Mechanisms of Stem Cell Maintenance and Cell Differentiation in the Intestinal Epithelium

A dissertation presented

by

Adrianna Katrina San Roman

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Developmental and Regenerative Biology

Harvard University

Cambridge, Massachusetts

December, 2014

Dissertation Advisor: Dr. Ramesh Shivdasani Adrianna Katrina San Roman

Mechanisms of Stem Cell Maintenance and Cell Differentiation in the Intestinal Epithelium

Abstract

Constant regeneration of the intestinal epithelium, a dynamic tissue with vital digestive and barrier functions, depends on proliferation of resident stem cells and their differentiation into mature cell types. This epithelium thus provides an ideal model to study stem cells and mechanisms of cell differentiation in an adult tissue.

The identification of a proliferative population of intestinal stem cells (ISCs) at the base of intestinal crypts presents the prospect of understanding their regulation by extrinsic and intrinsic factors. Although activation of Wnt signaling in ISCs is thought to be one crucial function of the ISC niche, the cellular source of Wnt ligands is uncertain. Chapter 2 addresses this question through genetic elimination of Wnt ligand secretion in candidate niche cell populations. The data reveal that Wnts originating in any of the sources considered in the literature – the epithelium (including Paneth cells) and sub-epithelial myofibroblasts – are not required for ISC function. These data support models of highly complex cell redundancy or alternative, non-Wnt ligands. Chapter 3 investigates the cell-intrinsic contributions of an intestine-restricted transcription factor (TF), CDX2, to important ISC behaviors. *Cdx2* loss in vivo perturbs ISC proliferation and differentiation, distinct from its functions in mature

enterocytes. Analysis of candidate direct CDX2 target genes in ISCs suggests that CDX2 modulates Fibroblast Growth Factor signaling, thus opening new avenues of investigation.

Although cells that differentiate from ISCs rely on gene expression changes mediated by TFs, loss of several individual TFs in vivo has modest effects on intestinal function. Chapter 4 characterizes the functional interactions of CDX2, which has many properties of a master regulator, with two other intestinal TFs: GATA4 and Hepatocyte Nuclear Factor 4 alpha (HNF4A). Analysis of compound mutant mouse intestines elucidated combinatorial roles for CDX2 with GATA4 in crypt cell proliferation and with HNF4A in enterocyte differentiation. Building on this foundation, Chapter 5 describes preliminary investigations into another TF, HNF1A, in intestinal gene regulation and its relationship with CDX2 in controlling cell differentiation and tissue architecture. These studies highlight the complexities of TF interactions and the functions of diverse TF complexes in controlling tissue-specific genes during cell differentiation.

Table of Contents

Acknowledgements

Although it is my name on this thesis, there are many people I would like to thank for helping me tackle the challenges of graduate school and for contributing to my accomplishments.

First, I would like to thank my graduate advisor, Ramesh Shivdasani for being an excellent scientific mentor, for supporting me in projects I was interested in pursuing and for sharing so much of his time. Thanks to all of the members of the Shivdasani lab past and present that helped make all my days in lab brighter. I especially thank those who were in the lab at the beginning as they taught me many techniques, woke me up early for breakfast and became great friends, including Mike Verzi, Tae-Hee Kim, Juliet Phillips, Rita Sulahian, Annouck Luyten, Justina Chen, and Paloma Cejas. Mike deserves particular credit for taking me on as a mentee during my rotation and mentoring me long after I joined, even after moving to Rutgers. Thanks to Janghee Woo, my fellow graduate student for paving the way for me in the lab. A big thanks also goes to the people in the lab more recently, who I have had the pleasure work with: Isabel Ferriero, Chenu Jayewickreme, Jing Liu, Alessia Cavazza, Unmesh Jhadav, Maddie Saxena, Kazu Murata, Nick O'Neill, and Xiao Han.

There are many others that I have to thank for my scientific development. My dissertation advisory committee, made up of Andrew Lassar, Stuart Orkin, and David Cohen, gave me much constructive advice and helped me to think about my research in new ways. Andrew deserves a special thank-you for being a wonderful committee chair and also for his unfailing support of all DRB students. Thanks to my dissertation defense committee members Andrew Lassar, Alan Cantor, Wolfram Goessling and Andrew Leiter for taking the time to read this thesis. Thanks to Boaz Aronson and Steve Krasinski who were collaborators on my first project in the lab and participated weekly in our joint lab meetings. I would like to thank the

DRB program, especially Spyros Artavanis-Tsakonas, Amy Wagers and Abha Ahuja for helping to create an inclusive student community that I hope will keep expanding over time. Thanks to my fellow DRB students for planning events,

I was very lucky during graduate school to have a solid support system to get me through the difficult times. I could always count on Kate, Maria and Danny in the BBS office to listen to me vent, give me advice and let me play with Tucker! Much of my support came from my classmates who went through this process along with me. Thank you especially to Leah Liu, Leah Silverstein, Ryan Lee, Peter Wang, Ben Morris, Diane Shao, Lauren Barclay, Natalie German, Amy Emerman, Kelli Carroll, Ryan Kuzmickas and Matt Owen for keeping lunch scheduled since 2009. Finally, thanks to Amy Rohlfing and Clare Malone who are amazing classmates, roommates, women and friends.

There are many mentors and friends from my high school and undergraduate years who influenced me to pursue my PhD. Thank you to my teachers in Port Jefferson, NY who sparked my interest in science and pushed me to be a hard worker. While at Williams College, my wonderful thesis advisor, Amy Gehring, was an inspiration and I thank her for her excellent teaching, mentoring and thoughtfulness. Thanks to my mentors Dave Jackson and Yoselin Benitez-Alfonso at Cold Spring Harbor Labs, and Gary Firestone and Crystal Marconett at UC Berkeley for exposing me to new areas of research and encouraging me to pursue graduate school. Thank you my friends from Williams: Emily Olsen, Joanna Hoffman, Quinlan Sievers and Lisa Chu, who have supported me for many years. A special thanks to Joanna for being a wonderful roommate, friend, and for putting up with so much chatter about Western blots. Thanks to my all my friends from home, especially Anique Schachner and Bennett Kramer who have been great supporters since we were in diapers.

James Harris deserves a very special thank-you for being a supportive partner throughout graduate school by reading drafts, listening to presentations and brainstorming with me. James, his family and many friends greatly enriched my time out of lab, and I am so grateful for all of the wonderful experiences we have had together.

Last, but certainly not least I would like to thank my large family and close family friends. Thank you for believing in me since I was small. Thank you to my sisters Ali and Amanda, whose support I value immensely, and whom I look up to for the way they handle difficult situations seemingly effortlessly. Finally, thanks to my Mom, Donna, and Dad, Gerry, for the life they created for me. I am so fortunate to have had all the opportunities that they sacrificed to make happen. They are both excellent mentors and I look up to them so much for their many accomplishments. Thanks for the support, encouragement and love.

Chapter 1: Introduction

The development and structure of the intestinal epithelium

The adult intestinal epithelium undergoes constant renewal during normal homeostasis, characterized by rapid cell proliferation, migration and differentiation (Cheng and Leblond, 1974a, b, c). These dynamic processes occur within the crypts of Lieberkühn and villi, the functional units of the small intestine. Morphologically, the crypts form tubular invaginations of the epithelium into the surrounding mesenchyme, while the villi form finger-like projections into the intestinal lumen (**Figure 1.1**). The crypt contains a proliferating intestinal stem cell (ISC) population, quiescent ISCs and transit-amplifying cells, a population of highly-proliferative intermediate progenitors (Cheng and Leblond, 1974a). As the transit-amplifying cells divide, their progeny undergo migration and terminal differentiation into the four mature cell types found within the intestine, including the secretory goblet and Paneth cells, hormone-producing enteroendocrine cells and absorptive enterocytes (**Figure 1.1**). While Paneth cells migrate down into the base of the crypts, goblet cells, enteroendocrine cells and enterocytes migrate up the villus and are shed into the intestinal lumen upon reaching the tip. A complete cycle takes 3-5 days (Potten and Loeffler, 1990), making the intestine one of the most rapidly cycling tissues in the body. This rapid tissue replacement ensures that the physiological functions of the intestinal epithelium, including absorption of water and nutrients, are maintained while avoiding damage from substances that pass through the intestine.

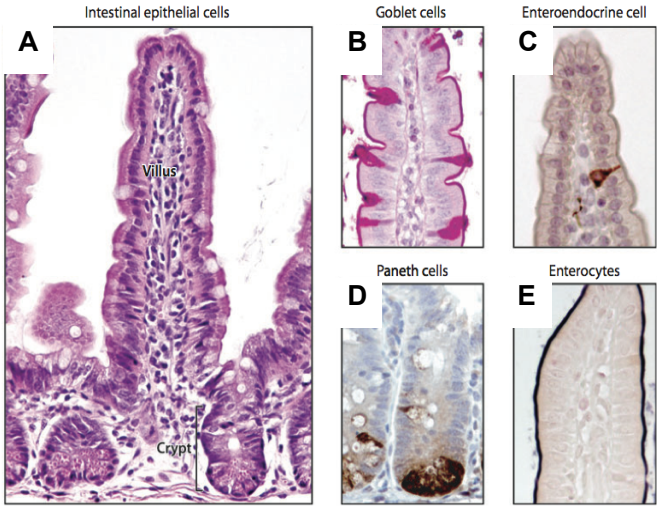


Figure 1.1 Cellular architecture of crypt-villus units of the small intestine. (A) H&E of mouse small intestine. (B-D) Four main differentiated cell types identified by various histochemical and immunohistochemical stains: (B) goblet (periodic acid-Schiff), (C) enteroendocrine (α-synaptophysin), (D) Paneth (α-lysozyme), and (E) enterocytes (alkaline phosphatase). Adapted from (van der Flier and Clevers, 2009).

The development of the murine gastrointestinal tract begins with the formation of the definitive endoderm during gastrulation (embryonic day (E) 6.5-7.5; **Figure 1.2A**) and is completed several weeks after birth (Arnold and Robertson, 2009; Wells and Melton, 1999; Zorn and Wells, 2009). During morphogenesis of the gut tube, it is regionalized on the anterior-posterior axis into foregut, midgut and hindgut, largely through epithelial-mesenchymal interactions (Zorn and Wells, 2009). From the foregut, the tissues of the esophagus, lung, liver, pancreas and stomach arise, while the midgut and hindgut develop into the small intestine and colon, respectively (E9.5-10.5; **Figure 1.2A**). At the foregut-midgut boundary, the homeodomain

transcription factor (TF) CDX2, a master regulator of the intestinal transcriptional program, plays a role in initially specifying the intestine (Gao et al., 2009).

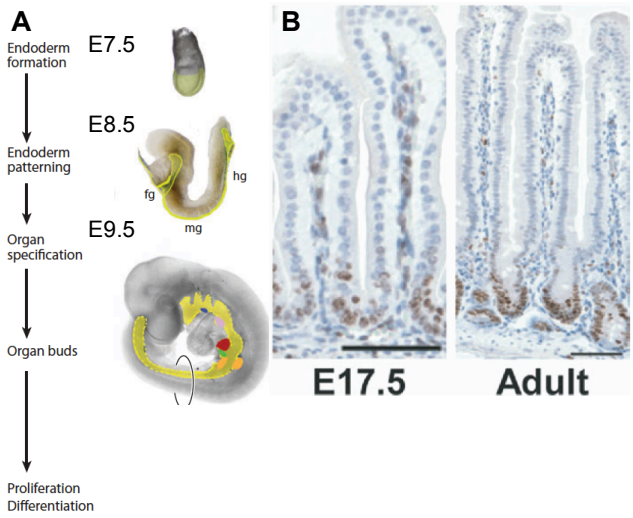


Figure 1.2. Development of the intestinal epithelium. (A) Stages of mouse endoderm (yellow) development. Adapted from (Zorn and Wells, 2009). (B) Villi form embryonically, followed by formation of crypts postnatally. At E17.5, proliferation (visualized by Ki67 immunostaining) is restricted to the intervillus space, while in the adult, proliferation occurs in the crypts. Adapted from (Kim et al., 2007).

After intestinal specification, proliferation combined with convergent extension and radial cell intercalation, contribute to gut tube elongation. These morphogenetic movements are dependent on the planar cell polarity pathway and continue well into neonatal life (Cervantes et al., 2009; Yamada et al., 2010). Between E9.5 and E14, the intestinal lining transitions from a simple cuboidal to a pseudo-stratified epithelium; a process relying on cellular actin-myosin and microtubule networks (Grosse et al., 2011), which is followed by a second transition to a columnar epithelium beginning in the anterior region and progresses as a wave posteriorly (from

E14-16). Villi begin to form during the second transition, which is influenced by the layers of tissue beneath the epithelium. The stiffness of the smooth muscle and condensation of hedgehog and platelet-derived growth factor (PDGF) responsive cells imparts mechanical and inductive cues, respectively (Karlsson et al., 2000; Shyer et al., 2013; Walton et al., 2012). Following villus formation, epithelial proliferation becomes restricted to the intervillus space, which will form the crypts of Lieberkühn postnatally (**Figure 1.2B**). Crypt morphogenesis has not been well-studied, however it is thought invasion of the mesenchyme by intervillus cells and the upward motion of the crypt-villus junction may play a role (Calvert and Pothier, 1990). To keep up with the demand of the elongating gut, and in response to injury, new crypts form by fission of single crypts that bud at the base and bifurcate into two new crypts, likely due to biomechanical strain of an excess of cells in the crypt (Clarke, 1972; Totafurno et al., 1987). Consistent with this mechanism, clonal patches of crypt/villus units can be observed in the epithelium (Cairnie and Millen, 1975; Cummins et al., 2008; Greaves et al., 2006; Gutierrez-Gonzalez et al., 2009; Kim and Shibata, 2004; Li et al., 1994; McDonald et al., 2006). The mature intestinal epithelium is achieved at the suckling-to-weaning transition, approximately 3-4 weeks after birth in the mouse.

The intestinal stem cell and its niche

To support the incredible regenerative capacities of the intestine, a robust stem cell pool must be maintained that is able to self-renew and differentiate into the four major cell types of the intestine. One such population of ISCs occupies the spaces between Paneth cells in the crypt, making up a group of cells historically termed the crypt base columnar cells (CBCs; **Figure 1.3A**; (Cheng and Leblond, 1974a)). The *leucine-rich repeat G-protein coupled receptor 5*

(*Lgr5*), a Wnt target gene and receptor for the Wnt-agonist R-spondin, marks this population (Barker et al., 2007; Carmon et al., 2011; de Lau et al., 2011; Glinka et al., 2011). To confirm *Lgr5* as an intestinal stem cell marker, Barker *et al*. generated a knock-in mouse strain (*Lgr5-EGFP-IRES-creERT2*) that expresses enhanced green fluorescent protein (EGFP) in the crypt base columnar cells (**Figures 1.3B,C**; (Barker et al., 2007)). In addition, lineage tracing and knockout of floxed alleles is enabled by the addition of an internal ribosomal entry site (*IRES*) followed by a tamoxifen-inducible Cre recombinase (Cre^{ERT2}; **Figure 1.3B**). By crossing to the *Rosa26-lox-STOP-lox-LacZ* Cre-recombinase reporter mouse ($R26R^{LacZ}$), it was shown that all mature cell types in the intestine could be formed from the $Lgr5^{+}$ population (**Figures 1.3D-F**). The evidence for *Lgr5* as a stem cell marker was further supported by an *in vitro* culture experiment in which entire crypt-villus organoids were formed from single $Lgr5^{+}$ stem cells, which can be passaged over many months (**Figures 1.3 G-I**; (Sato et al., 2009)).

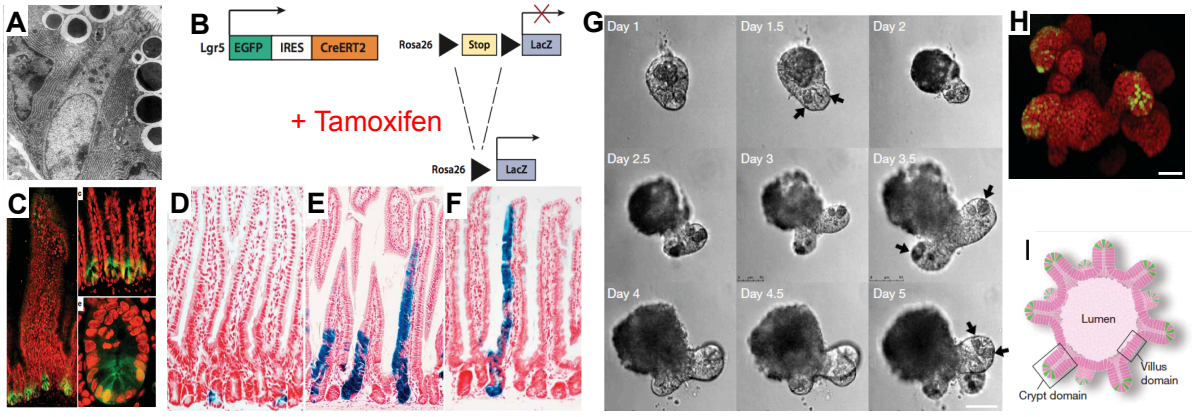


Figure 1.3. *Lgr5* is a marker of a population of ISCs residing at the crypt base. (A) Electron microscopy of CBCs sandwiched between Paneth cells. Adapted from (Cheng and Leblond, 1974a). **(B)** *Lgr5-EGFP-Cre*^{ERT2} knock-in allele and the *R26R*^{LacZ} reporter lineage tracing strategy. **(C)** *Lgr5*-driven GFP expression in the crypt base of small intestine marks CBCs. **(B-C)** Adapted from (Barker et al., 2007). **(D-F)** Lineage tracing using the *R26R*^{LacZ} allele 1 day **(D)**, 5 days **(E)**, and 60 days **(F)** after tamoxifen-induced recombination. Adapted from (Barker et al., 2007). **(G)** Formation of intestinal organoids *in vitro* from isolated intestinal crypts. **(H)** *Lgr5*-driven GFP expression is observed at the base of crypts protruding from organoids. **(I)** Schematic illustrating structure of intestinal organoids in which crypts containing *Lgr5*$^{+}$ ISCs protrude from the organoids, while differentiated cells of the villi are found inside the organoid surrounding a central lumen, into which cells are shed **(G-I)** Adapted from (Sato et al., 2009).

Subsequently, several groups published markers identifying other pools of ISCs that reside in the "+4" position from the crypt base and are quiescent. Interestingly, this was a population historically termed "label retaining cells" and postulated to be the intestinal stem cells (Potten et al., 1978). These markers include *Bmi1, Hopx, mTert* and *Lrig1* (**Figure 1.4**; (Montgomery et al., 2011; Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011)). Experts now postulate that both proliferative ($Lgr5^+$) and quiescent ISC populations exist to provide the constant epithelial renewal, and to respond to epithelial injury, respectively (Li and Clevers, 2010). In addition, it has been shown that these cells are interconvertible: $Bmi1^+$, $mTert^+$ and $Hopx^+$ cells can give rise to $Lgr5^+$ ISCs and vice versa (Montgomery et al., 2011; Takeda et al., 2011; Tian et al., 2011; Yan et al., 2012). Despite the interest in other ISC populations, $Lgr5^+$ CBCs constitute a "workhorse" population of ISCs that makes the principal contribution toward ongoing villus renewal, have been rigorously identified, characterized and can be isolated as a pure population. In this thesis, I will consider $Lgr5^+$ CBCs to represent a *bona fide* ISC population.

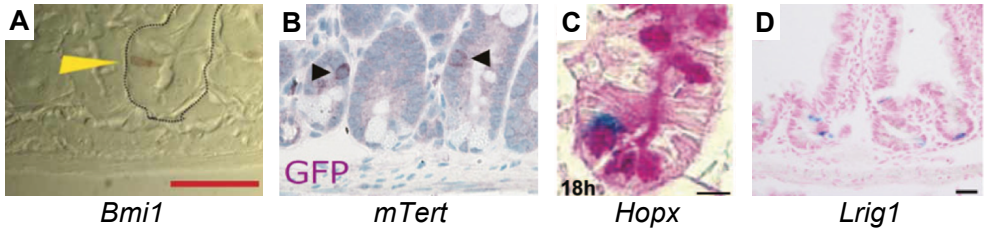


Figure 1.4. Markers of ISC populations at the "+4" position. (A,C,D) *Bmi1-Cre*^{ERT2} (A), *Hopx-Cre*^{ERT2} (C) or *Lrig1-Cre*^{ERT2} (D) mice were crossed with the *Rosa26*^{LacZ} reporter, injected with tamoxifen and LacZ staining was observed after 20, 18 and 24 hours, respectively, resulting in labeled cells mostly in the +4-5 position. Of note, *Lrig1*-marked cells were also found in the CBC zone as well as higher in the crypt. Adapted from (Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011). (B) Immunohistochemistry for *mTert*-driven GFP shows that these cells are highest in the +5-7 position. Adapted from (Montgomery et al., 2011).

