility to type 2 diabetes.* Mol Med Rep, 2016. **13**(3): p. 2241-6.

305. Vrieze, A., et al., *Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome.* Gastroenterology, 2012. **143**(4): p. 913-6.e7.

306. Qin, J., et al., *A metagenome-wide association study of gut microbiota in type 2 diabetes.* Nature, 2012. **490**(7418): p. 55-60.

307. Marcil, V., et al., *Modification in oxidative stress, inflammation, and lipoprotein assembly in response to hepatocyte nuclear factor 4alpha knockdown in intestinal epithelial cells.* J Biol Chem, 2010. **285**(52): p. 40448-60.

308. Palanker, L., et al., *Drosophila HNF4 regulates lipid mobilization and beta-oxidation.* Cell Metab, 2009. **9**(3): p. 228-39.

309. Barry, W.E. and C.S. Thummel, *The Drosophila HNF4 nuclear receptor promotes glucose-stimulated insulin secretion and mitochondrial function in adults.* Elife, 2016. **5**.

310. Soutoglou, E., N. Katrakili, and I. Talianidis, *Acetylation regulates transcription factor activity at multiple levels.* Mol Cell, 2000. **5**(4): p. 745-51.

311. Hong, Y.H., et al., *AMP-activated protein kinase regulates HNF4alpha transcriptional activity by inhibiting dimer formation and decreasing protein stability.* J Biol Chem, 2003. **278**(30): p. 27495-501.

312. Rha, G.B., et al., *Multiple binding modes between HNF4alpha and the LXXLL motifs of PGC-1alpha lead to full activation.* J Biol Chem, 2009. **284**(50): p. 35165-76.

313. Jager, S., et al., *AMP-activated protein kinase (AMPK) action in skeletal muscle via direct phosphorylation of PGC-1alpha.* Proc Natl Acad Sci U S A, 2007. **104**(29): p. 12017-22.

314. Huang, J., L.L. Levitsky, and D.B. Rhoads, *Novel P2 promoter-derived HNF4alpha isoforms with different N-terminus generated by alternate exon insertion.* Exp Cell Res, 2009. **315**(7): p. 1200-11.

315. Li, Y., et al., *RNA-Seq Analysis of Differential Splice Junction Usage and Intron Retentions by DEXSeq.* PLoS One, 2015. **10**(9): p. e0136653.

316. Yuan, X., et al., *Identification of an endogenous ligand bound to a native orphan nuclear receptor.* PLoS One, 2009. **4**(5): p. e5609.

317. Hertz, R., et al., *Fatty acyl-CoA thioesters are ligands of hepatic nuclear factor-4alpha.* Nature, 1998. **392**(6675): p. 512-6.

318. O'Shea, E.F., et al., *Production of bioactive substances by intestinal bacteria as a basis for explaining probiotic mechanisms: bacteriocins and conjugated linoleic acid.* Int J Food Microbiol, 2012. **152**(3): p. 189-205.

319. Mbodji, K., et al., *Adjunct therapy of n-3 fatty acids to 5-ASA ameliorates inflammatory score and decreases NF-kappaB in rats with TNBS-induced colitis.* J Nutr Biochem, 2013. **24**(4): p. 700-5.

320. Semenkovich, N.P., et al., *Impact of the gut microbiota on enhancer accessibility in gut intraepithelial lymphocytes.* Proc Natl Acad Sci U S A, 2016.

321. Westerfield, M., *The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio).* 4th ed. 2000: Univ. of Oregon Press, Eugene, OR.

322. Pham, L.N., et al., *Methods for generating and colonizing gnotobiotic zebrafish.* Nat Protoc, 2008. **3**(12): p. 1862-75.

323. Boyle, G., et al., *Comparative analysis of vertebrate diurnal/circadian transcriptomes.* PloS One, 2016. **In Press**.

324. Jao, L.E., S.R. Wente, and W. Chen, *Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system.* Proc Natl Acad Sci U S A, 2013. **110**(34): p. 13904-9.

325. Song, L. and G.E. Crawford, *DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells.* Cold Spring Harb Protoc, 2010. **2010**(2): p. pdb prot5384.

326. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.

327. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

328. Hu, Y., et al., *An integrative approach to ortholog prediction for disease-focused and other functional studies.* BMC Bioinformatics, 2011. **12**: p. 357.

329. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

330. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

331. Trapnell, C., et al., *Differential analysis of gene regulation at transcript resolution with RNA-seq.* Nat Biotech, 2013. **31**(1): p. 46-53.

332. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

333. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.

334. Gury-BenAri, M., et al., *The Spectrum and Regulatory Landscape of Intestinal Innate Lymphoid Cells Are Shaped by the Microbiome.* Cell, 2016. **166**(5): p. 1231-1246 e13.

335. Bienz, M., *TCF: transcriptional activator or repressor?* Current Opinion in Cell Biology, 1998. **10**(3): p. 366-372.

336. Benko, S., et al., *Molecular determinants of the balance between co-repressor and co-activator recruitment to the retinoic acid receptor.* J Biol Chem, 2003. **278**(44): p. 43797-806.

337. Lickwar, C.R., et al., *Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function.* Nature, 2012. **484**(7393): p. 251-5.

338. Lickwar, C.R., F. Mueller, and J.D. Lieb, *Genome-wide measurement of protein-DNA binding dynamics using competition ChIP.* Nat Protoc, 2013. **8**(7): p. 1337-53.

339. Li, J., et al., *Identification of the Flavonoid Luteolin as a Repressor of the Transcription Factor Hepatocyte Nuclear Factor 4α.* Journal of Biological Chemistry, 2015. **290**(39): p. 24021-24035.

340. Li, Q., et al., *Preservation of hepatocyte nuclear factor-4alpha contributes to the beneficial effect of dietary medium chain triglyceride on alcohol-induced hepatic lipid dyshomeostasis in rats.* Alcohol Clin Exp Res, 2013. **37**(4): p. 587-98.

341. Gil-Cardoso, K., et al., *Effects of flavonoids on intestinal inflammation, barrier integrity and changes in gut microbiota during diet-induced obesity.* Nutr Res Rev, 2016. **29**(2): p. 234-248.

342. Dean, S., et al., *Developmental and Tissue-Specific Regulation of Hepatocyte Nuclear Factor 4-α (HNF4-α) Isoforms in Rodents.* Gene Expression, 2010. **14**(6): p. 337-344.

343. Harries, L.W., et al., *The diabetic phenotype in HNF4A mutation carriers is moderated by the expression of HNF4A isoforms from the P1 promoter during fetal development.* Diabetes, 2008. **57**(6): p. 1745-52.

344. Eeckhoute, J., et al., *Hepatocyte nuclear factor 4 alpha isoforms originated from the P1 promoter are expressed in human pancreatic beta-cells and exhibit*

stronger transcriptional potentials than P2 promoter-driven isoforms. Endocrinology, 2003. **144**(5): p. 1686-94.

345. Briancon, N. and M.C. Weiss, *In vivo role of the HNF4alpha AF-1 activation domain revealed by exon swapping.* EMBO J, 2006. **25**(6): p. 1253-62.

346. Xie, X., et al., *Down-regulation of hepatic HNF4alpha gene expression during hyperinsulinemia via SREBPs.* Mol Endocrinol, 2009. **23**(4): p. 434-43.

347. Hansen, S.K., et al., *Genetic evidence that HNF-1alpha-dependent transcriptional control of HNF-4alpha is essential for human pancreatic beta cell function.* J Clin Invest, 2002. **110**(6): p. 827-33.

348. Austin, S. and J. St-Pierre, *PGC1alpha and mitochondrial metabolism--emerging concepts and relevance in ageing and neurodegenerative disorders.* J Cell Sci, 2012. **125**(Pt 21): p. 4963-71.

349. You, M., et al., *Transcriptional control of the human aldehyde dehydrogenase 2 promoter by hepatocyte nuclear factor 4: inhibition by cyclic AMP and COUP transcription factors.* Arch Biochem Biophys, 2002. **398**(1): p. 79-86.