The discovery of tools to identify and isolate *Lgr5*$^{+}$ ISCs has made it possible to investigate how the critical functions of this cell population are controlled, both extrinsically and intrinsically. In terms of extrinsic signals, attention has been focused on the local microenvironment surrounding the ISCs, termed the "niche." The ISC niche is the region at the base of the crypt and is comprised of many different cell types found in the crypt base epithelium and in the surrounding tissues, including Paneth cells, sub-epithelial myofibroblasts, non-muscle fibroblasts, endothelial cells, neurons, muscle cells and blood cells (Shaker and Rubin, 2010). Several signaling pathways activated by factors within the niche participate in ISC regulation including Wnt, bone morphogenetic protein (BMP), Notch and hedgehog (Hh) (Walker et al., 2009; Yen and Wright, 2006). The Wnt pathway has been most thoroughly studied due to activating mutations in the pathway found consistently in human colon cancer, most of which disrupt the *Apc* gene or stabilize β-catenin (Clarke, 2006; Gregorieff and Clevers, 2005; Harada

et al., 1999). Experiments that abrogate Wnt signaling through expression of the antagonist *Dickkopf-1* (*Dkk1*) or deletion of the effector *Tcf4* revealed loss of proliferation and marked epithelial defects (Korinek et al., 1998; Kuhnert et al., 2004; Pinto et al., 2003; van Es et al., 2012). In culture, isolated *Lgr5*$^{+}$ ISCs only proliferate and form intestinal organoids in the presence of the Wnt agonist R-spondin (Sato et al., 2009). Thus, two main questions regarding the ISC niche are which cell(s) secretes Wnt ligands and which Wnt ligand(s) is important.

There are two likely cell types that have been hotly debated in the field: the Paneth cells, which reside directly adjacent to the ISCs in the base of the crypt, and the intestinal subepithelial myofibroblasts, which sit below the crypts. Several papers have addressed this question, although the results are conflicting. Four studies ablated the Paneth cells in different ways: using a Paneth cell specific diphtheria toxin, *Gfi1*$^{-/-}$ mice, and two using *Sox9* conditional knockout mice and all showed no effect in the homeostasis of the intestine, resulting in the conclusion that Paneth cells must not be required in the ISC niche (Bastide et al., 2007; Garabedian et al., 1997; Mori-Akiyama et al., 2007; Shroyer et al., 2005). However, these studies were called into question due to two main observations. First, it was shown that Paneth cells are not completely absent in the diphtheria toxin model and *Gfi1*$^{-/-}$ mice. Second, by examining the *Sox9* conditional knockout crypts at early time-points, it was proposed that these crypts without Paneth cells are at an initial disadvantage, however crypts that escape Cre-mediated *Sox9* recombination replenish the epithelium with wild-type cells, making the phenotype look normal at the later time points initially reported (Sato et al., 2011). Additional arguments based upon *in vitro* experiments in which Paneth cells co-cultured with ISCs form better organoids than ISCs alone led to the conclusion that Paneth cells actually do constitute the ISC niche (Sato et al., 2011). Subsequently, two papers using *Atoh1* knockout mice, which genetically eliminate all secretory

cells including Paneth cells, showed no reduced ISC function *in vivo* (Durand et al., 2012; Kim et al., 2012). Most recently, it was shown that Wnt3 expression is required for *in vitro* culture of organoids, but not to maintain the epithelium *in vivo* and the requirement in organoids can be bypassed by co-culture with mesenchymal cells, suggesting that redundant sources of Wnt within the ISC niche (Farin et al., 2012). The cellular source of intestinal Wnt ligands within the ISC niche is investigated in **Chapter 2**.

In addition to extrinsic factors provided by the ISC niche, intrinsic factors also play an important role. The precise control of an ISC specific gene expression program is crucial for controlling cell functions and to respond to the extrinsic signals. Unsurprisingly many of the factors intrinsic to ISCs that seem to play a functional role are transcription factors (TFs). The first identified factor was the Wnt-responsive TF achaete scute-like 2 (ASCL2), the loss of which results in the elimination of the *Lgr5*$^{+}$ ISC population, and the deregulation of Wnt-responsive genes (van der Flier et al., 2009). Subsequently, several other TFs have been identified in controlling ISC functions during normal homeostasis, including KLF5, VDR and YY1 (Bell and Shroyer, 2014; Nakaya et al., 2014; Peregrina et al., 2014; Perekatt et al., 2014), and in response to injury, including ID1 (Zhang et al., 2014). **Chapter 3** will describe the requirement of another TF, CDX2, in controlling gene expression and ISC functions.

Epigenetic regulation of the intestinal epithelium

Cell differentiation in the intestinal epithelium requires the execution of lineage-specific gene expression programs, which involves changes in the expression levels of hundreds or thousands of genes. These programs are mediated through alterations in the chromatin structure, which can be stably inherited through differentiation and through tissue or lineage specific transcription

factor binding (Struhl, 1999). The activation of gene expression in eukaryotes is largely accomplished at distal cis-regulatory elements (CREs), known as enhancers, at which clusters of transcription factor binding sites are found. The nucleoprotein complexes formed at these enhancer regions are termed the "enhanceosome" and there has been much recent interest in understanding how this controls gene expression (Carey, 1998).

While we know that assembly of enhanceosomes is critical for proper gene regulation, it is challenging to find the locations of CREs and to determine what genes they regulate, as they can be found at large distances from the transcription start site, within introns and in any orientation (Visel et al., 2009). Initial attempts used a comparative genomic approach to mine conserved non-coding regions for CREs (Pennacchio et al., 2006; Visel et al., 2008). While this yielded some positional information, it did not provide any indication to the temporal or cell-type specific activity of these CREs (Pennacchio and Visel, 2010). Subsequently, more advanced approaches were developed to take into account general features of enhanceosomes, which have generally been more successful (Ong and Corces, 2011; Visel et al., 2009). Since many proteins need to gain access to the DNA in CREs, enhanceosomes tend to be relatively nucleosome free, a general property, which new techniques have successfully exploited. Some labs have used DNaseI hypersensitivity and other assays to try and determine regions of nucleosome free DNA (Crawford et al., 2006). Others have investigated the presence of unstable histone variants H3.3 and H2A.Z (Jin and Felsenfeld, 2007; Jin et al., 2009), while still others have looked at the presence of ATP-dependent chromatin remodelers which loosen the DNA from contact with histone proteins (Rada-Iglesias et al., 2011). Another useful strategy for finding CREs is looking for the specific combination of histone modifications that generally correlates with active chromatin or assessing the presence of histone-modifying factors (Heintzman et al., 2007).

Studies combining several of these strategies have identified temporal and spatial activity at enhancers during early development, establishing a proof of principle that the identification of enhancers in this way can predict gene expression (Rada-Iglesias et al., 2011).

Much work on the enhanceosome of various cell populations within the intestinal epithelium has been performed in the Shivdasani Lab. To begin to understand how gene expression changes are mediated during differentiation of the intestinal epithelium, the active cis-regulatory elements in progenitor versus differentiated cells were profiled (Verzi et al., 2010b). The Caco2 human colon cancer cell line was used as a model because the cells resemble a progenitor state while proliferating and mature enterocytes, the largest class of differentiated cells in the intestine, post-confluence. Chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-seq) using antibodies for histone marks correlated with active enhancers (H3K4Me2 and H3K27Ac) was used to identify locations of active CREs in each cell population by looking for a pattern of two marked nucleosomes flanking a central unmarked region (**Figure 1.5A**). By analyzing the DNA sequences in the center of active CREs, the motif for CDX2, a homeodomain transcription factor that is exclusively restricted to the epithelium in the adult (Silberg et al., 2000), was found to be present at many sites in both populations of cells. CDX2 was previously known to be required for proper specification of the intestine during development (Gao et al., 2009) and is considered to be a master regulator transcription factor in the intestine because its ectopic expression in the stomach or esophagus leads to the expression of intestinal markers (Liu et al., 2007; Silberg et al., 2002). Other transcription factor binding motifs enriched in active chromatin of proliferating cells were GATA and BACH1, while hepatocyte nuclear factor (HNF) 1 and HNF4A were enriched in differentiated cells (**Figures 1.5B,C)**.

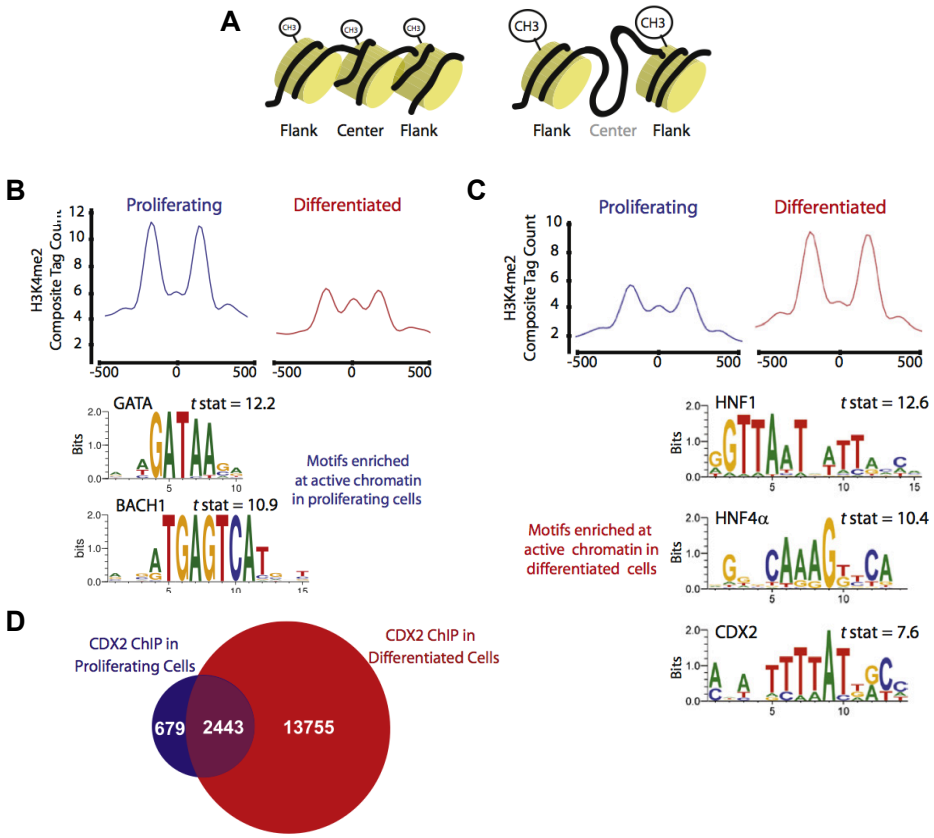


Figure 1.5. Identification of transcription factor motifs within H3K4me2-marked cis-regulatory elements in proliferating and differentiated Caco2 cells. (A) Schematic diagramming inactive chromatin (left) containing three adjacent nucleosomes bearing the H3K4me2 mark. Right, active chromatin is characterized by two well-positioned H3K4me2-marked nucleosomes flanking a central unmarked region in which TF motifs can be found. **(B)** Composite tag counts of H3K4me2 ChIP-seq data, showing sites that are active specifically in proliferating cells. At the center of these sites, the motifs for GATA and BACH1 are enriched. **(C)** Sites that are active specifically in differentiated cells contain the motifs for HNF1A, HNF4A and CDX2. **(D)** Venn diagram comparing ChIP-seq binding sites in proliferating and differentiated Caco2 cells reveals dynamic changes in CDX2 binding during cell differentiation. All figures adapted from (Verzi et al., 2010a).

To better understand the role of CDX2 in controlling gene expression during cell differentiation within the adult intestine, CDX2 binding sites were mapped using ChIP-seq in proliferating and differentiated cells. Interestingly, CDX2 bound to many different sites in each state (**Figure 1.5D**) and motifs for different transcription factors near CDX2 binding sites were differentially enriched in each state. In proliferating cells, the GATA motif was commonly found, while in differentiated cells, the HNF4A motif was found. To ascertain the role of CDX2 *in vivo*, a mouse strain was generated in which CDX2 was conditionally deleted in the intestine using a tamoxifen-inducible Cre-recombinase driven by the *Villin* gene, which is ubiquitously expressed in the intestinal epithelium (Madison et al., 2002). Phenotypic analysis of the knockout mice revealed that CDX2 is required for intestinal cell differentiation in the intestine. Additionally, many of the sites with H3K4me2-marked chromatin in wild-type epithelium lacked these marks upon CDX2 knockout, suggesting that the presence of CDX2 affects chromatin dynamics in the intestinal epithelium. This study provided a foundation for current studies in the Shivdasani Lab aimed at understanding the dynamics of CDX2 binding during cell differentiation *in vivo* (discussed in **Chapter 3**), and the mechanisms controlling these changes. One hypothesis is that the other transcription factors also predicted to bind at CREs in progenitor and differentiated cells with CDX2, including GATA, HNF4A and HNF1, may interact with CDX2 to mediate its functions at these sites. Indeed, **Chapters 4 and 5** present evidence for functional interactions of CDX2 with GATA4 in proliferating cells, and CDX2 with HNF4A and HNF1A in differentiated cells.

Conclusions and Significance

Elucidation of epigenetic regulation of cell differentiation within the intestinal epithelium has broad implications for the fields of genetics, developmental biology and medicine. Any information gained in this area has the potential to influence our understanding of pathologies of the intestine including cancer and maturity onset diabetes of the young, a disease that the TFs HNF4A and HNF1 contribute to when inactive. In addition, understanding the factors that regulate the functions of ISCs will facilitate their culture for cell-based regenerative therapies, which could be applied to patients suffering from congenital short bowel syndrome, inflammatory bowel disease or radiation damage (Shaker and Rubin, 2010). Furthermore, the knowledge of how transcription factors interact at enhanceosomes to regulate gene expression may result in new paradigms for understanding fundamental mechanisms behind the execution of genetic programs.

References

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nature reviews Molecular cell biology *10*, 91-103.

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

Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., Bibeau, F., Scherer, G., Joubert, D., Hollande, F.*, et al.* (2007). Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol *178*, 635-648.

Bell, K.N., and Shroyer, N.F. (2014). Krupple-Like Factor 5 Is Required for Proper Maintenance of Adult Intestinal Crypt Cellular Proliferation. Dig Dis Sci.

Cairnie, A.B., and Millen, B.H. (1975). Fission of crypts in the small intestine of the irradiated mouse. Cell Tissue Kinet *8*, 189-196.

Calvert, R., and Pothier, P. (1990). Migration of fetal intestinal intervillous cells in neonatal mice. The Anatomical record *227*, 199-206.

Carey, M. (1998). The enhanceosome and transcriptional synergy. Cell *92*, 5-8.

Carmon, K.S., Gong, X., Lin, Q., Thomas, A., and Liu, Q. (2011). R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling. Proc Natl Acad Sci U S A *108*, 11452-11457.

Cervantes, S., Yamaguchi, T.P., and Hebrok, M. (2009). Wnt5a is essential for intestinal elongation in mice. Dev Biol *326*, 285-294.

Cheng, H., and Leblond, C.P. (1974a). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. I. Columnar cell. The American journal of anatomy *141*, 461-479.

Cheng, H., and Leblond, C.P. (1974b). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. III. Entero-endocrine cells. The American journal of anatomy *141*, 503-519.

Cheng, H., and Leblond, C.P. (1974c). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. Am J Anat *141*, 537-561.

Clarke, A.R. (2006). Wnt signalling in the mouse intestine. Oncogene *25*, 7512-7521.

Clarke, R.M. (1972). The effect of growth and of fasting on the number of villi and crypts in the small intestine of the albino rat. J Anat *112*, 27-33.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D*., et al.* (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res *16*, 123-131.

Cummins, A.G., Catto-Smith, A.G., Cameron, D.J., Couper, R.T., Davidson, G.P., Day, A.S., Hammond, P.D., Moore, D.J., and Thompson, F.M. (2008). Crypt fission peaks early during

infancy and crypt hyperplasia broadly peaks during infancy and childhood in the small intestine of humans. J Pediatr Gastroenterol Nutr *47*, 153-157.

de Lau, W., Barker, N., Low, T.Y., Koo, B.-K., Li, V.S.W., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M.*, et al.* (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature *476*, 293-297.

Durand, A., Donahue, B., Peignon, G., Letourneur, F., Cagnard, N., Slomianny, C., Perret, C., Shroyer, N.F., and Romagnolo, B. (2012). Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). Proc Natl Acad Sci U S A *109*, 8965-8970.

Farin, H.F., Van Es, J.H., and Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. Gastroenterology *143*, 1518-1529 e1517.

Gao, N., White, P., and Kaestner, K.H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev Cell *16*, 588-599.

Garabedian, E.M., Roberts, L.J., McNevin, M.S., and Gordon, J.I. (1997). Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. J Biol Chem *272*, 23729-23740.

Glinka, A., Dolde, C., Kirsch, N., Huang, Y.L., Kazanskaya, O., Ingelfinger, D., Boutros, M., Cruciat, C.M., and Niehrs, C. (2011). LGR4 and LGR5 are R-spondin receptors mediating Wnt/beta-catenin and Wnt/PCP signalling. EMBO reports *12*, 1055-1061.

Greaves, L.C., Preston, S.L., Tadrous, P.J., Taylor, R.W., Barron, M.J., Oukrif, D., Leedham, S.J., Deheragoda, M., Sasieni, P., Novelli, M.R.*, et al.* (2006). Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. Proc Natl Acad Sci U S A *103*, 714-719.

Gregorieff, A., and Clevers, H. (2005). Wnt signaling in the intestinal epithelium: from endoderm to cancer. Genes & development *19*, 877-890.

Grosse, A.S., Pressprich, M.F., Curley, L.B., Hamilton, K.L., Margolis, B., Hildebrand, J.D., and Gumucio, D.L. (2011). Cell dynamics in fetal intestinal epithelium: implications for intestinal growth and morphogenesis. Development *138*, 4423-4432.

Gutierrez-Gonzalez, L., Deheragoda, M., Elia, G., Leedham, S.J., Shankar, A., Imber, C., Jankowski, J.A., Turnbull, D.M., Novelli, M., Wright, N.A., *et al.* (2009). Analysis of the clonal architecture of the human small intestinal epithelium establishes a common stem cell for all lineages and reveals a mechanism for the fixation and spread of mutations. The Journal of pathology *217*, 489-496.

Harada, N., Tamai, Y., Ishikawa, T., Sauer, B., Takaku, K., Oshima, M., and Taketo, M.M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. The EMBO journal *18*, 5931-5942.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Karlsson, L., Lindahl, P., Heath, J.K., and Betsholtz, C. (2000). Abnormal gastrointestinal development in PDGF-A and PDGFR-(alpha) deficient mice implicates a novel mesenchymal structure with putative instructive properties in villus morphogenesis. Development (Cambridge, England) *127*, 3457-3466.

Kim, B.M., Mao, J., Taketo, M.M., and Shivdasani, R.A. (2007). Phases of canonical Wnt signaling during the development of mouse intestinal epithelium. Gastroenterology *133*, 529-538.

Kim, K.-M., and Shibata, D. (2004). Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. BMC Gastroenterol *4*, 8.

Kim, T.-H., Escudero, S., and Shivdasani, R.A. (2012). Intact function of Lgr5 receptor-expressing intestinal stem cells in the absence of Paneth cells. Proc Natl Acad Sci U S A *109*, 3932-3937.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P.J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet *19*, 379-383.

Kuhnert, F., Davis, C.R., Wang, H.-T., Chu, P., Lee, M., Yuan, J., Nusse, R., and Kuo, C.J. (2004). Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A *101*, 266-271.

Li, L., and Clevers, H. (2010). Coexistence of quiescent and active adult stem cells in mammals. Science (New York, NY) *327*, 542-545.

Li, Y.Q., Roberts, S.A., Paulus, U., Loeffler, M., and Potten, C.S. (1994). The crypt cycle in mouse small intestinal epithelium. J Cell Sci *107 (Pt 12)*, 3271-3279.

Liu, T., Zhang, X., So, C.K., Wang, S., Wang, P., Yan, L., Myers, R., Chen, Z., Patterson, A.P., Yang, C.S.*, et al.* (2007). Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. Carcinogenesis *28*, 488-496.

Madison, B.B., Dunbar, L., Qiao, X.T., Braunstein, K., Braunstein, E., and Gumucio, D.L. (2002). Cis elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. The Journal of biological chemistry *277*, 33275-33283.

McDonald, S.A.C., Preston, S.L., Greaves, L.C., Leedham, S.J., Lovell, M.A., Jankowski, J.A.Z., Turnbull, D.M., and Wright, N.A. (2006). Clonal expansion in the human gut: mitochondrial DNA mutations show us the way. Cell cycle (Georgetown, Tex) *5*, 808-811.

Montgomery, R.K., Carlone, D.L., Richmond, C.A., Farilla, L., Kranendonk, M.E.G., Henderson, D.E., Baffour-Awuah, N.Y., Ambruzs, D.M., Fogli, L.K., Algra, S.*, et al.* (2011). Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. Proc Natl Acad Sci U S A *108*, 179-184.

Mori-Akiyama, Y., van den Born, M., van Es, J.H., Hamilton, S.R., Adams, H.P., Zhang, J., Clevers, H., and de Crombrugghe, B. (2007). SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. YGAST *133*, 539-546.

Nakaya, T., Ogawa, S., Manabe, I., Tanaka, M., Sanada, M., Sato, T., Taketo, M.M., Nakao, K., Clevers, H., Fukayama, M.*, et al.* (2014). KLF5 regulates the integrity and oncogenicity of intestinal stem cells. Cancer Res *74*, 2882-2891.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature reviews Genetics *12*, 283-293.

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D.*, et al.* (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499-502.

Pennacchio, L.A., and Visel, A. (2010). Limits of sequence and functional conservation. Nat Genet *42*, 557-558.

Peregrina, K., Houston, M., Daroqui, C., Dhima, E., Sellers, R.S., and Augenlicht, L.H. (2014). Vitamin D is a determinant of mouse intestinal Lgr5 stem cell functions. Carcinogenesis.

Perekatt, A.O., Valdez, M.J., Davila, M., Hoffman, A., Bonder, E.M., Gao, N., and Verzi, M.P. (2014). YY1 is indispensable for Lgr5+ intestinal stem cell renewal. Proc Natl Acad Sci U S A *111*, 7695-7700.

Pinto, D., Gregorieff, A., Beghtel, H., and Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev *17*, 1709-1713.

Potten, C.S., HUME, W., REID, P., and CAIRNS, J. (1978). SEGREGATION OF DNA IN EPITHELIAL STEM-CELLS. Cell *15*, 899-906.

Potten, C.S., and Loeffler, M. (1990). Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. Development (Cambridge, England) *110*, 1001-1020.

Powell, A.E., Wang, Y., Li, Y., Poulin, E.J., Means, A.L., Washington, M.K., Higginbotham, J.N., Juchheim, A., Prasad, N., Levy, S.E.*, et al.* (2012). The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. Cell *149*, 146-158.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet *40*, 915-920.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415-418.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., *et al.* (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262-265.

Shaker, A., and Rubin, D.C. (2010). Intestinal stem cells and epithelial-mesenchymal interactions in the crypt and stem cell niche. Translational research : the journal of laboratory and clinical medicine *156*, 180-187.

Shroyer, N.F., Wallis, D., Venken, K.J.T., Bellen, H.J., and Zoghbi, H.Y. (2005). Gfi1 functions downstream of Math1 to control intestinal secretory cell subtype allocation and differentiation. Genes Dev *19*, 2412-2417.

Shyer, A.E., Tallinen, T., Nerurkar, N.L., Wei, Z., Gil, E.S., Kaplan, D.L., Tabin, C.J., and Mahadevan, L. (2013). Villification: how the gut gets its villi. Science *342*, 212-218.

Silberg, D.G., Sullivan, J., Kang, E., Swain, G.P., Moffett, J., Sund, N.J., Sackett, S.D., and Kaestner, K.H. (2002). Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. Gastroenterology *122*, 689-696.

Silberg, D.G., Swain, G.P., Suh, E.R., and Traber, P.G. (2000). Cdx1 and cdx2 expression during intestinal development. Gastroenterology *119*, 961-971.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1-4.

Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion Between Intestinal Stem Cell Populations in Distinct Niches. Science (New York, NY) *334*, 1420-1424.

Tian, H., Biehs, B., Warming, S., Leong, K.G., Rangell, L., Klein, O.D., and de Sauvage, F.J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature.

Totafurno, J., Bjerknes, M., and Cheng, H. (1987). The crypt cycle. Crypt and villus production in the adult intestinal epithelium. Biophys J *52*, 279-294.

van der Flier, L.G.v.d., and Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annu Rev Physiol *71*, 241-260.

van der Flier, L.G.v.d., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Beghtel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.

van Es, J.H., Haegebarth, A., Kujala, P., Itzkovitz, S., Koo, B.K., Boj, S.F., Korving, J., van den Born, M., van Oudenaarden, A., Robine, S.*, et al.* (2012). A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. Mol Cell Biol *32*, 1918-1927.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010a). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell *19*, 713-726.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010b). Differentiation-Specific Histone Modifications Reveal Dynamic Chromatin Interactions and Partners for the Intestinal Transcription Factor CDX2. Dev Cell *19*, 713-726.

Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., and Pennacchio, L.A. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet *40*, 158-160.

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. Nature *461*, 199-205.

Walker, M.R., Patel, K.K., and Stappenbeck, T.S. (2009). The stem cell niche. The Journal of pathology *217*, 169-180.

Walton, K.D., Kolterud, A., Czerwinski, M.J., Bell, M.J., Prakash, A., Kushwaha, J., Grosse, A.S., Schnell, S., and Gumucio, D.L. (2012). Hedgehog-responsive mesenchymal clusters direct patterning and emergence of intestinal villi. Proc Natl Acad Sci U S A *109*, 15817-15822.

Wells, J.M., and Melton, D.A. (1999). Vertebrate endoderm development. Annu Rev Cell Dev Biol *15*, 393-410.

Yamada, M., Udagawa, J., Matsumoto, A., Hashimoto, R., Hatta, T., Nishita, M., Minami, Y., and Otani, H. (2010). Ror2 is required for midgut elongation during mouse development. Dev Dyn *239*, 941-953.

Yan, K.S., Chia, L.A., Li, X., Ootani, A., Su, J., Lee, J.Y., Su, N., Luo, Y., Heilshorn, S.C., Amieva, M.R.*, et al.* (2012). The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. Proc Natl Acad Sci U S A *109*, 466-471.

Yen, T.-H., and Wright, N.A. (2006). The gastrointestinal tract stem cell niche. Stem cell reviews *2*, 203-212.

Zhang, N., Yantiss, R.K., Nam, H.S., Chin, Y., Zhou, X.K., Scherl, E.J., Bosworth, B.P., Subbaramaiah, K., Dannenberg, A.J., and Benezra, R. (2014). ID1 Is a Functional Marker for Intestinal Stem and Progenitor Cells Required for Normal Response to Injury. Stem cell reports *3*, 716-724.

Zorn, A.M., and Wells, J.M. (2009). Vertebrate endoderm development and organ formation. Annu Rev Cell Dev Biol *25*, 221-251.

Chapter 2: Wnt secretion from epithelial cells and sub-epithelial myofibroblasts is not required in the mouse intestinal stem cell niche *in vivo*

Adrianna K. San Roman, Chenura D. Jayewickreme, L. Charles Murtaugh, and Ramesh A. Shivdasani

This work was published in *Stem Cell Reports (2014) 2:127-134.*

Contributions:

A. K. San Roman and R. A. Shivdasani developed the research plan and wrote the manuscript, with input from C. D. Jayewickreme and L. C. Murtaugh. All experiments and data analysis were performed collaboratively between A. K. San Roman and C. D. Jayewickreme. L. C. Murtaugh provided *Porcn* knockout mice. All work was conducted under the direction of R. A. Shivdasani.

Wnt secretion from epithelial cells and sub-epithelial myofibroblasts is not required in the mouse intestinal stem cell niche *in vivo*

Adrianna K. San Roman[1,2,5], Chenura D. Jayewickreme[1,5], L. Charles Murtaugh[3], and Ramesh A. Shivdasani[1,4,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA; [2]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA 02215, USA; [3]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA; [4]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215, USA; [5]Equal contribution

***Corresponding author:**
Ramesh A. Shivdasani, M.D., Ph.D.
Dana-Farber Cancer Institute
450 Brookline Avenue
Boston, MA 02215, U.S.A.
Tel.: (617) 632-5746 Fax: (617) 582-7198
Email: ramesh_shivdasani@dfci.harvard.edu

Summary

Wnt signaling is a crucial aspect of the intestinal stem cell niche required for crypt cell proliferation and differentiation. Paneth cells or sub-epithelial myofibroblasts are leading candidate sources of the required Wnt ligands, but this has not been tested *in vivo*. To abolish Wnt ligand secretion, we used *Porcupine* (*Porcn*) conditional-null mice crossed to strains expressing inducible Cre recombinase in the epithelium, including Paneth cells (*Villin-Cre*^{ERT2}); in smooth muscle, including sub-epithelial myofibroblasts (*Myh11-Cre*^{ERT2}); and simultaneously in both compartments. Elimination of Wnt secretion from any of these compartments did not disrupt tissue morphology, cell proliferation, differentiation, or Wnt pathway activity. Thus, Wnt ligand secretion from these cell populations is dispensable for intestinal homeostasis, revealing that a minor cell type or significant and unexpected redundancy is responsible for physiologic Wnt signaling *in vivo*.

Introduction

Intestinal crypts house self-renewing stem cells and transit-amplifying progenitors that depend on Wnt signaling. Expression of endogenous pathway antagonists such as *Dickkopf-1* (*Dkk1*) reduces Wnt signaling, arrests stem and progenitor cell proliferation, and impairs secretory cell differentiation (Kuhnert et al., 2004; Pinto et al., 2003). Additionally, although it is unclear exactly when Wnt signaling begins in the developing intestine (Kim et al., 2007), mice lacking *Tcf4*, a transcriptional effector of the Wnt pathway, show marked epithelial defects (Korinek et al., 1998; van Es et al., 2012). Conversely, constitutive Wnt activity drives excessive cell replication and tumors, including human colorectal cancer (Korinek et al., 1997; Morin et al.,

1997). Intestinal stem cells (ISCs) located at the base of mouse small intestine crypts express the Wnt-responsive gene *Lgr5* (Barker et al., 2007), which encodes a co-receptor for Wnt-agonist R-spondins (Carmon et al., 2011; de Lau et al., 2011; Glinka et al., 2011), and are the origin for epithelial tumors spurred by constitutive Wnt activity (Barker et al., 2009). Although Wnt signaling is thus an imperative aspect of intestinal homeostasis, the *bona fide* physiologic Wnt source within the ISC niche *in vivo* is unknown.

Current hypotheses for this source draw on diverse observations. Paneth cells are a favored epithelial source because they reside in intimate contact with $Lgr5^+$ ISCs (Barker et al., 2007; Cheng and Leblond, 1974) and their co-culture with the latter improves mouse organoid formation (Sato et al., 2011). *Wnts 3*, *6*, and *9b* are highly expressed at the crypt base and *Wnt3*, in particular, is present at higher levels in Paneth cells than in ISCs (Gregorieff et al., 2005; Sato et al., 2011). However, partial (Bastide et al., 2007; Garabedian et al., 1997; Mori-Akiyama et al., 2007; Shroyer et al., 2005) or complete, irreversible (Durand et al., 2012; Kim et al., 2012) Paneth cell ablation preserves mouse crypt homeostasis *in vivo*, suggesting extra-epithelial Wnt sources. A likely mesenchymal source is the sub-epithelial myofibroblasts (SEMFs) that envelop intestinal crypts (Powell et al., 2011) and support intestinal epithelial growth *ex vivo* and in tissue xenografts, likely through Wnt secretion (Lahar et al., 2011). SEMFs and other sub-epithelial cells express several *Wnts*, including *Wnt2b*, *Wnt4*, and *Wnt5a* (Gregorieff et al., 2005). Non-SEMF mesenchymal cells, such as non-muscle fibroblasts, endothelial cells, neurons or blood leukocytes, could also provide the required Wnts. Finally, epithelial and sub-epithelial cells may serve as redundant sources, much as renal development and function reveal complex, overlapping sites of Wnt production (Kispert et al., 1998; Park et al., 2007).

We used *Porcupine (Porcn*) conditional knockout mice to identify the source of intestinal Wnts. *Porcn* encodes an O-acyltransferase that palmitoylates all vertebrate Wnts at a conserved serine residue and is necessary for their secretion (Chen et al., 2009; Najdi et al., 2012; Proffitt and Virshup, 2012). *PORCN* mutations cause focal dermal hypoplasia in humans (Grzeschik et al., 2007; Wang et al., 2007) and the tissue defects in *Porcn*-null mice resemble those observed upon loss of single *Wnts*, including *Wnt3, Wnt3a, Wnt5a*, and *Wnt7b* (Barrott et al., 2011; Biechele et al., 2013; Biechele et al., 2011; Liu et al., 2012). Thus, our approach circumvented the problem of redundancy that afflicts studies of single *Wnt* gene disruption. We used *Villin-Cre^{ERT2}* mice (el Marjou et al., 2004) to delete *Porcn* in the intestinal epithelium and *Myh11-Cre^{ERT2}* mice (Wirth et al., 2007) for deletion in sub-epithelial cells. Loss of *Porcn* from all epithelial or smooth muscle (including SEMF) cells, alone or in combination, produced none of the defects in crypt cell proliferation, differentiation, or Wnt target gene expression expected from Wnt deficiency. This rigorous genetic study therefore points to some other minor cell type as a source of Wnt ligands in the mammalian intestine.

Experimental Procedures

Mice

Villin-Cre^{ERT2}, *Myh11-Cre^{ERT2}, Rosa26-lox-STOP-lox-YFP*, and *Porcn$^{Fl/Y}$* mouse strains were described previously (Barrott et al., 2011; el Marjou et al., 2004; Srinivas et al., 2001; Wirth et al., 2007). To induce recombination of conditional alleles, mice were injected intraperitoneally (IP) with 1-2 mg Tamoxifen (TAM, Sigma) dissolved in sunflower oil (Sigma) on 5 consecutive days.

Tissue harvests

Small intestines were dissected and flushed with cold phosphate buffered saline (PBS). The proximal and distal 1/3 were fixed overnight in 4% paraformaldehyde, then washed in PBS and embedded in paraffin or frozen for immunohistochemical analysis. The first 1 cm of the middle 1/3 was reserved for whole intestine (unfractionated) DNA or mRNA analysis. The remainder was used for isolating epithelium by incubating in 5mM EDTA, shaking by hand, and passage through 70 μm filters to separate crypts and villi.

DNA isolation and genotyping

Epithelium or whole intestine was agitated in SNET buffer (0.2% SDS, 0.2 M NaCl, 100 mM Tris pH 8, and 5 mM EDTA) with 15 μg Proteinase K at 55°C for 2 hours (epithelium) or overnight (whole intestine) and DNA was isolated. *Porcn* gene recombination was assessed using primers specifically detecting the wild-type, floxed and deleted allele, as described in Supplemental Information.

RNA isolation and gene expression analysis

RNA was isolated using TRIzol reagent (Invitrogen) and RNeasy mini kits (Qiagen), followed by treatment with DNaseI and reverse transcription using the Superscript III First-Strand Synthesis System (Invitrogen). Quantitative RT-PCR was performed using FastStart Universal SYBR Green Master (Roche) and gene-specific primers; data analysis is described in Supplemental Information.

Histochemistry and immunohistochemistry

Some mice received 1 mg bromodeoxyuridine (BrdU, Sigma) 1 hour before euthanasia. 5 μm paraffin tissue sections were stained with Hematoxylin & Eosin, Alcian blue, or specific antibodies (Ab) to Ki67, BrdU, Chromogranin A, β-catenin, Cleaved Caspase 3, GFP and Lysozyme. 10 μm frozen tissue sections were co-stained with Ki67, GFP, or SMA Ab, followed by incubation with Cy3-, FITC- or Alexa546-conjugated anti-IgG. See Supplemental Information for antibody sources and conditions.

Detailed descriptions of all materials and methods can be found in the Supplemental Information.

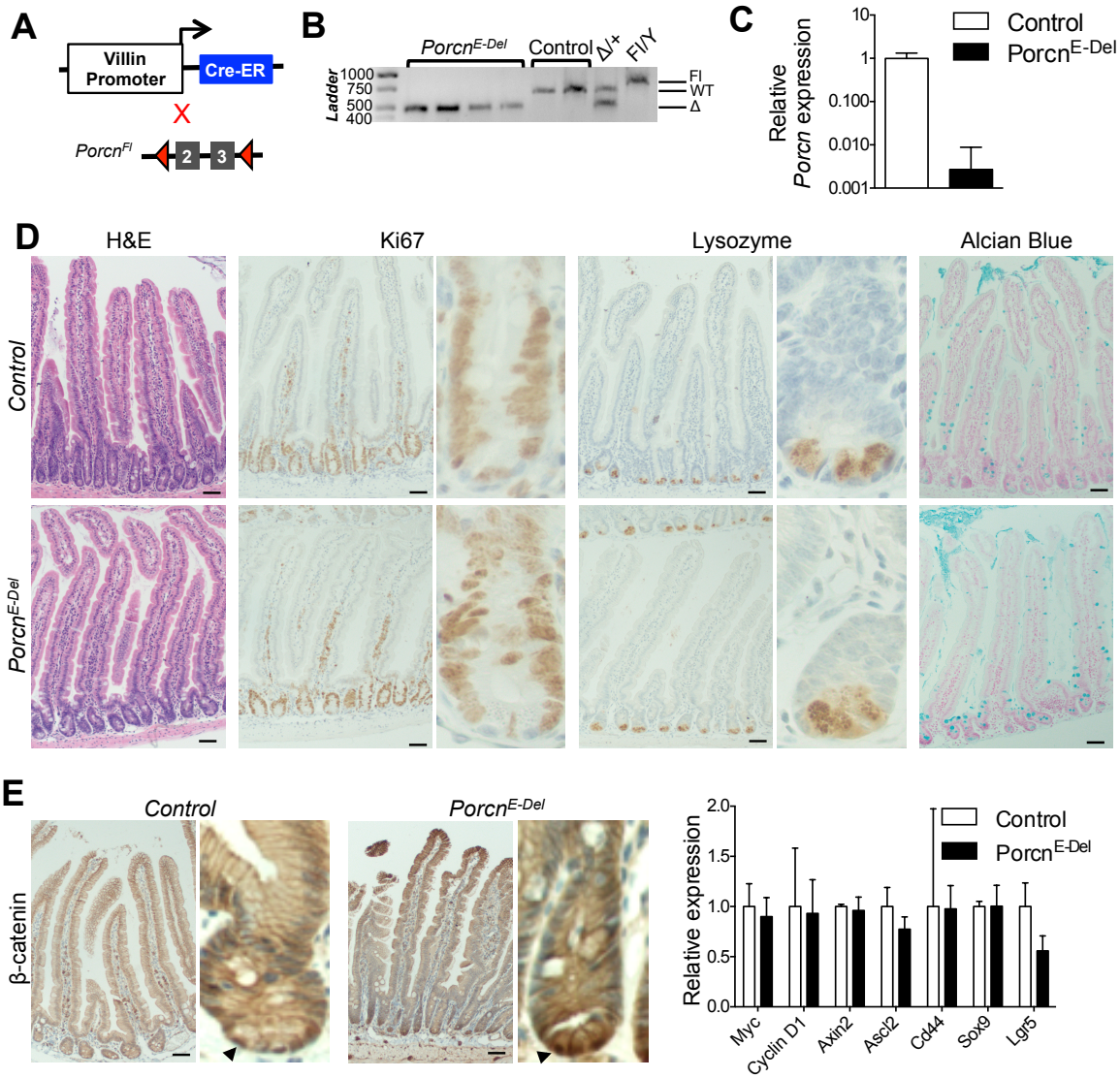
Results and Discussion

To test if the Wnt ligand(s) required for intestinal homeostasis originate in epithelial (including Paneth) cells, we crossed *Villin-Cre*^{ERT2} and *Porcn*$^{Fl/Y}$ mice (**Figure 2.1A**) (Barrott et al., 2011; el Marjou et al., 2004). LoxP sites flank exons 2 and 3 in the *Porcn*^{Fl} allele, poised for Cre recombinase-mediated deletion of the first 3 transmembrane domains to produce a null allele. As *Porcn* is X-linked, hemizygote males carry a single null allele. Although *Villin-Cre* is active in Paneth cells (Kim et al., 2012; Verzi et al., 2010), we examined *Porcn*$^{Fl/Y}$;*Villin-Cre*^{ERT2} (*Porcn*$^{E\text{-}Del}$ for epithelium-deleted) males at least 8 weeks after tamoxifen (TAM) exposure to ensure that Paneth cells were derived from *Porcn*-null ISCs. PCR on genomic DNA and qRT-PCR on mRNA isolated from the small intestine epithelium verified efficient *Porcn* deletion (**Figure 2.1B**) and loss of *Porcn* transcripts (**Figure 2.1C**, note logarithmic scale). *Porcn*$^{E\text{-}Del}$ mice thrived without weight loss or morbidity (**Figure S2.1A**), and tissue morphology and crypt

cell replication were intact. Both Paneth cells and goblet cells, which require Wnt signaling to differentiate (Pinto et al., 2003; van Es et al., 2005), were undisturbed (**Figures 2.1D, S2.1B,C**). Nuclear β-catenin was present in crypt base cells and the levels of well-characterized Wnt target mRNAs, including *Axin2*, *Myc*, *Cyclin D1*, *Cd44, Sox9* and *Lgr5* were equal in $Porcn^{E\text{-}Del}$ and wild-type small intestine crypt epithelium (**Figure 2.1E** and **Table S2.1**). These data reveal that intestinal epithelium is not an essential source of Wnt ligands *in vivo.*

Figure 2.1. *Porcn* deletion in epithelial cells spares intestinal epithelial morphology and function. (A) Strategy for *Porcn* deletion by TAM-inducible Cre recombinase driven by epithelium-specific *Villin* promoter. **(B)** Genomic PCR of epithelium isolated from 4 $Porcn^{E\text{-}Del}$ intestines indicates complete recombination of the floxed allele (485-bp band), compared to two $Porcn^{+/Y}$ control mice (685-bp band from the wild-type allele). For reference, DNA is also shown from single $Porcn^{Del/X}$ (485-bp and 685-bp products) and $Porcn^{Fl/Y}$;*Cre*- (762-bp band from the unrecombined floxed allele) mice. **(C)** qRT-PCR analysis of *Porcn* mRNA in isolated $Porcn^{E\text{-}Del}$ intestinal crypts (*N*=4) reveals reduction by orders of magnitude compared to controls (*N*=2). **(D)** Histology and immunohistochemistry (IHC) of $Porcn^{E\text{-}Del}$ (*N*=5) and control ($Porcn^{+/Y}$;*Villin-*Cre^{ERT2}, *N*=3) mice reveals no abnormalities. Left to right: hematoxylin and eosin (H&E) staining, Ki67 IHC, Lysozyme IHC, and Alcian blue staining. High-magnification images are shown to the right of low-magnification views. **(E)** Left, β-catenin IHC in control and $Porcn^{E\text{-}Del}$ mice shows nuclear staining crypt base cells (arrows). Right, qRT-PCR analysis of Wnt target mRNAs in isolated control (N=2) and $Porcn^{E\text{-}Del}$ (N=4) crypt epithelium demonstrates unperturbed Wnt pathway activity. Bars represent mean ± standard error of the mean (SEM) of biological replicates; all scale bars, 50 μm.