350. Sun, K., et al., *Phosphorylation of a conserved serine in the deoxyribonucleic acid binding domain of nuclear receptors alters intracellular localization.* Mol Endocrinol, 2007. **21**(6): p. 1297-311.

351. Veto, B., et al., *The transcriptional activity of hepatocyte nuclear factor 4 alpha is inhibited via phosphorylation by ERK1/2.* PLoS One, 2017. **12**(2): p. e0172020.

352. Reddy, S., et al., *Mitogen-activated protein kinase regulates transcription of the ApoCIII gene. Involvement of the orphan nuclear receptor HNF4.* J Biol Chem, 1999. **274**(46): p. 33050-6.

353. Zhou, W., et al., *SUMOylation of HNF4alpha regulates protein stability and hepatocyte function.* J Cell Sci, 2012. **125**(Pt 15): p. 3630-5.

354. Giudici, M., et al., *Nuclear Receptor Coregulators in Metabolism and Disease.* Handb Exp Pharmacol, 2016. **233**: p. 95-135.

355. Anbalagan, M., et al., *Post-translational modifications of nuclear receptors and human disease.* Nucl Recept Signal, 2012. **10**: p. e001.

356. Ghisletti, S., et al., *Cooperative NCoR/SMRT interactions establish a corepressor-based strategy for integration of inflammatory and anti-inflammatory signaling pathways.* Genes Dev, 2009. **23**(6): p. 681-93.

357. Torres-Padilla, M.E., F.M. Sladek, and M.C. Weiss, *Developmentally regulated N-terminal variants of the nuclear receptor hepatocyte nuclear factor 4alpha mediate multiple interactions through coactivator and corepressor-histone deacetylase complexes.* J Biol Chem, 2002. **277**(47): p. 44677-87.

358. Abedin, S.A., et al., *Elevated NCOR1 disrupts a network of dietary-sensing nuclear receptors in bladder cancer cells.* Carcinogenesis, 2009. **30**(3): p. 449-56.

359. Mottis, A., L. Mouchiroud, and J. Auwerx, *Emerging roles of the corepressors NCoR1 and SMRT in homeostasis.* Genes Dev, 2013. **27**(8): p. 819-35.

360. Nepelska, M., et al., *Commensal gut bacteria modulate phosphorylation-dependent PPARgamma transcriptional activity in human intestinal epithelial cells.* Sci Rep, 2017. **7**: p. 43199.

361. Rawls, J.F., *Enteric infection and inflammation alter gut microbial ecology.* Cell Host & Microbe, 2007. **2**: p. 73-74.

362. Wajant, H., K. Pfizenmaier, and P. Scheurich, *Tumor necrosis factor signaling.* Cell Death Differ, 0000. **10**(1): p. 45-65.

363. Wang, Y. and J. Adjaye, *A cyclic AMP analog, 8-Br-cAMP, enhances the induction of pluripotency in human fibroblast cells.* Stem Cell Rev, 2011. **7**(2): p. 331-41.

364. Boulange, C.L., et al., *Impact of the gut microbiota on inflammation, obesity, and metabolic disease.* Genome Med, 2016. **8**(1): p. 42.

365. Samuel, B.S., et al., *Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16767-72.

366. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes, 2012. **3**(4): p. 289-306.

367. Regueiro, L., M. Carballa, and J.M. Lema, *Microbiome response to controlled shifts in ammonium and LCFA levels in co-digestion systems.* J Biotechnol, 2016. **220**: p. 35-44.

368. Stoffel, M. and S.A. Duncan, *The maturity-onset diabetes of the young (MODY1) transcription factor HNF4alpha regulates expression of genes required for glucose transport and metabolism.* Proc Natl Acad Sci U S A, 1997. **94**(24): p. 13209-14.

369. Zupancic, M.L., et al., *Analysis of the gut microbiota in the old order Amish and its relation to the metabolic syndrome.* PLoS One, 2012. **7**(8): p. e43052.

370. Fandriks, L., *Roles of the gut in the metabolic syndrome: an overview.* J Intern Med, 2017. **281**(4): p. 319-336.