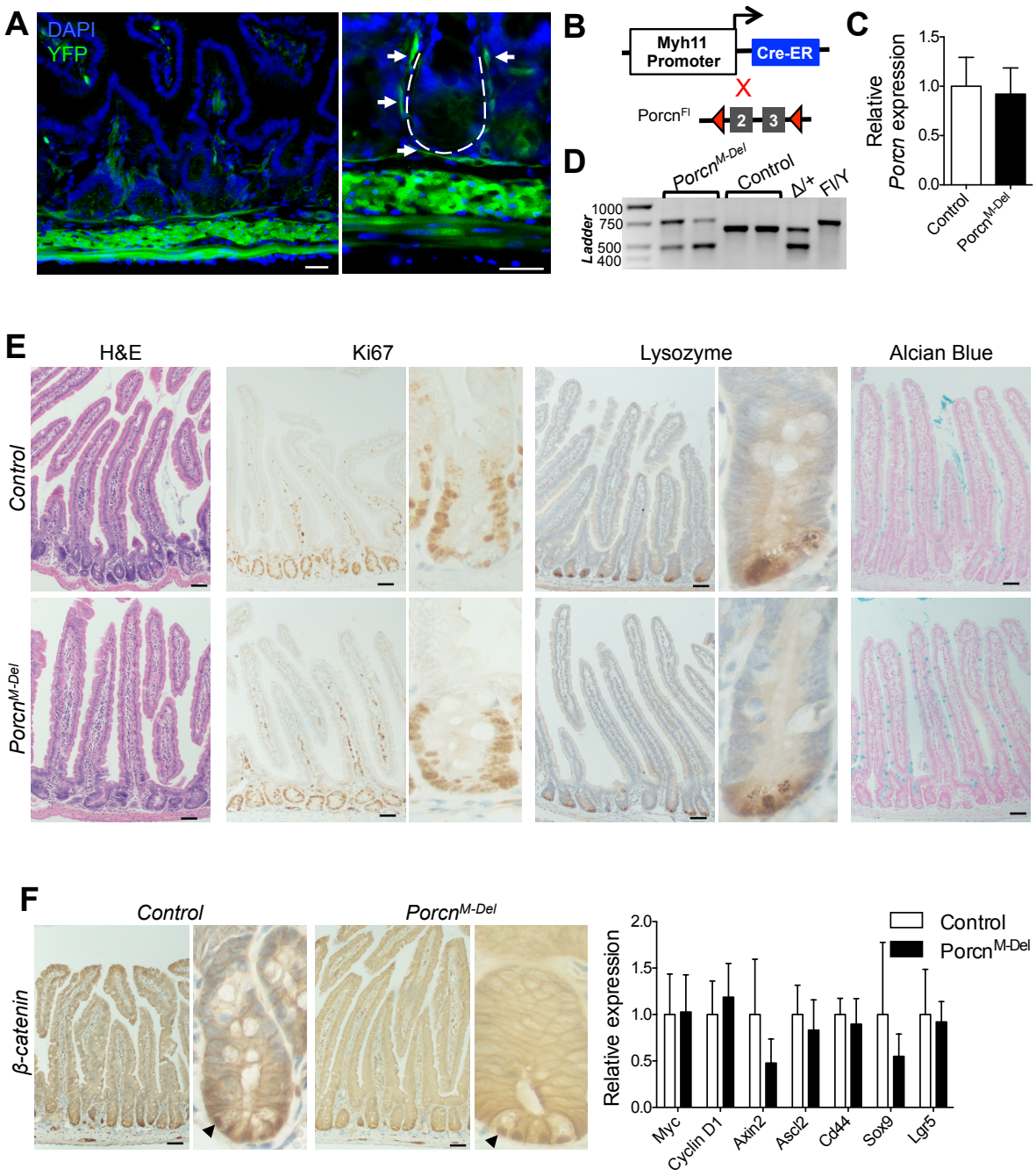
Figure 2.1 (Continued)



To determine if sub-epithelial cells, such as SEMFs, are a required physiologic Wnt source, we used *Myh11-Cre*^{ERT2} mice. *Myh11*–encoded smooth muscle myosin heavy chain is expressed in much of the intestinal sub-epithelium (Wirth et al., 2007). To define the precise expression domain, we crossed *Myh11-Cre*^{ERT2} to *Rosa26-lox-STOP-lox-YFP* (*Rosa26R*) reporter mice and treated with TAM. YFP expression was evident in circular and longitudinal muscle layers as well as in SEMF 5 days post Cre induction, co-localizing with α-smooth muscle actin (**Figures 2A, S2.2**). Most sub-epithelial cells expressed YFP but the lamina propria was not totally marked, indicating that *Myh11-Cre*^{ERT2} is expressed in most, but not all sub-epithelial lineages. We then crossed *Myh11-Cre*^{ERT2} and *Porcn*$^{Fl/Y}$ mice to ablate Wnt ligand secretion specifically in SEMF and smooth muscle (**Figure 2.2B**), while preserving *Porcn* expression in the epithelium (**Figure 2.2C**). To estimate the mesenchymal cell fraction with recombined *Porcn*, we genotyped *Porcn* alleles in unfractionated *Porcn*$^{Fl/Y}$*;Myh11-Cre*^{ERT2} *(Porcn*$^{M\text{-}Del}$) intestines, expecting no recombination in epithelial, serosal, endothelial or blood cells, which lack *Myh11* expression. In this light, a substantial contribution from recombined null *Porcn* DNA implied efficient recombination in most *Myh11*+ cells (**Figure 2.2D**), a point we demonstrate with greater confidence below (**Figure 2.3B**). Three weeks following Cre activation in *Porcn*$^{M\text{-}Del}$ males we did not observe weight loss (**Figure S2.1A**), implying preserved intestinal function. Gross and microscopic intestinal anatomy were intact, and *Porcn*$^{M\text{-}Del}$ did not differ from control mice in the numbers of proliferating Ki67+ crypt cells; Paneth or other secretory cells (**Figures 2.2E, S2.1B,C**); expression of nuclear β-catenin; or levels of Wnt target genes in crypt epithelium (**Figure 2.2F**). Thus, disruption of a sub-epithelial *Myh11*+ cell source of Wnts also did not perturb intestinal homeostasis.

Figure 2.2. *Porcn* loss in intestinal smooth muscle cells and SEMFs does not adversely affect epithelial morphology and function. **(A)** *Myh11-Cre^{ERT2};Rosa26R* mice activate Cre recombinase in all muscle layers, including subepithelial myofibroblasts. The dotted line in the high-magnification image (right) outlines a crypt and arrows point to the thin SEMF layer of YFP+ cells enveloping the crypt. Scale bars, 30 μm. See also **Figure S2.2**. **(B)** Strategy to induce muscle cell-specific recombination in *Porcn$^{Fl/Y}$;Myh11-Cre^{ERT2}* mice. **(C)** qRT-PCR analysis of isolated crypt epithelium shows no significant difference in *Porcn* mRNA expression in *Porcn$^{M\text{-}Del}$* (*N=4*) mice compared to control (*Porcn$^{+/Y}$;Myh11-Cre^{ERT2}, N=4*) mice, *P*=0.35. **(D)** PCR from whole intestine genomic DNA, revealing the expected proportion of recombined *Porcn* in smooth muscle, with residual unrecombined *Porcn* DNA contributed by epithelial and other cell types that lack *Myh11-Cre^{ERT2}* expression. **(E)** Histology and immunostains on *Porcn$^{M\text{-}Del}$* (*N*=7) and control (*Porcn$^{+/Y}$;Myh11-Cre^{ERT2}*, *N*=6) intestines reveal no abnormalities. **(F)** Left, β-catenin IHC in control and *Porcn$^{M\text{-}Del}$* mice shows nuclear staining in crypt base cells (arrows). Right, qRT-PCR reveals intact Wnt target expression in crypts isolated from Porcn$^{M\text{-}Del}$ intestines, compared to controls (*N*=4; statistics in **Table S2.1**). Bars represent mean ± SEM of biological replicates; scale bars E-F, 50 μm.

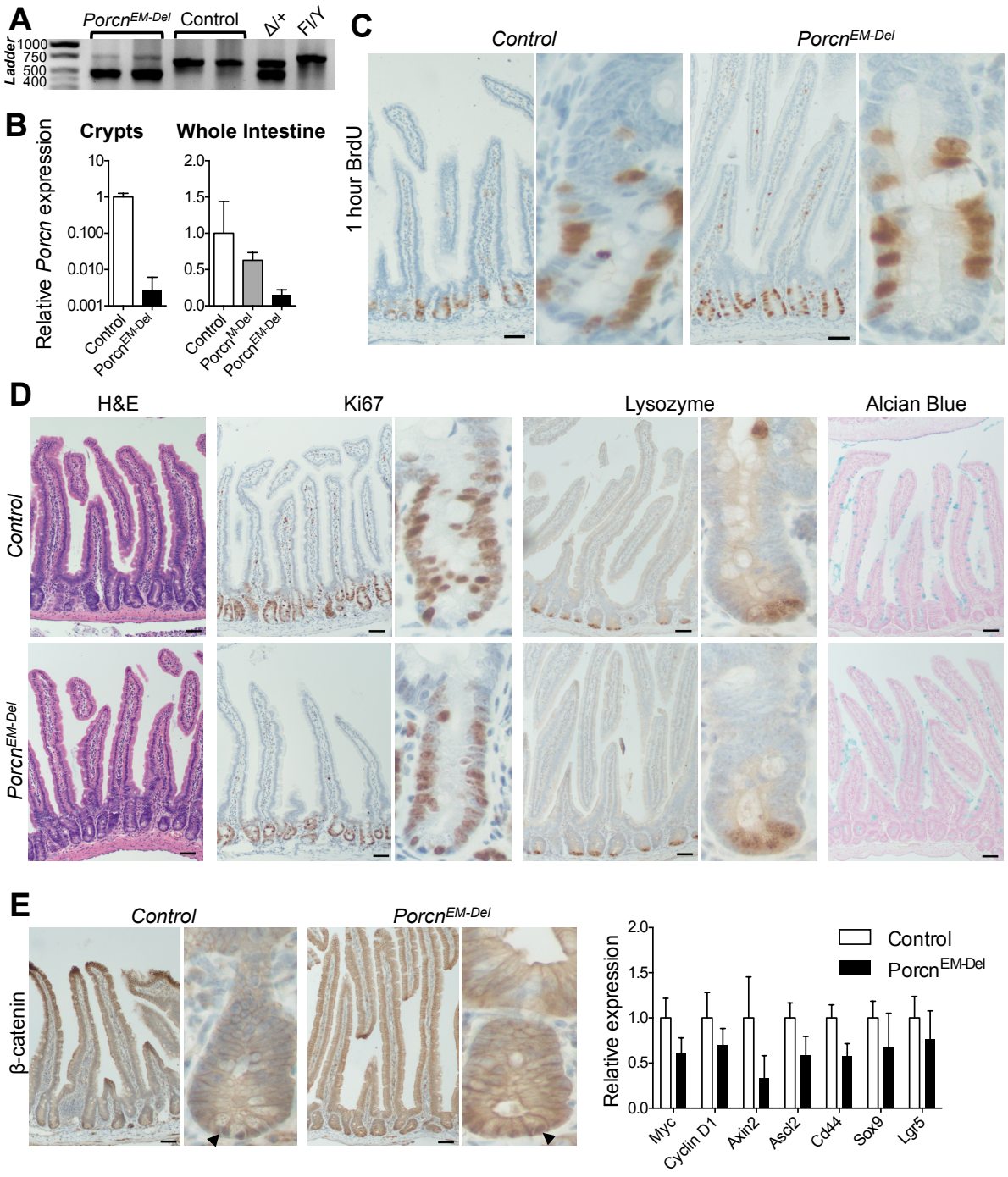
Figure 2.2 (Continued)



Intestinal organoid cultures from isolated $Lgr5^{+}$ ISCs require supplemental WNT3A, which is thought to derive *in vivo* from Paneth cells (Sato et al., 2011). However, mouse intestines devoid of epithelial *Wnt3* are normal and co-culture of $Lgr5^{+}$ ISCs with mesenchymal cells can substitute for WNT3A in organoid formation, suggesting redundant Wnt sources (Farin et al., 2012). To test this idea *in vivo*, we crossed $Porcn^{Fl/Y}$ mice onto a compound *Villin-*Cre^{ERT2}*;Myh11-*Cre^{ERT2} background and treated conditional-null mice with TAM to force *Porcn* loss from both compartments ($Porcn^{EM\text{-}Del}$). Three weeks post Cre activation, we observed nearly complete recombination of the floxed mutant allele in unfractionated intestines (**Figure 2.3A**), producing total absence of *Porcn* mRNA in the epithelium (**Figure 2.3B**). Unfractionated intestines showed 40% less *Porcn* mRNA in $Porcn^{M\text{-}Del}$ mice and 85% less in compound $Porcn^{EM\text{-}Del}$ mice (**Figure 2.3B**). Together, these data reveal highly efficient *Porcn* depletion in both compartments, with the expected residual contribution from *Villin*$^{-}$*;Myh11*$^{-}$ cells. Compound mutant mice showed no weight loss (**Figure S2.1A**) or clinical compromise; decrease in proliferation as assessed by Ki67 expression or bromodeoxyuridine (BrdU) uptake (**Figures 2.3C,D**); or defects in intestine morphology (**Figures 2.3D, S2.1B,C**). Strong immunostaining for nuclear β-catenin provided direct evidence for sustained Wnt signaling activity and qRT-PCR analysis revealed subtle, statistically insignificant deficits in Wnt target transcripts (**Figure 2.3E**). Thus, combined loss of *Porcn* in the gut epithelium and the dominant sub-epithelial compartment, including SEMFs, preserves all measurable Wnt-dependent functions.

Figure 2.3. Consequences of simultaneous loss of *Porcn* from epithelial and muscle cells. (A) DNA from whole intestines in TAM-treated *Porcn*$^{Fl/Y}$;*Villin-Cre*^{ERT2};*Myh11-Cre*^{ERT2} mice shows significant recombination at the floxed *Porcn* allele, with a minimal contribution of unrecombined DNA from non-epithelial, non-muscle cells. **(B)** *Left,* qRT-PCR analysis of RNA from crypts of *Porcn*$^{EM\text{-}Del}$ mutants (*N=3*) compared to controls (*Porcn*$^{+/Y}$;*Villin-Cre*^{ERT2};*Myh11-Cre*^{ERT2}; *N=3*) shows total *Porcn* deficiency, *P*=0.0013. *Right,* qRT-PCR of unfractionated *Porcn*$^{M\text{-}Del}$ (*N=2*) intestines shows lower *Porcn* expression than controls (*Porcn*$^{+/Y}$;*Villin-Cre*^{ERT2};*Myh11-Cre*^{ERT2} and *Porcn*$^{+/Y}$;*Myh11-Cre*^{ERT2}, N=4), which is further reduced in *Porcn*$^{EM\text{-}Del}$ (*N=2*) mice. **(C)** Equivalent BrdU uptake in control and *Porcn*$^{EM\text{-}Del}$ intestinal crypts. **(D)** Histology and immunostains from *Porcn*$^{EM\text{-}Del}$ and control intestines reveal intact morphology, cell proliferation, and differentiation (*N=3*). **(E)** Left, β-catenin IHC in control and *Porcn*$^{EM\text{-}Del}$ mice shows nuclear staining in crypt base cells (arrows). Right, qRT-PCR analysis of control (*N=3*) and *Porcn*$^{EM\text{-}Del}$ (*N=3*) intestines indicates a statistically insignificant (**Table S2.1**) reduction in levels of Wnt target transcripts. Bars in graphs represent mean ± SEM of biological replicates; all scale bars, 50 μm.

Figure. 2.3 (Continued)



These findings conflict with previous results from forced expression of Wnt antagonists, which produced crypt atrophy, villus shortening, loss of Wnt target transcripts, and failure of secretory cell differentiation within 2-4 days (Kuhnert et al., 2004). There are several possible reasons for the absence of similar defects in *Porcn*^{Del} intestines. First, Wnts necessary for intestinal homeostasis might circumvent the requirement for PORCN in secretion. This is unlikely because two independent studies reveal that PORCN seems necessary to palmitoylate all human Wnts; this in turn is required for Wnts to bind the carrier protein Wntless for secretory transport (Coombs et al., 2010; Liu et al., 2013) and to bind Frizzled receptors (Janda et al., 2012; Najdi et al., 2012). Second, the potent effect of Wnt antagonists on crypt functions might not reflect the native activity of Wnts *per se*, but rather of R-spondin or another family of ligands. Available evidence, however, indicates that R-spondins act in conjunction with, and not separate from, Wnts (Niehrs, 2012). A third possibility is that Wnt reserves in the sub-epithelial basement membrane or elsewhere persisted for the duration of our experiments. Wnts can bind heparan sulfate proteoglycans present on the surface of Wnt-recipient cells, a proposed mechanism to prevent their diffusion and allow prolonged activity (Mikels and Nusse, 2006). However, even in the absence of information on intestinal Wnt concentrations, turnover or reserves, secreted Wnts are unlikely to have persisted for the length of our studies. *Porcn* deletion was efficient and we deliberately examined mice weeks after Cre activation, giving time for Wnt reserves to decay. Moreover, in a tissue that self-renews continually and responds quickly to injury or cell loss (Clevers, 2013), Wnts probably turn over rapidly to allow responsive homeostasis.

We therefore favor the final possibility: that a cell type that evaded Cre-mediated *Porcn* deletion is a sufficient source of essential, intestine-active Wnts. That cell is unlikely to reside in the epithelium, where *Villin-Cre*^{ERT2} mice drove efficient Cre expression, leaving no intact *Porcn*

DNA in isolated epithelial cells. By contrast, although *Myh11*-expressing smooth muscle cells or SEMF are not a required physiologic Wnt source, gut mesenchyme contains diverse additional cell types, including endothelium, non-muscle fibroblasts, leukocytes, lymphocytes, and neurons (Powell et al., 2011). Endothelial and neuronal contributions cannot readily be assessed in mice because their deficiencies are lethal early (Dumont et al., 1994; Enomoto et al., 1998) and intestinal functions seem intact in *Rag2*-null mice (Shinkai et al., 1992), indicating that lymphocytes either provide no essential Wnts or act redundantly with other cells. This extent of specificity or redundancy in intestinal Wnt source(s) challenges the prevailing view and reveals unanticipated complexity in control of intestinal self-renewal.

Acknowledgements

We thank Sylvie Robine for *Villin*$^{Cre\text{-}ERT2}$ mice. This work was supported by National Institute of Health grants R01DK081113 and R01DK082889 to R.A.S. and a National Science Foundation Graduate Research Fellowship to A.S.R.

References

Barker, N., Ridgway, R.A., van Es, J.H., van de Wetering, M., Beghtel, H., van den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J., and Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. Nature *457*, 608-611.

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

Barrott, J.J., Cash, G.M., Smith, A.P., Barrow, J.R., and Murtaugh, L.C. (2011). Deletion of mouse Porcn blocks Wnt ligand secretion and reveals an ectodermal etiology of human focal dermal hypoplasia/Goltz syndrome. Proc Natl Acad Sci U S A *108*, 12752-12757.

Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., Bibeau, F., Scherer, G., Joubert, D., Hollande, F.*, et al.* (2007). Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol *178*, 635-648.

Biechele, S., Cockburn, K., Lanner, F., Cox, B.J., and Rossant, J. (2013). Porcn-dependent Wnt signaling is not required prior to mouse gastrulation. Development *140*, 2961-2971.

Biechele, S., Cox, B.J., and Rossant, J. (2011). Porcupine homolog is required for canonical Wnt signaling and gastrulation in mouse embryos. Dev Biol *355*, 275-285.

Carmon, K.S., Gong, X., Lin, Q., Thomas, A., and Liu, Q. (2011). R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling. Proc Natl Acad Sci U S A *108*, 11452-11457.

Chen, B., Dodge, M.E., Tang, W., Lu, J., Ma, Z., Fan, C.-W., Wei, S., Hao, W., Kilgore, J., Williams, N.S.*, et al.* (2009). Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nat Chem Biol *5*, 100-107.

Cheng, H., and Leblond, C.P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. Am J Anat *141*, 537-561.

Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. Cell *154*, 274-284.

Coombs, G.S., Yu, J., Canning, C.A., Veltri, C.A., Covey, T.M., Cheong, J.K., Utomo, V., Banerjee, N., Zhang, Z.H., Jadulco, R.C.*, et al.* (2010). WLS-dependent secretion of WNT3A requires Ser209 acylation and vacuolar acidification. J Cell Sci *123*, 3357-3367.

de Lau, W., Barker, N., Low, T.Y., Koo, B.-K., Li, V.S.W., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M.*, et al.* (2011). Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. Nature *476*, 293-297.

Dumont, D.J., Gradwohl, G., Fong, G.H., Puri, M.C., Gertsenstein, M., Auerbach, A., and Breitman, M.L. (1994). Dominant-negative and targeted null mutations in the endothelial

receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo. Genes Dev *8*, 1897-1909.

Durand, A., Donahue, B., Peignon, G., Letourneur, F., Cagnard, N., Slomianny, C., Perret, C., Shroyer, N.F., and Romagnolo, B. (2012). Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). Proc Natl Acad Sci U S A *109*, 8965-8970.

el Marjou, F., Janssen, K.-P., Chang, B.H.-J., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186-193.

Enomoto, H., Araki, T., Jackman, A., Heuckeroth, R.O., Snider, W.D., Johnson, E.M., Jr., and Milbrandt, J. (1998). GFR alpha1-deficient mice have deficits in the enteric nervous system and kidneys. Neuron *21*, 317-324.

Farin, H.F., Van Es, J.H., and Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. Gastroenterology *143*, 1518-1529 e1517.

Garabedian, E.M., Roberts, L.J., McNevin, M.S., and Gordon, J.I. (1997). Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. J Biol Chem *272*, 23729-23740.

Glinka, A., Dolde, C., Kirsch, N., Huang, Y.L., Kazanskaya, O., Ingelfinger, D., Boutros, M., Cruciat, C.M., and Niehrs, C. (2011). LGR4 and LGR5 are R-spondin receptors mediating Wnt/beta-catenin and Wnt/PCP signalling. EMBO Rep *12*, 1055-1061.

Gregorieff, A., Pinto, D., Beghtel, H., Destrée, O., Kielman, M., and Clevers, H. (2005). Expression pattern of Wnt signaling components in the adult intestine. Gastroenterology *129*, 626-638.

Grzeschik, K.-H., Bornholdt, D., Oeffner, F., König, A., del Carmen Boente, M., Enders, H., Fritz, B., Hertl, M., Grasshoff, U., Höfling, K*., et al.* (2007). Deficiency of PORCN, a regulator of Wnt signaling, is associated with focal dermal hypoplasia. Nat Genet *39*, 833-835.

Janda, C.Y., Waghray, D., Levin, A.M., Thomas, C., and Garcia, K.C. (2012). Structural basis of Wnt recognition by Frizzled. Science *337*, 59-64.

Kim, B.M., Mao, J., Taketo, M.M., and Shivdasani, R.A. (2007). Phases of canonical Wnt signaling during the development of mouse intestinal epithelium. Gastroenterology *133*, 529-538.

Kim, T.-H., Escudero, S., and Shivdasani, R.A. (2012). Intact function of Lgr5 receptor-expressing intestinal stem cells in the absence of Paneth cells. Proc Natl Acad Sci U S A *109*, 3932-3937.

Kispert, A., Vainio, S., and McMahon, A.P. (1998). Wnt-4 is a mesenchymal signal for epithelial transformation of metanephric mesenchyme in the developing kidney. Development *125*, 4225-4234.

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P.J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet *19*, 379-383.

Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science *275*, 1784-1787.

Kuhnert, F., Davis, C.R., Wang, H.-T., Chu, P., Lee, M., Yuan, J., Nusse, R., and Kuo, C.J. (2004). Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. Proc Natl Acad Sci U S A *101*, 266-271.

Lahar, N., Lei, N.Y., Wang, J., Jabaji, Z., Tung, S.C., Joshi, V., Lewis, M., Stelzner, M., Martín, M.G., and Dunn, J.C.Y. (2011). Intestinal subepithelial myofibroblasts support in vitro and in vivo growth of human small intestinal epithelium. PLoS One *6*, e26898.

Liu, J., Pan, S., Hsieh, M.H., Ng, N., Sun, F., Wang, T., Kasibhatla, S., Schuller, A.G., Li, A.G., Cheng, D*., et al.* (2013). Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. Proc Natl Acad Sci U S A *110*, 20224-20229.

Liu, W., Shaver, T.M., Balasa, A., Ljungberg, M.C., Wang, X., Wen, S., Nguyen, H., and Van den Veyver, I.B. (2012). Deletion of Porcn in mice leads to multiple developmental defects and models human focal dermal hypoplasia (Goltz syndrome). PLoS One *7*, e32331.

Mikels, A.J., and Nusse, R. (2006). Wnts as ligands: processing, secretion and reception. Oncogene *25*, 7461-7468.

Mori-Akiyama, Y., van den Born, M., van Es, J.H., Hamilton, S.R., Adams, H.P., Zhang, J., Clevers, H., and de Crombrugghe, B. (2007). SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. YGAST *133*, 539-546.

Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science *275*, 1787-1790.

Najdi, R., Proffitt, K., Sprowl, S., Kaur, S., Yu, J., Covey, T.M., Virshup, D.M., and Waterman, M.L. (2012). A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. Differentiation *84*, 203-213.

Niehrs, C. (2012). The complex world of WNT receptor signalling. Nat Rev Mol Cell Biol *13*, 767-779.

Park, J.S., Valerius, M.T., and McMahon, A.P. (2007). Wnt/beta-catenin signaling regulates nephron induction during mouse kidney development. Development *134*, 2533-2539.

Pinto, D., Gregorieff, A., Beghtel, H., and Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. Genes Dev *17*, 1709-1713.

Powell, D.W., Pinchuk, I.V., Saada, J.I., Chen, X., and Mifflin, R.C. (2011). Mesenchymal cells of the intestinal lamina propria. Annu Rev Physiol *73*, 213-237.

Proffitt, K.D., and Virshup, D.M. (2012). Precise Regulation of Porcupine Activity Is Required for Physiological Wnt Signaling. J Biol Chem *287*, 34167-34178.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415-418.

Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A.M.*, et al.* (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855-867.

Shroyer, N.F., Wallis, D., Venken, K.J.T., Bellen, H.J., and Zoghbi, H.Y. (2005). Gfi1 functions downstream of Math1 to control intestinal secretory cell subtype allocation and differentiation. Genes Dev *19*, 2412-2417.

Srinivas, S., Watanabe, T., Lin, C.S., William, C.M., Tanabe, Y., Jessell, T.M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol *1*, 4.

van Es, J.H., Haegebarth, A., Kujala, P., Itzkovitz, S., Koo, B.K., Boj, S.F., Korving, J., van den Born, M., van Oudenaarden, A., Robine, S.*, et al.* (2012). A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. Mol Cell Biol *32*, 1918-1927.

van Es, J.H., Jay, P., Gregorieff, A., van Gijn, M.E., Jonkheer, S., Hatzis, P., Thiele, A., van den Born, M., Begthel, H., Brabletz, T.*, et al.* (2005). Wnt signalling induces maturation of Paneth cells in intestinal crypts. Nat Cell Biol *7*, 381-386.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell *19*, 713-726.

Wang, X., Reid Sutton, V., Omar Peraza-Llanes, J., Yu, Z., Rosetta, R., Kou, Y.-C., Eble, T.N., Patel, A., Thaller, C., Fang, P.*, et al.* (2007). Mutations in X-linked PORCN, a putative regulator of Wnt signaling, cause focal dermal hypoplasia. Nat Genet *39*, 836-838.

Wirth, A., Benyó, Z., Lukasova, M., Leutgeb, B., Wettschureck, N., Gorbey, S., Örsy, P., Horváth, B., Maser-Gluth, C., Greiner, E.*, et al.* (2007). G12-G13–LARG–mediated signaling in vascular smooth muscle is required for salt-induced hypertension. Nat Med *14*, 64-68.

Chapter 3: CDX2 is a critical regulator of intestinal stem cell functions

Contributions:
A. K. San Roman conceived of this project, with guidance from R. A. Shivdasani. Our collaborators A. Tovaglieri and D. Breault performed organoid culture experiments. A. K. San Roman carried out all other experiments and analysis and wrote this chapter.

Summary

Intestinal stem cells (ISCs) renew the adult gut epithelium and are a useful model to study stem cell biology and control of cell differentiation. Self-renewal of ISCs and their differentiation requires control of genes by transcription factors (TFs) and chromatin through mechanisms that are not well understood. One pool of ISCs, marked by *Lgr5*, resides at the base of the crypts, cycles rapidly, and can be isolated using *Lgr5-GFP* knock-in mice. To gauge the scope of gene expression changes during cell differentiation, we performed global RNA-seq analysis of $Lgr5^{+}$ ISCs and purified enterocytes, the most abundant differentiated cell type. This analysis identified genes enriched in ISCs or enterocytes and others expressed equally in both populations. One gene highly expressed in both, *Cdx2*, encodes a homeodomain TF that maintains active chromatin at thousands of enterocyte genes. Abundant *Cdx2* expression in ISCs suggested a previously unappreciated role in this cell. We observed replication defects of $Cdx2^{-/-}$ mouse ISCs, and an inability to produce mature cell types or to generate intestinal organoids in vitro. These findings reveal functions for CDX2 in ISCs distinct from its requirement to activate enterocyte genes for nutrient digestion and absorption. To identify direct CDX2 targets, we performed ChIP-seq, and RNA-seq of wild-type and $Cdx2^{-/-}$ ISCs. Underscoring the role for CDX2 in cell proliferation, analysis of genes dysregulated in $Cdx2^{-/-}$ ISCs revealed significant aberration in genes related to the cell cycle. We detected CDX2 binding near 10% of dysregulated genes, including *Fgfbp1*, a factor involved in promoting fibroblast growth factor signaling, which may account for the observed defects. This study reveals a broad role for CDX2 in maintaining stem cell properties and

functions and identifies specific transcriptional targets that may mediate canonical stem cell behaviors.

Introduction

The intestinal epithelium is a continuously renewing tissue that relies on a population of adult intestinal stem cells (ISCs) to differentiate into mature cell types to maintain the physiological functions of the intestine. This regenerative capacity requires the precise execution of gene expression programs during ISC maintenance and differentiation to mature cell types, but the mechanisms controlling these processes remain unclear. The intestinal epithelium represents an excellent system to study cell differentiation and stem cell dynamics due to its highly compartmentalized structure. The crypts of Lieberkühn contain stem cells, which give rise to a highly proliferative population of transit-amplifying (TA) cells. TA cells produce the mature cell types found in the villi, which are structures that protrude into the lumen of the intestine, thus increasing the surface area. Absorptive enterocytes make up the majority of cells in the villi, followed by mucous secreting goblet cells and hormone secreting enteroendocrine cells. The final major differentiated cell type is the Paneth cell, which migrates to the base of the crypt and is nestled between crypt base columnar cells (CBCs).

One particular challenge to studying cell differentiation in the intestinal epithelium is that several populations of ISCs have been described. There are at least two types of ISC populations in intestinal crypts: the proliferative CBCs, marked by *Lgr5*, and quiescent cells located at the +4 position or higher that respond to epithelial

injury. Markers of quiescent ISCs include *Bmi1, Lrig1, Hopx,* and *mTert* (Montgomery et al., 2011; Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011). The relationships between these types of ISCs is bidirectional; studies have demonstrated that $Bmi1^+$, $mTert^+$ and $Hopx^+$ cells can give rise to $Lgr5^+$ ISCs and vice versa (Montgomery et al., 2011; Takeda et al., 2011; Tian et al., 2011; Yan et al., 2012). In our study we chose to use $Lgr5^+$ cells for several reasons. First, $Lgr5^+$ CBCs contribute to all mature villus cell lineages over many months *in vivo* (Barker et al., 2007). Second, in culture, $Lgr5^+$ CBCs are uniquely capable of generating "organoids" that self-renew for months (Sato et al., 2009). Third, $Lgr5^+$ CBCs can integrate into the mouse intestine and give rise to normal epithelial tissue when transplanted (Yui et al., 2012). Fourth, $Lgr5^+$ CBCs are monoclonal (a crypt hallmark) and can give rise to other ISC populations (Barker et al., 2007; Takeda et al., 2011). Finally, recent data show that $Lgr5^+$ CBCs express the candidate ISC markers *Bmi1*, *mTert*, *Hopx* and *Lrig1* (Itzkovitz et al., 2011; Muñoz et al., 2012).

Since their identification in 2007, there has been considerable effort to identify factors controlling functions of $Lgr5^+$ ISCs. Not surprisingly, several transcription factors (TFs) have been described to be specifically required for ISC functions in normal and during times of injury. These include ASCL2, KLF5, ID1, YY1, and VDR (Bell and Shroyer, 2014; Nakaya et al., 2014; Peregrina et al., 2014; Perekatt et al., 2014; van der Flier et al., 2009; Zhang et al., 2014). Many of these studies used genetic knockout mice to show specific defects in the functions of ISCs, however the specific transcriptional mechanisms by which these TFs control ISC genes have not been worked out, largely

because it is difficult to isolate the large cell numbers required for TF chromatin-immunoprecipitation and sequencing (ChIP-seq) experiments.

To better understand the cell differentiation process and functions of ISCs, we compared the global gene expression profile of purified absorptive enterocytes to *Lgr5*+ ISCs to by RNA-seq. In addition to the genes that are differentially expressed during differentiation, we were interested to find that CDX2, a homeodomain TF that we had previously studied in the context of mature enterocyte gene expression (Verzi et al., 2010), was highly and equally expressed in both populations, suggesting it may have an unappreciated role in the ISCs. By crossing our $Cdx2^{Fl/Fl}$ mice to *Lgr5-GFP-*CRE^{ERT2} knock-in mice (Barker et al., 2007), we identified a cell autonomous proliferation defect in ISCs. Lineage tracing revealed that *Cdx2*-null ISCs are unable to differentiate into mature cell types of the intestine, and instead form undifferentiated cystic structures. In addition, crypts lacking *Cdx2* are unable to form mature organoids in culture, corroborating previous reports (Stringer et al., 2012). To find the mechanism by which CDX2 affects ISC functions, we performed chromatin immunoprecipitation and sequencing (ChIP-seq) and gene expression profiling by RNA-seq on isolated control and $Cdx2^{-/-}$ ISCs. After identifying direct target genes of CDX2 in the ISCs, we honed in on the fibroblast growth factor (FGF) signaling pathway and plan to investigate its contributions to ISC functions further in the future.

Experimental Procedures

Mouse strains, conditional knockouts and genetic lineage tracing

To deplete *Cdx2* in *Lgr5*$^{+}$ ISCs or in the entire intestinal epithelium, *Cdx2*$^{Fl/Fl}$ mice were crossed with either *Lgr5-EGFP-Cre*^{ERT2} or *Villin-Cre*^{ERT2} mice, respectively (Barker et al., 2007; el Marjou et al., 2004; Verzi et al., 2010). For ISC isolation, *Cdx2*$^{Fl/Fl}$*;Lgr5-EGFP-Cre*^{ERT2} and control (*Cdx2*$^{+/+}$*;Lgr5-EGFP-Cre*^{ERT2}) mice were injected with 2mg tamoxifen (TAM) for 4 consecutive days and Lgr5-GFP^{High} cells, representing the crypt base columnar stem cells, were harvested on the following day via fluorescence activated cell sorting (FACS). Whole epithelium devoid of *Cdx2* was generated by injecting *Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} and control (non-Cre *Cdx2*$^{Fl/Fl}$*)* mice with 1mg TAM for 5 consecutive days and intestines were harvested at least 3 days later. To obtain pure enterocytes, *Atoh1*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} mice were injected with 1mg TAM for 5 consecutive days to eliminate secretory cells and villus epithelium was harvested 3 weeks later. For lineage tracing experiments, *Cdx2*$^{Fl/Fl}$*;Lgr5-EGFP-Cre*^{ERT2}*;Rosa26-lox-STOP-lox-YFP* and control (*Cdx2*$^{+/+}$*;Lgr5-EGFP-Cre*^{ERT2}*;Rosa26-lox-STOP-lox-YFP*) mice were injected with 2mg TAM for 5 consecutive days and intestines were harvested at each of the following time points after the first tamoxifen injection: 10 days, 2 weeks, 1 month, 3 months, and 6 months.

Isolation of Lgr5^{+} *ISCs, crypts and villi*

ISCs were harvested from mice containing the *Lgr5-EGFP-Cre*^{ERT2} allele, which is expressed in a mosaic fashion throughout the intestine. As the frequency of marked ISCs is higher in the proximal intestine, we used the proximal half of the small intestine

to isolate crypts for sorting. Intestines were isolated and flushed with cold phosphate-buffered saline (PBS), splayed open and the surface was gently scraped using glass slides to remove the majority of villi. The remaining tissue was incubated in 5mM EDTA in PBS at 4°C for 30 minutes on a shaker with intermittent vigorous shaking by hand. Tissues were transferred to fresh 5mM EDTA in PBS at 4°C for 15 additional minutes and shaken by hand for 2 minutes to remove the crypt epithelium. Isolated epithelium from both steps were pooled and disaggregated by treatment with 4X TrypLE (Invitrogen) in DMEM without phenol red (Mediatech) at 37°C for 45 minutes on a rotator. Cells were pelleted, washed in PBS, stained with Live/Dead cell viability dye (Invitrogen) and sorted using the FACSAria III instrument (BD Biosciences). Single, live cells expressing high levels of GFP were collected.

To harvest whole crypts and villi, intestines were isolated and flushed as above, but villi were not scraped. Following EDTA incubations and shaking, the detached epithelium was poured through a 70 μm filter, upon which the villi are captured and crypts flow through. Villi and crypts were pelleted and snap-frozen in liquid nitrogen for protein analysis or resuspended in TRIzol reagent for RNA isolation, and stored at -80°C until ready for use.

Cell cycle analysis of isolated Lgr5$^+$ ISCs

For cell cycle analysis of isolated *Lgr5*$^+$ ISCs, we followed the ISC isolation and cell sorting protocol, with the addition of 10 μg/ml Hoechst Dye (Invitrogen) during trypsinization. Cell cycle analysis was performed by gating on single, Lgr5-GFP^{High} cells, followed by detection with the UV laser for Hoechst Dye. Data were analyzed with

Flowjo software (Tree Star) using the Watson Pragmatic Test to determine cell cycle phase distribution. This experiment was repeated 3 times on different mice with similar results, and one representative experiment is presented here.

RNA-Seq

Approximately 100,000 ISCs or villus cells were collected from a single mouse and mRNA was extracted using oligo $d(T)_{25}$ magnetic beads (New England Biolabs). On-column DNaseI treatment was subsequently performed using the RNeasy kit (Qiagen). mRNA was quantified using Ribogreen (Life Technologies) and 500 picograms of mRNA was used as input for cDNA synthesis and library generation using the Encore Complete RNA-Seq Library System (Nugen) modified for ½ volume reactions. Size selection of the libraries for 200-450 base pair fragments was performed using Pippin Prep (Sage Science). Sequencing was performed (50 bp, single end reads) using the Illumina Hi-Seq 2000 and sequence tags were aligned to the *Mus musculus* reference genome build 9 (mm9). The Tuxedo software package was used to align reads, assemble transcripts, and determine differences in transcript levels using an FDR of 0.05 (Trapnell et al., 2012). Two biological replicates were sequenced and compared for each condition. Some 3' bias was noted for the samples, however this was comparable between samples and the results are highly concordant with data on ISC expression from other labs such as (Munoz et al., 2012; Sheaffer et al., 2014).