371. Trico, D., et al., *Oxidized derivatives of linoleic acid in pediatric metabolic syndrome: is their pathogenic role modulated by the genetic background and the gut microbiota?* Antioxid Redox Signal, 2017.

372. Despres, J.P., et al., *Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk.* Arterioscler Thromb Vasc Biol, 2008. **28**(6): p. 1039-49.

373. Sun, H., et al., *Modulation of Microbiota-Gut-Brain Axis by Berberine Resulting in Improved Metabolic Status in High-Fat Diet-Fed Rats.* Obes Facts, 2016. **9**(6): p. 365-378.

374. Ley, R.E., et al., *Microbial ecology: human gut microbes associated with obesity.* Nature, 2006. **444**(7122): p. 1022-3.

375. Goossens, G.H., *The role of adipose tissue dysfunction in the pathogenesis of obesity-related insulin resistance.* Physiol Behav, 2008. **94**(2): p. 206-18.

376. Corpeleijn, E., W.H. Saris, and E.E. Blaak, *Metabolic flexibility in the development of insulin resistance and type 2 diabetes: effects of lifestyle.* Obes Rev, 2009. **10**(2): p. 178-93.

377. Hayhurst, G.P., et al., *Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis.* Mol Cell Biol, 2001. **21**(4): p. 1393-403.

378. Bugianesi, E., et al., *Insulin resistance in nonalcoholic fatty liver disease.* Curr Pharm Des, 2010. **16**(17): p. 1941-51.

379. Cobbina, E. and F. Akhlaghi, *Non-alcoholic fatty liver disease (NAFLD) - pathogenesis, classification, and effect on drug metabolizing enzymes and transporters.* Drug Metab Rev, 2017: p. 1-15.

380. Duparc, T., et al., *Hepatocyte MyD88 affects bile acids, gut microbiota and metabolome contributing to regulate glucose and lipid metabolism.* Gut, 2017. **66**(4): p. 620-632.

381. Sartor, R.B., *Microbial influences in inflammatory bowel diseases.* Gastroenterology, 2008. **134**(2): p. 577-94.

382. Matijasic, M., et al., *Modulating Composition and Metabolic Activity of the Gut Microbiota in IBD Patients.* Int J Mol Sci, 2016. **17**(4).

383. Ibrahim, A., et al., *Anti-inflammatory and anti-angiogenic effect of long chain n-3 polyunsaturated fatty acids in intestinal microvascular endothelium.* Clin Nutr, 2011. **30**(5): p. 678-87.

384. Marion-Letellier, R., et al., *Polyunsaturated fatty acids in inflammatory bowel diseases: a reappraisal of effects and therapeutic approaches.* Inflamm Bowel Dis, 2013. **19**(3): p. 650-61.

385. Wang, L., et al., *CPAG: software for leveraging pleiotropy in GWAS to reveal similarity between human traits links plasma fatty acids and intestinal inflammation.* Genome Biol, 2015. **16**: p. 190.

386. MacFie, T.S., et al., *DUOX2 and DUOXA2 form the predominant enzyme system capable of producing the reactive oxygen species H2O2 in active ulcerative colitis and are modulated by 5-aminosalicylic acid.* Inflamm Bowel Dis, 2014. **20**(3): p. 514-24.

387. Kuhn, R., et al., *Interleukin-10-deficient mice develop chronic enterocolitis.* Cell, 1993. **75**(2): p. 263-74.

388. Jacob, F., *Evolution and tinkering.* Science, 1977. **196**(4295): p. 1161-6.

389. Su, D., et al., *Vitamin D Signaling through Induction of Paneth Cell Defensins Maintains Gut Microbiota and Improves Metabolic Disorders and Hepatic Steatosis in Animal Models.* Front Physiol, 2016. **7**: p. 498.

390. Bhattacharya, N., et al., *Normalizing Microbiota-Induced Retinoic Acid Deficiency Stimulates Protective CD8(+) T Cell-Mediated Immunity in Colorectal Cancer.* Immunity, 2016. **45**(3): p. 641-55.

391. Chen, W.S., et al., *Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos.* Genes Dev, 1994. **8**(20): p. 2466-77.