Gene ontology and network analysis

Significantly differentially expressed genes from RNA-seq were analyzed using DAVID functional annotation clustering software (default options, medium stringency) to identify significantly enriched gene ontology clusters (scores > 1.3) (Huang da et al., 2009). To represent clusters, we chose a representative gene ontology (GO) term and list it with the enrichment score. Gene regulatory networks were generated from the down-regulated genes in *Cdx2*-null ISCs using Metacore software (Thomson Reuters).

qRT-PCR

RNA was isolated using TRIzol reagent (Invitrogen) and the RNeasy kit (Qiagen), with on-column DNase treatment. RNA was reversed transcribed (Superscript III, Invitrogen) and transcripts were assessed by quantitative RT-PCR with FastStart Universal SYBR Green Master Mix (Roche) using the following specific primers:

Gene	Forward	Reverse
Fgf1	GAAGCATGCGGAGAAGAACTG	CGAGGACCGCGCTTACA
Fgf2	GTCACGGAAATACTCCAGTTGGT	CCCGTTTTGGATCCGAGTTT
Fgf3	CTGGAGATTACTGCGGTGGAA	CCAGGTACCGCCCAGAAA
Fgf4	TGCCTTTCTTTACCGACGAGT	GCGTAGGATTCGTAGGCGTT
Fgf6	GGATATAAGGACCCATGGATGAAG	GCAGAAGCGTTCTCTTAGTGTAAAATC
Fgf7	TGGGCACTATATCTCTAGCTTGC	GGGTGCGACAGAACAGTCT
Fgf8	GGTCTCTACATCTGCATGAACAAGA	CTTGCCTTTGCCGTTGCT
Fgf9	TGTGGACACCGGAAGGAGATA	GCCGTTTAGTCCTGGTCCC
Fgf10	TTTGGTGTCTTCGTTCCCTGT	TAGCTCCGCACATGCCTTC
Fgf16	GTGTTTTCCGGGAACAGTTTGA	GGTGAGCCGTCTTTATTCAGG
Fgf18	AAGGAGTGCGTGTTCATTGAG	AGCCCACATACCAACCAGAGT
Fgf19	AGAGAACAGCTCCAGGACCA	TTCTCCATCCTGTCGGAATC
Fgf22	CTCTGTGGACTGTAGGTTCCG	GAGGCGTATGTGTTGTAGCC
Fgf23	ATGCTAGGGACCTGCCTTAGA	AGCCAAGCAATGGGGAAGTG

Statistical significance was calculated by *t*-test adjusted for multiple comparisons using Graphpad Prism.

Histology, immunofluorescence and immunohistochemistry

Intestinal segments were splayed open and fixed in 4% paraformaldehyde overnight at 4°C. For histological analysis, fixed tissues were washed in PBS and dehydrated in 70% ethanol, and processed for paraffin embedding as swiss rolls. Hematoxylin and eosin (H&E), alkaline phosphatase, alcian blue staining and immunohistochemistry were performed according to standard procedures and as described previously (Verzi et al., 2010). For immunohistochemistry, tissues were probed with goat anti-CRS4C (1:1000, gift of Andre Ouellette, University of Southern California, Los Angeles, CA) or pre-treated in a pressure cooker in 10 mM sodium citrate buffer (pH 6) for antigen retrieval followed by antibodies against mouse anti-CDX2 (1:20, Biogenex), mouse anti-KI67 (1:1,000, Vector labs), rabbit anti-KI67 (1:200, Santa Cruz), rabbit anti-Cleaved Caspase 3 (1:1,000, Cell Signaling), rabbit anti-Chromogranin A (1:1,000, DNA Star), mouse anti-GFP (1:100, Santa Cruz), rabbit anti-GFP/YFP (1:100, Cell Signaling), mouse anti-β Catenin (1:250, BD Biosciences), mouse anti-Ezrin (1:100, Neomarker), and rabbit anti-NHERF (1:100, Abcam). Quantification of KI67+ cells was performed by counting 25 crypts in at least 3 mice.

Electron microscopy

Mouse ilea were flushed with PBS, fixed overnight or longer in EM fixative (2% formaldehyde, 2.5 % glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4), and embedded in Taab 812 Resin (Marivac Ltd., Nova Scotia, Canada). 80 nm sections were cut, stained with 0.2% Lead Citrate, viewed, and imaged with a Philips Technai BioTwin Spirit Electron Microscope at an accelerating voltage of 80 kV.

Intestinal organoid culture

Crypts from the proximal small intestine of *Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} and control (no Cre, *Cdx2*$^{Fl/Fl}$) mice were isolated and cultured in matrigel and organoid growth factor media, without Wnt, as previously described (Sato et al., 2009). 4-OH-Tamoxifen (4-OHT; Sigma) was added to the media at a concentration of 1 μm for the first two days to activate the *Cre*^{ERT2}. Afterwards, media without 4-OHT was replaced and changed every other day until day 8, when the organoids were fixed for 40 minutes in 4% PFA. Organoids were rinsed in PBS and 70% ethanol and resuspended in HistoGel (Thermo Scientific). Gel-embedded organoids were then treated the same as tissues for processing, paraffin embedding, sectioning and immunohistochemistry.

ChIP-seq

Lgr5-GFP^{High} ISCs were collected via FACS and chromatin was cross-linked in 1% formaldehyde at room temperature for 25 minutes. 5 million cells were pooled for ChIP, performed as previously described (Verzi et al., 2013), but using ½ reaction volumes and 1.5 μg of CDX2 antibody (Bethyl). Immunoprecipitated and input (non-immunoprecipitated) DNA was quantified using Picogreen (Life Technologies) and the ThruPLEX-FD Prep Kit (Rubicon Genomics) was used to generate libraries, according to the manufacturer's instructions. Size selection of libraries for 200-450 base pair fragments was performed using Pippin Prep (Sage Science) and sequenced using the Illumina Hi-Seq 2000 (50 bp, single end reads). Sequences were aligned to the *Mus musculus* reference genome build 9 (mm9). Binding peaks were identified using MACS

software, version 1.4.2, (Zhang et al., 2008) considering the non-immunoprecipitated input sample and local background. Motif calling within ChIP-Seq samples was performed using SeqPos; distribution of binding peaks within genomic features using CEAS; and conservation scores of binding sites using Conservation Plot, all tools in the Cistrome project (http://cistrome.org/ap/) (Liu et al., 2011). Peak to gene assignments were performed using GREAT and the single nearest gene within 100 kb was considered.

Protein extraction and Western blotting

Whole cell protein extracts were obtained by lysing mouse epithelium pellets in RIPA buffer, supplemented with 1mM PMSF, 2mM sodium orthovanadate, and protease inhibitors (Roche) on ice for 15 minutes. The concentration of the resulting lysates were measured using the Pierce BCA Protein Assay Kit for equal loading, boiled in Laemmli sample buffer for 5 minutes, resolved on a 10% SDS-PAGE gel, and transferred to nitrocellulose membranes. For immunoblotting, membranes were probed overnight at 4°C with the following primary antibodies: mouse anti-Actin (1:200, Santa Cruz); rabbit anti-FGFR3 (1:200, Santa Cruz); or rabbit anti-FGFR4 (1:100, Santa-Cruz).

Results

Gene expression changes from ISCs to enterocytes

To profile the global gene expression changes during cell differentiation, we collected purified *Lgr5*$^{+}$ ISCs from *Lgr5-EGFP-ires-CreERT2* mice by flow cytometry and compared them to a pure population of enterocytes, collected from the villus fraction

of tamoxifen-induced *Atoh1* knockout mice, which lack all secretory cells. We collected samples from 2 mice for each cell type and performed global RNA-seq analysis to quantitatively measure transcript expression during the cell differentiation process. To assess the results, we plotted the average fragments per kilobase per million mapped reads (FPKM) levels for each transcript (**Figure 3.1A**). Notably, enterocyte marker genes, *sucrose isomaltase (Sis), alkaline phosphatase (Alpi/Akp3), lactase (Lct),* and *fatty acid binding protein 1 (Fabp1)* were highly expressed and enriched in enterocytes, while ISC markers *Lgr5, Ascl2, Notch1, Axin2, Myc* and *Olfm4* were enriched in ISCs (**Figure 3.1A**). We find markers of other ISC populations expressed highly in our $Lgr5^{+}$ ISC and enterocyte data including *Bmi1, Lrig1, Prom1, Hopx* and *DCAMKL-1* (**Figure 3.1B**), consistent with recent reports (Itzkovitz et al., 2012; Munoz et al., 2012). Gene ontology analysis reveals genes highly enriched in ISCs (fold change > 4) are involved in cell cycle regulation, DNA replication and the microtubule cytoskeleton, while those enriched in enterocytes (fold change > 4) are primarily involved in metabolic processes, oxidation-reduction, and transport (**Figure 3.1C**). As these cell populations have been relatively well studied, the observation of many expected categories highlight the reliability of the RNA-seq assay for defining cell type specific gene modules. Through this analysis we also identified genes that are expressed at high levels in both cell populations (grey dots in **Figure 3.1A**). One such gene was *Cdx2*, an intestinal-specific homeobox gene we previously implicated in control of expression of genes in terminally differentiated enterocytes, but its role has not been well-characterized in ISCs.

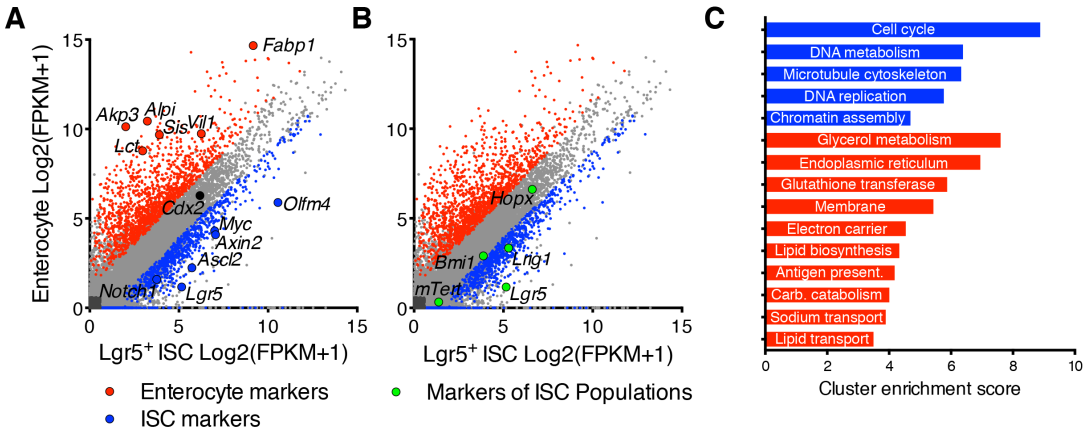


Figure 3.1: RNA-seq profiling of ISCs and enterocytes reveals gene expression changes. (A,B) RNA-seq data from isolated *Lgr5*$^{+}$ ISCs and purified enterocytes from *Atoh1*$^{-/-}$ villi reveals gene expression changes during the differentiation axis. Each point represents one gene. Genes with FPKM values <0.5 in both samples were considered not expressed (10,667 genes; dark grey). 9,749 genes had FPKM values >0.5 in at least one cell type, however were not differentially expressed (light grey). 1,639 genes were enriched in enterocytes (red), while 1,157 were enriched in ISCs (blue) with Q<0.05 and fold-change > 2. **(A)** To validate the RNA-seq results, validated markers of enterocytes and ISCs are superimposed with larger dots. **(B)** With this data we also observed expression of marker genes of other ISC populations described in the literature (large green dots). **(C)** We performed a gene ontology analysis using DAVID software to profile the categories of genes enriched in ISCs (blue) and enterocytes (red) using stringent cutoffs: Q<0.05, fold change > 4.

Loss of CDX2 impairs proliferation and differentiation of ISCs

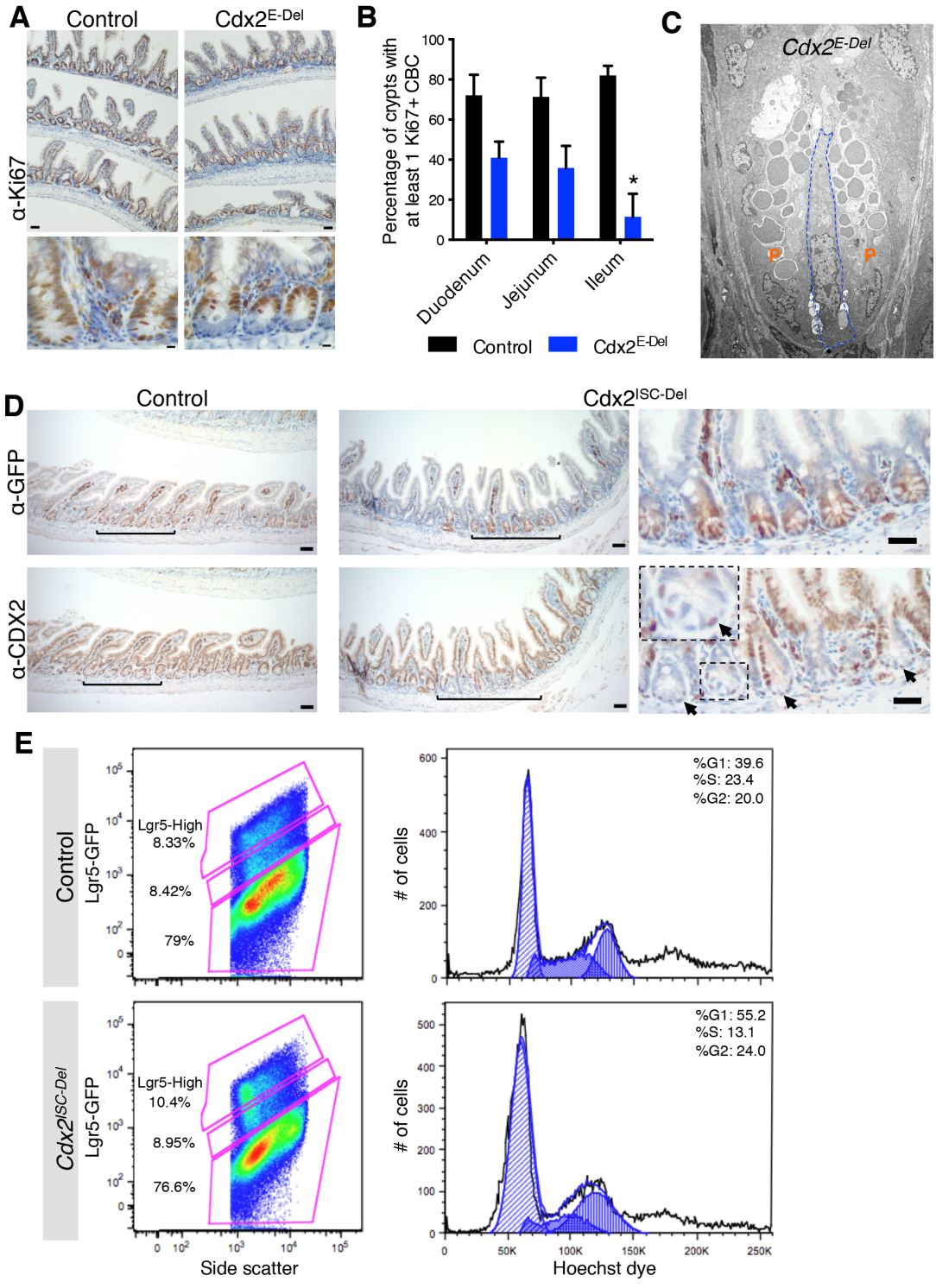
To investigate the consequences of *Cdx2* loss for intestinal stem cells, we used conditional knockout mice in which *Cdx2* was deleted throughout the intestinal epithelium upon tamoxifen injection (referred to here as $Cdx2^{E\text{-}DEL}$ for epithelium deleted; $Cdx2^{Fl/Fl}$;$Villin\text{-}Cre^{ERT2}$). Proliferation was assessed by immunohistochemistry (IHC) and quantification of CBCs marked with KI67. Compared to controls, (mice without $Villin\text{-}Cre^{ERT2}$), the number of crypts with at least one $KI67^{+}$ CBC was reduced in $Cdx2^{E\text{-}Del}$ mice, with the most significant results in the distal 1/3 of the small intestine (ileum), where CDX2 levels are highest in wild type mice (**Figures 3.2A,B**). This defect was highly restricted to the stem cell (CBC) zone, as transient amplifying (TA) cells higher in the crypt retain much proliferation (**Figure 3.2A**). The loss of KI67+ cells in the CBC zone was not due to absence of the CBCs themselves, as their nuclei remain visible by histology and CBCs can be identified by electron microscopy (**Figure 3.2C**).

To visualize and isolate $Cdx2^{-/-}$ ISCs for analysis, we crossed $Cdx2^{Fl/Fl}$ mice with $Lgr5\text{-}EGFP\text{-}ires\text{-}Cre^{ERT2}$ mice. After 4 days of tamoxifen injection, loss of CDX2 was observed in clusters of *Lgr5-GFP* positive crypts (**Figure 3.2D**; these mice are referred to as $Cdx2^{ISC\text{-}Del}$). Notably, as Paneth cells turn over slowly, they retained CDX2 expression at this time point (**Figure 3.2D**), meaning phenotypes observed in ISCs were likely cell autonomous. To further investigate the proliferation defect, we analyzed the DNA content in Lgr5-GFP^{High} cells. Despite the ileum having the most significant proliferation defect, we chose to use the proximal half of the small intestine for this experiment because the *Lgr5-knockin* allele is mosaic and there is a much higher proportion of crypts expressing the allele proximally (Barker et al., 2007). With this caveat in mind, we

profiled the cell cycle phase distribution in Lgr5-GFP^{High} cells based on the intensity of Hoechst dye staining and we detected a decrease in the percentage of cells in S-phase, with an increase in cells in G1 (**Figure 3.2E**). Together, our immunohistochemistry and flow cytometry experiments point to a cell-autonomous requirement for CDX2 in controlling ISC proliferation.

Figure 3.2. CDX2 is required for ISC proliferation. **(A)** Immunohistochemistry (IHC) for KI67 in the ileum of control and $Cdx2^{E\text{-}Del}$ mice reveals specific loss of proliferation in CBC in most crypts, while TA-zone proliferation is retained. Top, scale bars 50 um; bottom scale bars 10 um. **(B)** Quantification of percentage of crypts with a KI67+ CBC throughout the intestine shows the defect is most severe in the ileum. Significance was assessed by t-test; *, $P<0.05$. **(C)** Electron microscopy on $Cdx2^{E\text{-}Del}$ ileum confirms the presence of CBCs. **(D)** Four days after TAM injection, ISCs in Lgr5-GFP-Cre+ patches of crypts (brackets) in $Cdx2^{ISC\text{-}Del}$ mice are negative for CDX2, however Paneth cells in these crypts retain expression (arrows; inset shows magnified view of one crypt base). Scale bars, 50 μm. **(E)** *Left,* flow cytometry analysis from the proximal ½ of the small intestine reveals no difference in the percentage of Lgr5-GFP^{High} cells between control and $Cdx2^{ISC\text{-}Del}$ mice. *Right,* histograms of DNA content (via Hoechst dye staining) are analyzed using the Watson Pragmatic Test to identify the percentage of cells in each phase of the cell cycle. Notably, $Cdx2^{-/-}$ ISCs have a reduction in cells in S-phase and an increase in cells in G1.