392. Duncan, S.A., et al., *Expression of transcription factor HNF-4 in the extraembryonic endoderm, gut, and nephrogenic tissue of the developing mouse embryo: HNF-4 is a marker for primary endoderm in the implanting blastocyst.* Proc Natl Acad Sci U S A, 1994. **91**(16): p. 7598-602.

393. Sucov, H.M., et al., *RXR alpha mutant mice establish a genetic basis for vitamin A signaling in heart morphogenesis.* Genes Dev, 1994. **8**(9): p. 1007-18.

394. Barak, Y., et al., *PPAR gamma is required for placental, cardiac, and adipose tissue development.* Mol Cell, 1999. **4**(4): p. 585-95.

395. Kubota, N., et al., *PPAR gamma mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance.* Mol Cell, 1999. **4**(4): p. 597-609.

396. Frazier, T.H., J.K. DiBaise, and C.J. McClain, *Gut microbiota, intestinal permeability, obesity-induced inflammation, and liver injury.* JPEN J Parenter Enteral Nutr, 2011. **35**(5 Suppl): p. 14S-20S.

397. Cani, P.D., et al., *Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice.* Diabetes, 2008. **57**(6): p. 1470-81.

398. Wong, A.C., J.M. Chaston, and A.E. Douglas, *The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis.* ISME J, 2013. **7**(10): p. 1922-32.

399. Shin, S.C., et al., *Drosophila Microbiome Modulates Host Developmental and Metabolic Homeostasis via Insulin Signaling.* Science, 2011. **334**(6056): p. 670-674.

400. Wong, A.C.-N., A.J. Dobson, and A.E. Douglas, *Gut microbiota dictates the metabolic response of Drosophila to diet.* The Journal of Experimental Biology, 2014. **217**(11): p. 1894-1901.

401. Dutta, D., et al., *Regional Cell-Specific Transcriptome Mapping Reveals Regulatory Complexity in the Adult Drosophila Midgut.* Cell Rep, 2015. **12**(2): p. 346-58.

402. Ablain, J. and L.I. Zon, *Tissue-specific gene targeting using CRISPR/Cas9.* Methods Cell Biol, 2016. **135**: p. 189-202.

403. Melancon, E., et al., *Best practices for germ-free derivation and gnotobiotic zebrafish husbandry.* Methods Cell Biol, 2017. **138**: p. 61-100.

404. Kabiri, Z., et al., *Stroma provides an intestinal stem cell niche in the absence of epithelial Wnts.* Development, 2014. **141**(11): p. 2206-2215.

405. Yissachar, N., et al., *An Intestinal Organ Culture System Uncovers a Role for the Nervous System in Microbe-Immune Crosstalk.* Cell. **168**(6): p. 1135-1148.e12.

406. Lindemans, C.A., et al., *Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration.* Nature, 2015. **528**(7583): p. 560-4.

407. Basak, O., et al., *Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells.* Cell Stem Cell, 2017. **20**(2): p. 177-190 e4.

408. White, M.D., et al., *Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo.* Cell, 2016. **165**(1): p. 75-87.

409. Patwardhan, R.P., et al., *Massively parallel functional dissection of mammalian enhancers in vivo.* Nat Biotech, 2012. **30**(3): p. 265-270.

410. Nelson, C.E., et al., *In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy.* Science, 2016. **351**(6271): p. 403-407.

411. Black, Joshua B., et al., *Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells.* Cell Stem Cell. **19**(3): p. 406-414.

412. Johansson, M.E., et al., *Normalization of Host Intestinal Mucus Layers Requires Long-Term Microbial Colonization.* Cell Host Microbe, 2015. **18**(5): p. 582-92.

413. Stockinger, S., M.W. Hornef, and C. Chassin, *Establishment of intestinal homeostasis during the neonatal period.* Cell Mol Life Sci, 2011. **68**(22): p. 3699-712.

414. Chassin, C., et al., *miR-146a mediates protective innate immune tolerance in the neonate intestine.* Cell Host Microbe, 2010. **8**(4): p. 358-68.