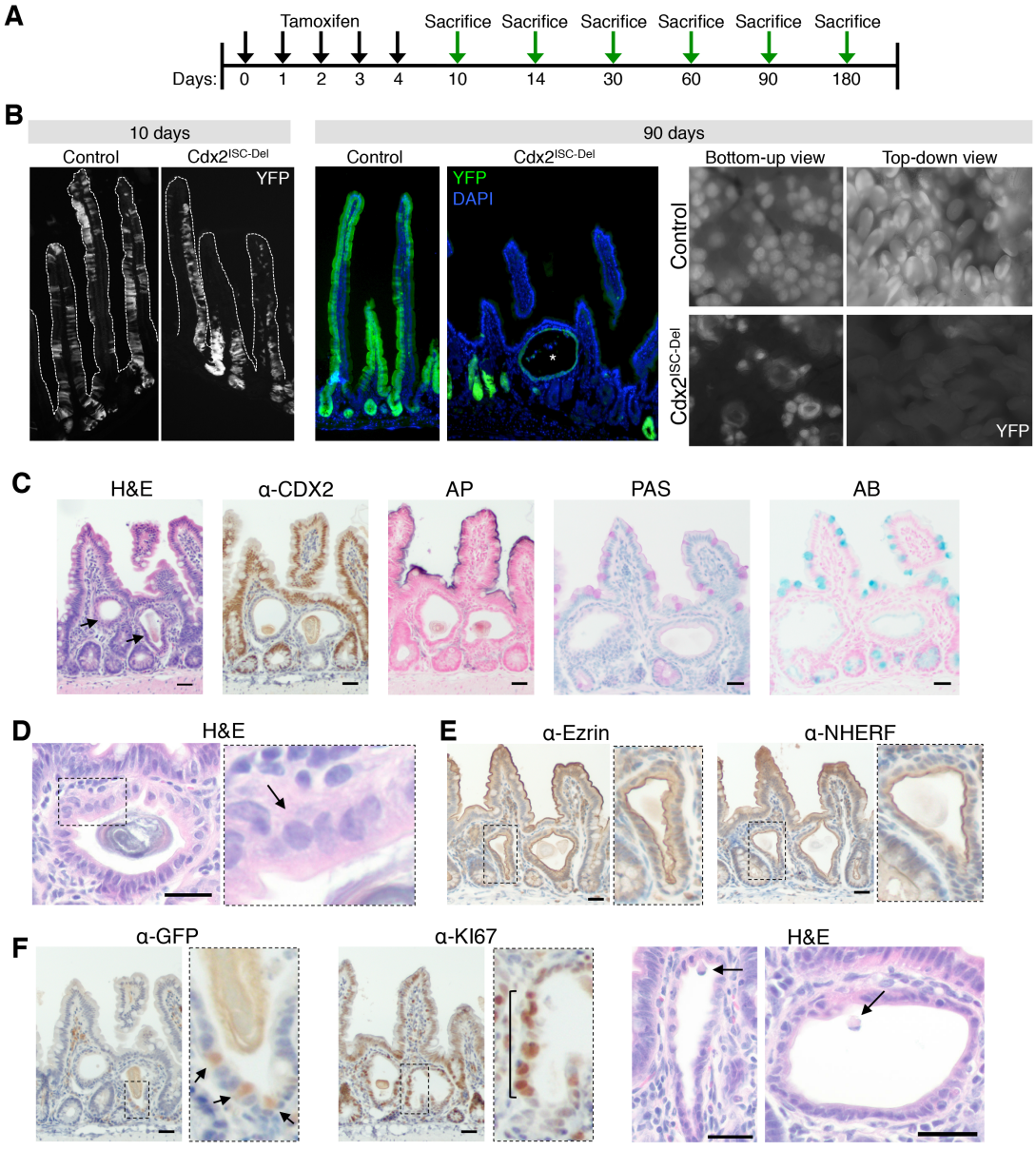
Figure 3.2 (Continued)



To understand whether loss of CDX2 affects cell differentiation from ISCs we bred the *Rosa26-lox-STOP-lox-YFP* Cre reporter allele [referred to here as $R26R^{YFP}$; (Srinivas et al., 2001)] into our $Cdx2^{Fl/Fl}$;*Lgr5-EGFP-ires-CreERT2* line. We injected mice with TAM for 5 days and observed YFP tracing at several intervals afterwards (**Figure 3.3A**). At 10 days post injection, tracing was observed throughout the crypts and villi in both control ($Cdx2^{+/+}$;*Lgr5-EGFP-ires-*Cre^{ERT2};$R26R^{YFP}$) and, although less frequently, in $Cdx2^{ISC\text{-}Del}$ mice (**Figure 3.3B**). However, beginning at one month and up to six months after injection, tracing was never observed in the villi; instead traced cysts and some crypts remained (**Figure 3.3B**). This experiment expands on previously published results describing similar cystic structures by providing a cell lineage for these cysts; all the cells descended from a *Cdx2*-null ISC. The cysts are completely lacking CDX2 and do not contain markers of mature villus cells including alkaline phosphatase, which marks absorptive enterocytes, periodic acid-Schiff (PAS) or Alcian blue stain, which mark goblet cells (**Figure 3.3C**). Columnar cell shape is not retained in the cells lining the cysts; nuclei are circular and cells more cuboidal (**Figure 3.3D**). However, cells lining cysts retain apico-basal polarity, as shown by apical staining of Ezrin and NHERF (**Figure 3.3E**). Interestingly, the cysts retain some qualities of intestinal crypts: Lgr5-GFP staining can be observed in the base of cysts, proliferation is observed along the sides in a transit amplifying-like zone, and cells slough off the top of cysts into the lumen, perhaps contributing to the observed debris in some cysts (**Figure 3.3F**).

Figure 3.3. CDX2 is required for differentiation of ISCs into mature cell types. (A) Schematic of lineage tracing experiment. *Cdx2*$^{Fl/Fl}$*;Lgr5-EGFP-Cre*^{ERT2}*;R26R*^{YFP} and control mice were injected with TAM for 5 days, and sacrificed at the indicated times for analysis of lineage tracing. **(B)** Native YFP fluorescence at 10 days indicates that traced *Cdx2*$^{-/-}$ crypts continue to contribute to villi (marked with dotted lines), as in controls. In contrast, at 90 days tracing is confined to cysts in *Cdx2*$^{ISC\text{-}Del}$ mice. Observation of whole tissues from the outside of the intestine through the muscle layers (bottom-up) reveals halo-like structures corresponding to the *Cdx2*$^{-/-}$ cysts. A view from the lumen of the intestine looking down on the villi (top-down) normally reveals ribbons of YFP-labeled cells, which are never observed in *Cdx2*$^{ISC\text{-}Del}$ mice. **(C)** H&E of the cysts reveals accumulation of debris in the lumen. Antibody staining for CDX2 confirms that traced cysts are *Cdx2*$^{-/-}$, and are surrounded by CDX2-positive cells in crypts that did not express the *Lgr5-EGFP-Cre*^{ERT2} allele. Staining for alkaline phosphatase and periodic acid-Schiff (PAS) or Alcian blue (AB) reveals that cysts lack mature enterocytes and goblet cells, respectively. **(D)** Close-up of a cyst reveals changes in cell morphology within the cells lining the cyst, compared to columnar cells of the normal epithelium (area in box is magnified to the right; arrow highlights round nuclear shape). **(E)** IHC for Ezrin and NHERF reveal normal apical staining in the cysts (on the lumen side), suggesting that these cells retain their apico-basal polarity. **(F)** Cysts also retain information on the differentiation axis, as cells can be detected with IHC for Lgr5-GFP towards the base (arrows), KI67 on the sides (bracket), and cells are observed being extruded from the top of the cyst into the lumen (arrows). Areas in boxes are magnified to the right of the image. Scale bars (**C,E,F**: GFP/KI67), 30 μm; (**D,F**: H&E), 50 μm.

Figure 3.3 (Continued)



Another property of *Lgr5*$^{+}$ ISCs is that they can form organoids when isolated and cultured in a defined media. To test whether loss of *Cdx2* affects organoid formation, we isolated intestinal crypts from *Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} and control (no *Villin-Cre*^{ERT2}) mice and exposed them to 4-OH-Tamoxifen (4-OHT) in culture from 0-2 days (**Figure 3.4A**). After 8 days of culture, organoids from control mice and grown in normal media contained large organoids with many protruding crypts (**Figure 3.4B**). *Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} crypts exposed to 4-OHT from 0-2 days formed either none or a few organoids, which were small, round, contained few crypt protrusions, and completely lacked CDX2 (**Figures 3.4B-C**). Together, the lack of proliferation, lineage tracing and differentiation into mature intestinal cell types, and organoid forming ability describe a role for CDX2 in controlling critical ISC functions, which is unique from its role in mature cell types.

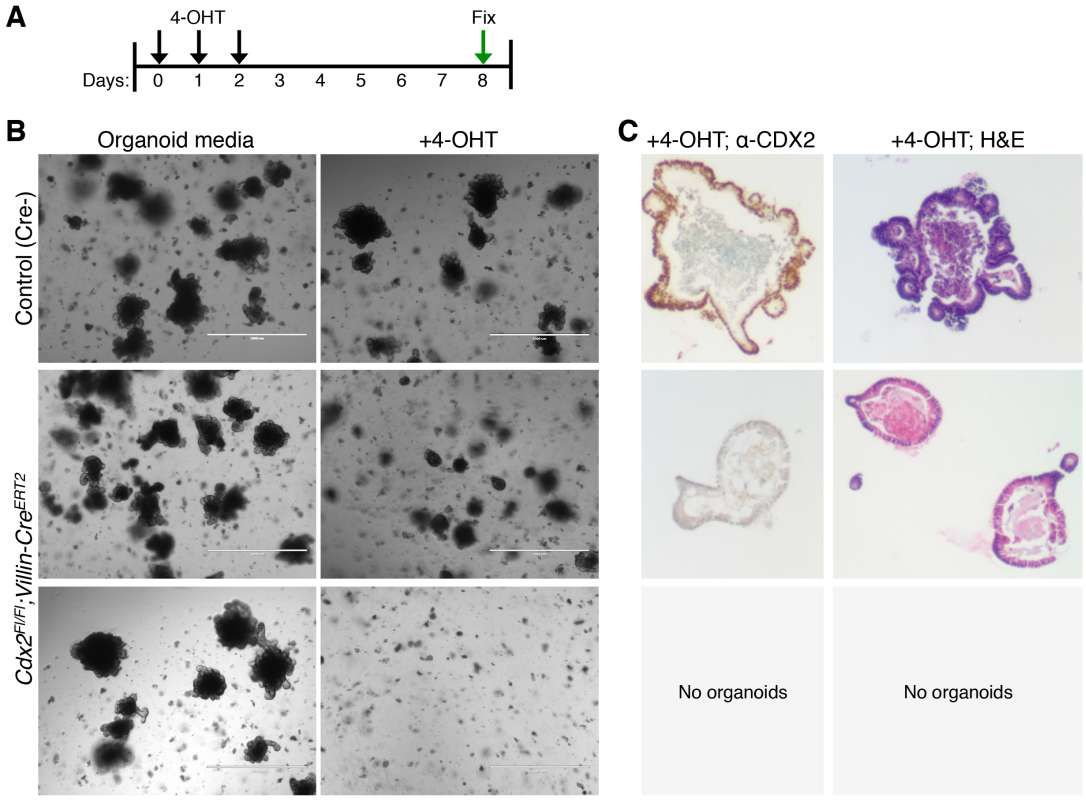


Figure 3.4. CDX2 is required for organoid formation from crypts. (A) Experimental scheme for organoid culture assay. Control and *Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} crypts were isolated from the proximal half of the small intestine and seeded in Matrigel for culture. Organoid growth media was added and 4-OHT was supplemented in some wells from 0-2 days. **(B)** On day 8, organoids from two different *Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} mice were comparable to controls without 4-OHT. However, in wells supplemented with 4-OHT, organoids were either very small (middle) or non-existent (bottom). Scale bars, 1 mm. **(C)** IHC for CDX2 and H&E on organoids supplemented with 4-OHT shows that *Cdx2*$^{-/-}$ organoids form few crypts and have a circular shape.

CDX2 loss in ISCs affects genes regulating proliferation

As our data suggests that CDX2 may play a unique role in ISCs, we wanted to follow up on the mechanism by which it may have these effects. Since CDX2 has been described as a transcriptional activator in mature villus cells (Verzi et al., 2013), we decided to investigate which genes are dysregulated upon CDX2 knockout in ISCs. We collected ISCs from control (*Cdx2*$^{+/+}$*;Lgr5-EGFP-ires-Cre*^{ERT2}) and *Cdx2*$^{\text{ISC-Del}}$ (*Cdx2*$^{Fl/Fl}$*;Lgr5-EGFP-ires-Cre*^{ERT2}) mice by flow cytometry after 4 days of tamoxifen injection and isolated mRNA for sequencing (**Figure 3.5A**). We confirmed our knockout by observing the aligned reads to *Cdx2* exon 2, which is removed via recombination, although the rest of the mRNA is produced (**Figure 3.5B**). We observed an approximately equal distribution of genes upregulated and down-regulated at least 1.5 fold by the loss of *Cdx2* (448 vs. 464, respectively; $P<0.05$; **Figure 3.5C**). Genes down-regulated upon *Cdx2* loss were enriched for terms related to mitosis, while up-regulated genes varied more in functions, resulting in low cluster enrichment scores (**Figure 3.5D**).

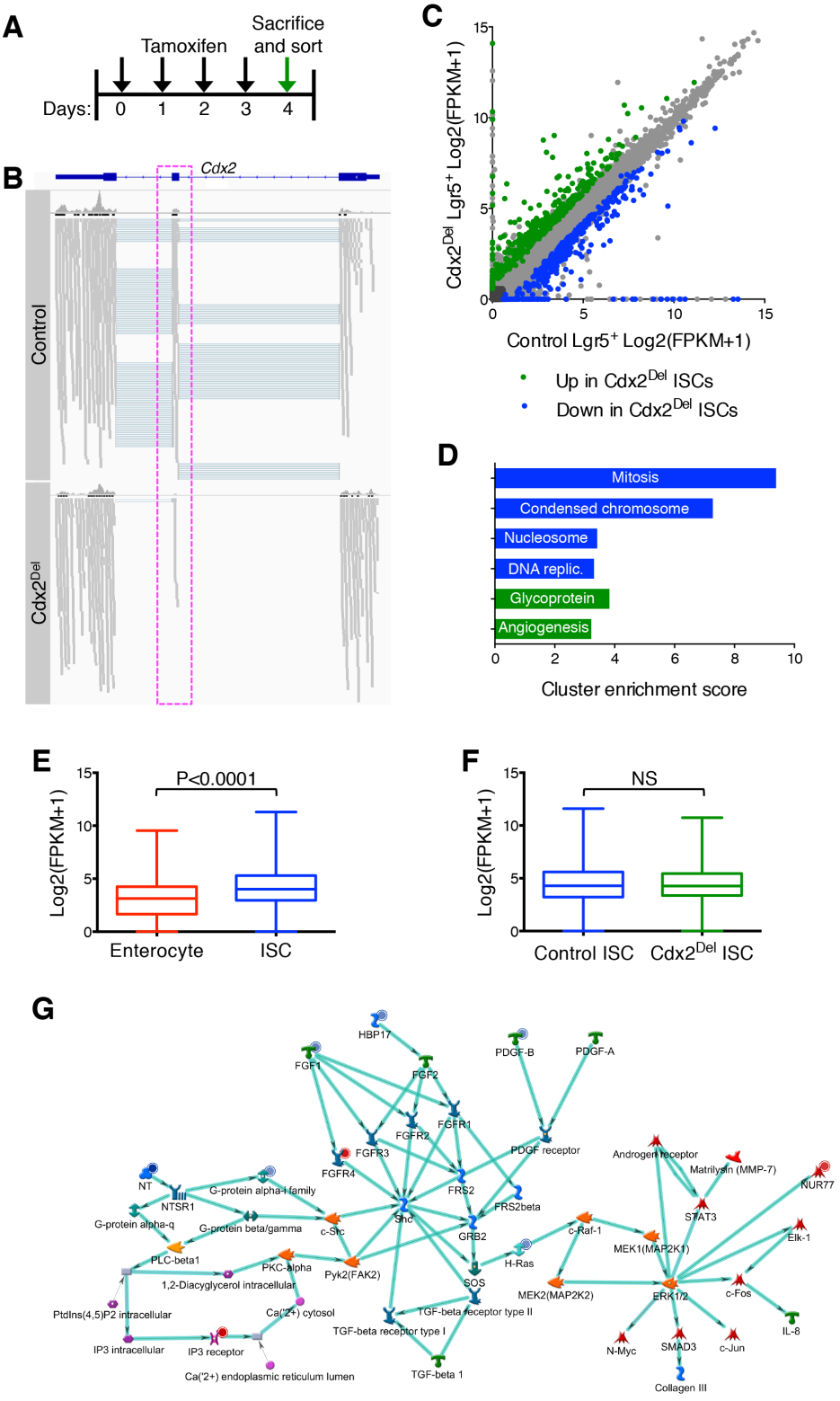
As the Wnt signaling pathway is related to proliferation in ISCs, we decided to analyze specifically whether any of the Wnt target genes are affected by *Cdx2* loss. We imagined that this could occur either through direct regulation of Wnt target genes by CDX2, as has been suggested in cell lines (Verzi et al., 2010a) or through the control of upstream members of the pathway. We used a curated list of 202 intestinal Wnt target genes to assess this pathway (van de Wetering et al., 2002). When we compare the average FPKM values for the 202 Wnt target genes in enterocytes and ISCs, we see that these genes are expressed at significantly higher levels in ISCs (**Figure 3.5E**). The same analysis on the *Cdx2*^{Del} ISCs versus control showed no significant difference (**Figure**

3.5F) and we found that only 10 Wnt target genes were significantly perturbed upon *Cdx2* loss.

To analyze what cell signaling pathways might be affected by *Cdx2* loss and follow up on the mechanism by which it controls ISC functions, we generated gene networks based upon down-regulated genes using Metacore software. The most significant (P=7.86x10^{-8}) down-regulated network assembled was the cellular response to fibroblast growth factor stimulus (**Figure 3.5G**), in which several genes were significantly dysregulated in our data. Although this is a promising pathway for follow-up, we cannot say from expression data alone that the dysregulated genes were direct targets of CDX2. To accomplish this task, we needed to find CDX2 binding sites.

Figure 3.5. Loss of *Cdx2* perturbs hundreds of genes related to cell proliferation, perhaps through the FGF signaling pathway. (A) For collection of control and *Cdx2*$^{-/-}$ ISCs, mice were injected for 4 days and cells were collected the following day by flow cytometry. **(B)** Analysis of aligned reads in the isolated cells confirms the excision of *Cdx2* exon 2 (magenta box) in isolated *Cdx2*$^{-/-}$ Lgr5-GFP^{High} cells. Grey bars represent single aligned reads. Teal lines indicate reads spanning two exons. **(C)** RNA-seq data from control and *Cdx2*^{Del} ISCs. 448 genes go up in *Cdx2*$^{-/-}$ ISCs (green), while 464 go down (blue). Genes that did not meet the cutoffs (P<0.05, fold-change > 1.5) are indicated in grey. **(D)** Gene ontology clusters from down-regulated (blue) and up-regulated (green) genes in *Cdx2*^{Del} ISCs indicate that CDX2 controls genes involved in proliferation. **(E)** Comparison of average FPKM values from 202 Wnt target genes in enterocytes (red) and ISCs (blue) indicate a significant increase in ISCs (P<0.0001). **(F)** The average FPKM values for the Wnt target genes are not affected in the *Cdx2*^{Del} ISCs. **(G)** Metacore software identified the fibroblast growth factor pathway due to enrichment of genes down regulated in *Cdx2*^{Del} ISCs. Blue circles next to genes indicate significant down-regulation, while red circles indicate up-regulation.

Figure 3.5 (Continued)



Identification of CDX2 binding sites target genes in ISCs by combining ChIP-seq and RNA-seq data

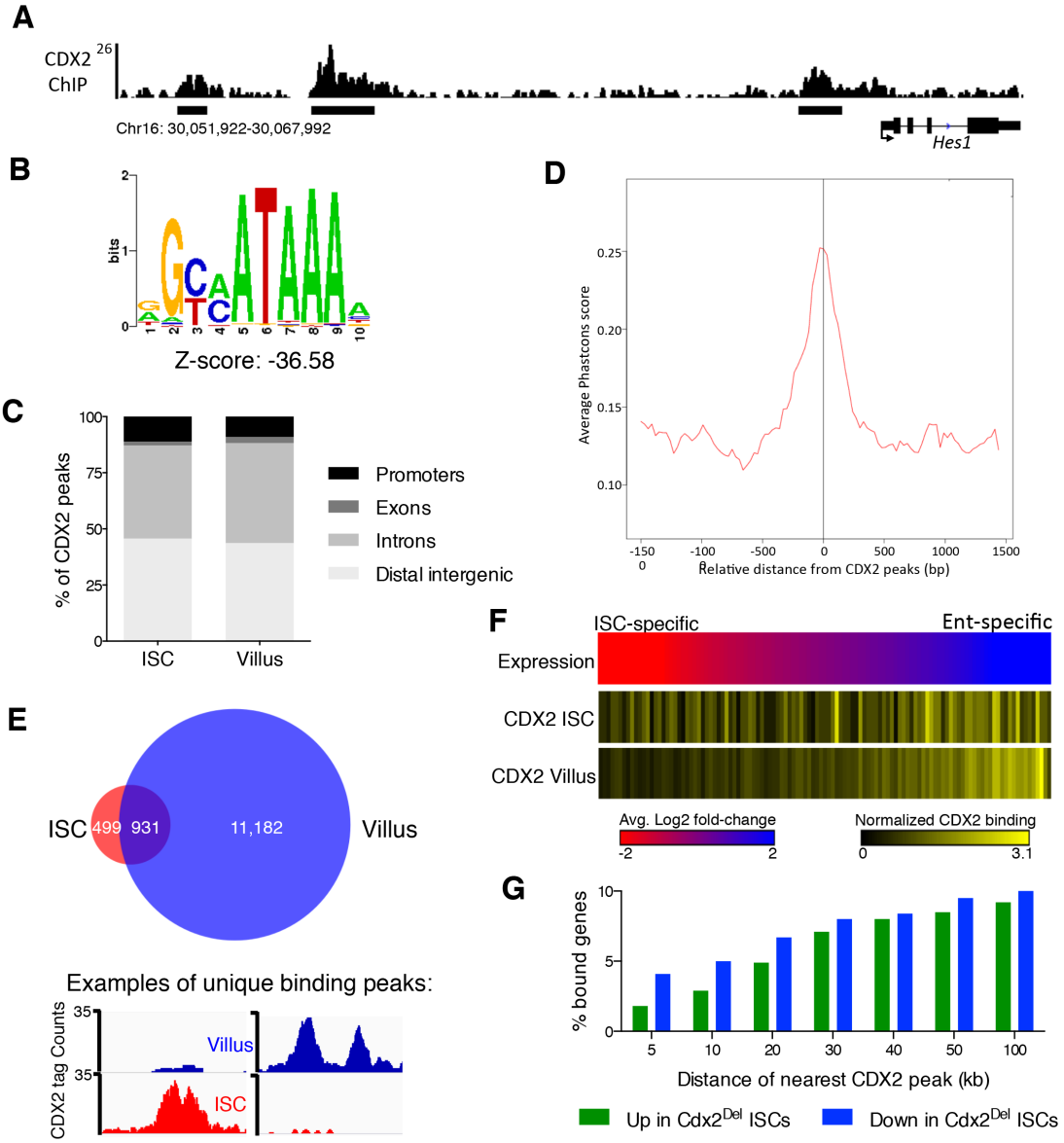
To understand where CDX2 is binding in the ISC population, we performed CDX2 ChIP-Seq on 5 million isolated control (*Cdx2*$^{+/+}$*;Lgr5-EGFP-ires-Cre*^{ERT2}) ISCs. We used the MACS algorithm to call peaks, which identified 1,442 high-confidence regions enriched for CDX2 binding (example data in **Figure 3.6A**). We had previously identified 12,121 CDX2 peaks in differentiated villus cells. In a previous study using the Caco2 cell line to model differentiation, more CDX2 binding sites were also identified in the differentiated cells, making the small number in ISCs unlikely to be due to technical reasons (Verzi et al., 2010b). To validate the peaks, we searched for enriched motifs around peak summits, which showed CDX2 to be the top hit (**Figure 3.6B**). We found that approximately 85% of peaks are within intronic or distal intergenic regions (potential enhancers), while 10% are at the promoter (2 kb upstream and 1 kb downstream from the TSS), which is very similar to what was found in CDX2 ChIP-seq in mouse villus (**Figure 3.6C**). Suggesting functional relevance, the CDX2 peaks are found in evolutionarily conserved regions, as measured by the average PhastCons score (**Figure 3.6D**). Together, these results show that we have generated a robust CDX2 binding profile using a small number of ISCs.

To understand how CDX2 gene regulates cell differentiation, we first looked at overlap between called peaks in ISCs and villus (**Figure 3.6E**). Overall, about one-third of the CDX2 binding sites identified in ISCs were from the villus sites, and two-thirds overlapped by at least 1 base-pair, indicating the possibility of unique regulation in this cell population. To understand the differential binding contributes to functional

regulation of differentiation, we associated peaks with the single nearest gene within 30 kb using GREAT software. To associate the CDX2 binding with gene expression changes during cell differentiation, we binned groups of 100 genes and ordered them from ISC-specific to enterocyte specific. Based on these groups, we found the average number of associated CDX2 binding peaks in ISC and villus and normalized them to the average of all genes. This analysis reveals that CDX2 binding in the villus is associated with genes enriched in enterocytes, while binding in ISCs is not enriched along this axis, but rather is distributed (**Figure 3.6F**). This analysis suggests that CDX2 binds genes in ISCs that will be expressed in the future enterocytes. To test whether this binding represents repression of these genes in ISCs, we associated the CDX2 binding data with the control and *Cdx2*^{Del} RNA-seq data. We find that unlike in villi where CDX2 is mainly an activator (Verzi et al., 2013), CDX2 equally binds a percentage of genes that goes up and down upon its loss (**Figure 3.6G**).

Figure 3.6. CDX2 ChIP-Seq defines differential occupancy in ISCs and villi. (A) ChIP-seq signal near the *Hes1* gene; the Y-axis represents aligned ChIP-seq tag counts. 3 regions of identified CDX2 peaks are indicated below (black bars). **(B)** The CDX2 motif was the top-scoring in an unbiased search within CDX2 ISC peaks, with a very significant Z-score of -36.58. **(C)** Association of CDX2 binding sites to UCSC annotated promoters, exons, introns, and distal intergenic regions reveals a similar distribution for CDX2 peaks within ISCs and villi. Most of the binding occurs in intronic or distal intergenic regions, which represent putative enhancer elements. **(D)** Plot showing average PhastCons score around CDX2 ISC peaks indicates that these regions are evolutionarily conserved. **(E)** Venn diagram of CDX2 binding sites in ISC and villus show that 1/3 of the ISC sites are unique. Below, representative unique binding sites are represented. Y-axis represents aligned ChIP-seq tag counts. **(F)** Genes were ordered based upon fold change comparing enterocyte to ISC expression. Genes were then binned into groups of 100. The average number of CDX2 binding sites within 30kb of all the genes was calculated for each group. Binding data was normalized in each sample by dividing by the average CDX2 signal across all genes. The data show that in the villus, CDX2 peaks cluster near genes specifically expressed in enterocytes, however the reverse is not observed for ISCs, which have a more broad distribution. **(G)** CDX2 binding near genes within the indicated distances was combined with gene expression changes upon loss of *Cdx2* in ISCs. This data shows that a similar percentage of genes bound within the 30-100kb window are significantly up and down regulated.

Figure 3.6 (continued)



CDX2 may control ISC proliferation through the FGF pathway

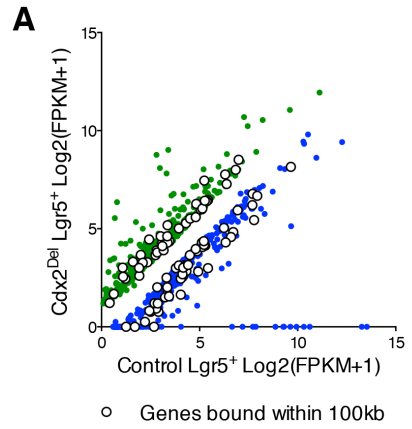
To better understand what genes CDX2 may regulate to mediate its control of ISC functions, we focused on putative direct target genes: those significantly dysregulated upon CDX2 loss and also bound by CDX2 within 100 kb. There were 41 up-regulated and 47 down-regulated genes that fit these criteria (**Figure 3.7B**). Several of the genes on this list may play a role in ISC regulation, however we decided to focus on *Fgfbp1 (Hbp17)* (**Figure 3.7C**) based upon the Metacore analysis that identified the FGF pathway in the gene expression data (**Figure 3.5G**). *Fgfbp1* encodes a secreted FGF binding protein that activates the FGF signaling pathway, promoting proliferation (Beer et al., 2005; Czubayko et al., 1994; Tassi et al., 2001).

To determine whether the CDX2 could be acting through the FGF pathway, we sought to characterize this pathway in the ISCs. First, we surveyed our RNA-seq data and found that in addition to *Fgfbp1* several members of the FGF pathway were expressed (FPKM>0.5; **Figure 3.7D**), including FGF receptors 2-4, however *Fgfbp1* is the only gene with nearby CDX2 binding (**Figure 3.7C**). All the factors (except FGFBP3) were expressed at higher levels in ISCs, with *Fgf1,* and *Fgfbp1* being significantly enriched (>4 fold) in ISCs, compared to enterocytes (**Figure 3.7D**). Interestingly, upon *Cdx2* knockout there was a slight increase in mRNA for *Fgfbp3*, *Fgfr2*, *Fgfr3* and a significant increase in *Fgfr4* (**Figure 3.7D**). The increase in FGFR3 and FGFR4 was confirmed by Western blot, where an increase in the glycosylated versions of the protein (found on the cell surface) was evident, without a relative increase in phosphorylated receptors (**Figure 3.7D**). This up-regulation may indicate compensation for the lack of FGF signaling due to reduced FGFBP1 and FGF1 levels.

FGFBP1 acts through releasing FGF ligands from interactions with heparan sulfate proteoglycans in the extracellular matrix and making them available to bind their receptors (Wu et al., 1991). Thus, we were interested in determining what ligands could be mediating these effects with FGFBP1. Based upon the receptors expressed in the ISCs (FGFR2, isoform IIIb; FGFR3, isoforms IIIb and IIIc; and FGFR4) and reports from the literature, we were able to narrow down the possible FGF ligands to FGF1, 2, 3, 4, 5, 7, 8, 9, 10, 16, 18, 19, 22 and 23. To test which might be expressed in the intestinal stem cell niche, we extracted RNA from whole intestinal tissue and performed qPCR for each gene. We compared the results to qPCR on RNA extracted from crypt epithelium, where expression should be low based on our RNA-seq results, although other cell types in the crypt, such as Paneth cells may contribute signal. Unfortunately, primers for FGF6 and FGF10 did not work to amplify any transcript, however we did have success with several other of the primer sets. We found that *Fgf1* and *Fgf19* were expressed at significantly lower levels in the whole intestine compared to isolated crypts, suggesting they are relatively epithelium-specific (**Figure 3.7F**). We honed in on *Fgf4* and *Fgf7*, which were both highly expressed in the whole intestine compared to crypts, with *Fgf7* being a very significant result (**Figure 3.7F**). Together, this evidence begins to characterize the involvement of the FGF pathway in ISC proliferation *in vivo*, as a result of CDX2 binding.

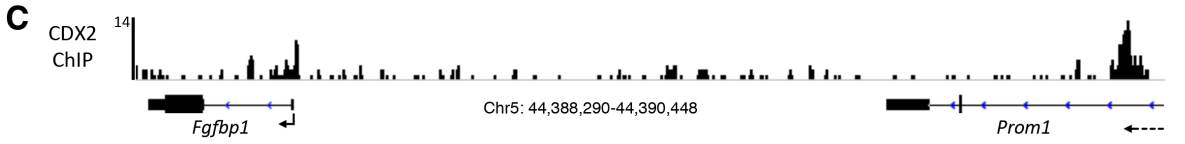
Figure 3.7. Direct target genes of CDX2 in ISCs include FGFBP1, flagging the FGF pathway for further investigation. (A) Significantly dysregulated genes upon *Cdx2* loss in ISCs that contain CDX2 binding sites within 100 kb (superimposed white dots) are good candidate CDX2 target genes in ISCs. **(B)** List of CDX2 target genes in ISCs. **(C)** CDX2 binding signal near *Fgfbp1* in ISCs. There is a significant binding peak in the *Prom1* intron that may regulate *Fgfbp1* expression. Y-axis, sequence tag counts from CDX2 ChIP-seq. **(D)** FGF pathway members expressed in control ISCs (FPKM>0.5) are listed with data on gene expression changes upon *Cdx2* loss in ISCs and along the differentiation axis. Significant results are shaded in blue (down regulated in $Cdx2^{Del}$ ISCs), green (up regulated in $Cdx2^{Del}$ ISCs) or yellow (enriched in ISCs versus enterocytes). **(E)** Western blots indicating an increase in the glycosylated versions (asterisk) of FGFR3 (~105 kDa band) and FGFR4 (~100 kDa band). The top band in both cases (arrow) refers to the phosphorylated (active) receptors. **(F)** Survey of FGF expression by qRT-PCR on RNA isolated from whole intestine (grey bars) and crypts (black bars). Data were normalized to *Gapdh* mRNA and expressed relative to the crypt samples. Significant results (assessed by t-test adjusted for multiple comparisons with an FDR cutoff of 5%) are indicated with an asterisk. *Fgf1, P*=0.006; *Fgf7, P=0.0036; Fgf19, P=0.00016.*

Figure 3.7 (Continued)



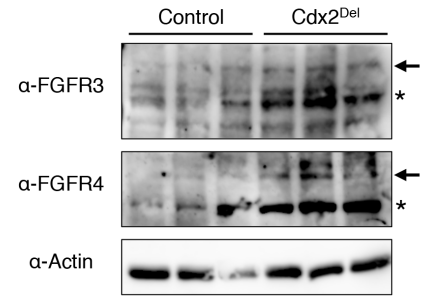
B

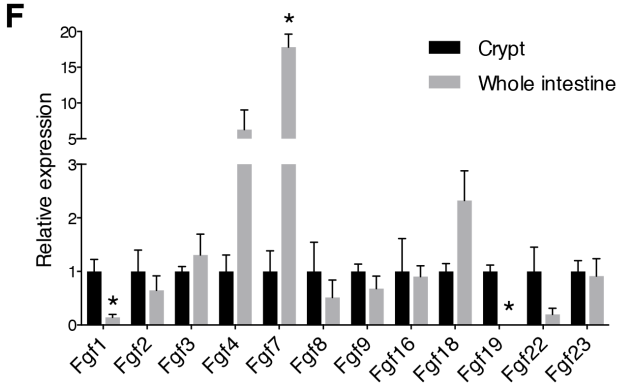
Down regulated targets				Up regulated targets			
1300002K09Rik	*Dach1*	*Kpna2*	*Reg3b*	*AA986860*	*Gfpt1*	*Lrch1*	*Stk38l*
5330417C22Rik	*Defb37*	*Krt20*	*Sema6d*	*Abcc2*	*Gpr55*	*Mansc1*	*Tgm2*
Abcb1a	*Engase*	*Magohb*	*Sis*	*Acnat1*	*Gstm1*	*Mecom*	*Wwc1*
Abp1	*Fam13a*	*Mmd*	*Slc38a1*	*Arid1b*	*Hmgcs2*	*Mecom*	*Zfp395*
Agpat4	*Fam189a2*	*Mtch2*	*Slco2b1*	*Asgr1*	*Hpgd*	*Mex3c*	*Zfpm1*
Bcl6	*Fgfbp1*	*Oat*	*Svip*	*Atf3*	*Hspa1b*	*Mical2*	
Bmp3	*Gcnt4*	*Osbpl3*	*Tifa*	*C4bp*	*Id1*	*Nt5dc3*	
Camk1d	*Glipr2*	*Pde3a*	*Tm4sf20*	*Clmn*	*Id2*	*Pag1*	
Cda	*Gls*	*Pde9a*	*Tnfrsf11a*	*Cpm*	*Itga6*	*Rasgef1b*	
Cdh17	*Greb1*	*Pnlip*	*Tubb2b*	*Cyp2d26*	*Jund*	*Scarb1*	
Cga	*H2afv*	*Pofut1*	*Uaca*	*Foxa2*	*Kif13b*	*Slc2a9*	
Ckap2l	*Isx*	*Rbp2*		*Foxq1*	*Lbh*	*Slc9a2*	



D

Gene	Control vs. Cdx2Del ISC				Enterocyte vs. ISC			CDX2 ISC Binding 100 kb
	Control FPKM	Cdx2Del FPKM	Log2 fold change	P-value	Enterocyte FPKM	Log2 fold change	Q-Value	
Fgf1	23.39	4.92	-2.25	5E-5	1.03	-4.21	0.0013	0
Fgfbp1	79.79	18.78	-2.09	5E-5	14.46	-2.16	0.012	1
Fgfbp3	0.79	2.52	1.67	0.14	1.88	1.55	0.45	0
Fgfr2	6.84	8.09	0.24	0.58	2.83	-0.98	0.24	0
Fgfr3	19.10	23.44	0.30	0.47	6.54	-1.25	0.069	0
Fgfr4	6.35	19.88	1.65	0.0034	0.03	-7.29	0.38	0





Discussion

This work supports a role for CDX2 in controlling a unique gene expression program in ISCs compared to villus. Additionally, through analysis of cell proliferation, differentiation and organoid formation, we conclude that CDX2 is required for normal functions of ISCs. Based upon the genes dysregulated and bound by CDX2, we determined direct target genes of CDX2 leading us to the FGF pathway. We note that several other direct targets of CDX2 also contribute to ISC functions and are worth future examination. One interesting down-regulated gene that is interesting is *Pofut1,* which is required to enable ligand binding to the notch receptor extracellular domain, and has a similar loss of proliferation in CBCs when deleted throughout the intestinal epithelium (Guilmeau et al., 2008).

Nevertheless, the FGF signaling pathway is a good candidate for regulating ISC functions for several reasons. First, there is evidence for FGFs, especially FGF7, having mitogenic effects in intestinal cells during normal homeostasis and during injury (Bajaj-Elliott et al., 1997; Brauchle et al., 1996; Finch et al., 1996). FGF ligands and FGFBP1 are both up-regulated in human colorectal cancers and the *Apc-min* mouse model (Ray et al., 2003; Tassi et al., 2006), and the inhibition of FGFBP1 has anti-tumor effects (Schulze et al., 2011). This is interesting because the intestinal stem cell is hypothesized to be the cell of origin for intestinal cancers.

In the future, we would like to follow up on whether the FGF pathway truly plays a role in the regulation of CDX2 ISCs through several approaches. First, we would like to survey the status of FGF pathway activation in control and $Cdx2^{Del}$ ISCs using phospho receptor-tyrosine kinase (RTK) arrays. The arrays have the added benefit of informing us on several other pathways that might be active in ISCs and potentially disturbed upon loss of *Cdx2*. We would also like to perform the RTK arrays on organoids to understand their activation status and to

assess the feasibility of performing further experiments in vitro. If we see that the FGF pathway is active in the organoids, we would like to treat it with a specific receptor tyrosine kinase inhibitor to abrogate FGF signaling and assess the consequences on organoid growth. In addition, we would like to supplement $Cdx2^{Del}$ organoids with recombinant FGF ligands and FGFBP1 to try to rescue their growth.

More broadly, the results from this study have several implications on our understanding of the contribution of CDX2 to ISC maintenance and cell differentiation. We have shown that CDX2 has unique occupancy in ISCs and that during cell differentiation it transitions to binding at thousands of additional sites. We saw that unlike in the villus, CDX2 binding in ISCs does not correlate well with cell-type specific expression, but rather it is distributed to genes that are expressed throughout differentiation. Another difference between CDX2 binding in ISCs and villus was the observation that CDX2 can act as both an activator and repressor in ISCs. These differences may be explained by CDX2 interacting with specific TFs in different compartments, as has been suggested previously (Verzi et al., 2010b). Alternatively, post-translational modifications or the differential availability of chromatin may also contribute. Further studies using purified ISC populations to profile other TFs and additional chromatin marks will be useful to determine whether these mechanisms play a role in the differential actions of CDX2 during intestinal cell differentiation.

Acknowledgements

We would like to thank the following people for helpful discussions and for sharing technical expertise: Mike Verzi, Tae-Hee Kim, Paloma Cejas, Li-Lun Ho, and members of the Center for Functional Cancer Epigenomics. Thanks to Sylvie Robine for providing the Villin-

Cre^{ERT2} mouse line. Additionally, we would like to thank Alessio Tovaglieri and David Breault for helpful discussions and for performing organoid assays. This work was supported by a National Cancer Institute F31 fellowship to A. K. San Roman.

References

Bajaj-Elliott, M., Breese, E., Poulsom, R., Fairclough, P.D., and MacDonald, T.T. (1997). Keratinocyte growth factor in inflammatory bowel disease. Increased mRNA transcripts in ulcerative colitis compared with Crohn's disease in biopsies and isolated mucosal myofibroblasts. Am J Pathol *151*, 1469-1476.

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

Beer, H.D., Bittner, M., Niklaus, G., Munding, C., Max, N., Goppelt, A., and Werner, S. (2005). The fibroblast growth factor binding protein is a novel interaction partner of FGF-7, FGF-10 and FGF-22 and regulates FGF activity: implications for epithelial repair. Oncogene *24*, 5269-5277.

Bell, K.N., and Shroyer, N.F. (2014). Krupple-Like Factor 5 Is Required for Proper Maintenance of Adult Intestinal Crypt Cellular Proliferation. Dig Dis Sci.

Brauchle, M., Madlener, M., Wagner, A.D., Angermeyer, K., Lauer, U., Hofschneider, P.H., Gregor, M., and Werner, S. (1996). Keratinocyte growth factor is highly overexpressed in inflammatory bowel disease. Am J Pathol *149*, 521-529.

Czubayko, F., Smith, R.V., Chung, H.C., and Wellstein, A. (1994). Tumor growth and angiogenesis induced by a secreted binding protein for fibroblast growth factors. J Biol Chem *269*, 28243-28248.

el Marjou, F., Janssen, K.P., Chang, B.H., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186-193.

Finch, P.W., Pricolo, V., Wu, A., and Finkelstein, S.D. (1996). Increased expression of keratinocyte growth factor messenger RNA associated with inflammatory bowel disease. Gastroenterology *110*, 441-451.

Guilmeau, S., Flandez, M., Bancroft, L., Sellers, R.S., Tear, B., Stanley, P., and Augenlicht, L.H. (2008). Intestinal deletion of Pofut1 in the mouse inactivates notch signaling and causes enterocolitis. Gastroenterology *135*, 849-860, 860 e841-846.

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

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2011). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat Cell Biol.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2012). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nature cell biology *14*, 106-114.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y*., et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Montgomery, R.K., Carlone, D.L., Richmond, C.A., Farilla, L., Kranendonk, M.E.G., Henderson, D.E., Baffour-Awuah, N.Y., Ambruzs, D.M., Fogli, L.K., Algra, S*., et al.* (2011). Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. Proc Natl Acad Sci U S A *108*, 179-184.

Muñoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.-K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S*., et al.* (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. The EMBO journal.

Munoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S*., et al.* (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. The EMBO journal *31*, 3079-3091.

Nakaya, T., Ogawa, S., Manabe, I., Tanaka, M., Sanada, M., Sato, T., Taketo, M.M., Nakao, K., Clevers, H., Fukayama, M*., et al.* (2014). KLF5 regulates the integrity and oncogenicity of intestinal stem cells. Cancer Res *74*, 2882-2891.

Peregrina, K., Houston, M., Daroqui, C., Dhima, E., Sellers, R.S., and Augenlicht, L.H. (2014). Vitamin D is a determinant of mouse intestinal Lgr5 stem cell functions. Carcinogenesis.

Perekatt, A.O., Valdez, M.J., Davila, M., Hoffman, A., Bonder, E.M., Gao, N., and Verzi, M.P. (2014). YY1 is indispensable for Lgr5+ intestinal stem cell renewal. Proc Natl Acad Sci U S A *111*, 7695-7700.

Powell, A.E., Wang, Y., Li, Y., Poulin, E.J., Means, A.L., Washington, M.K., Higginbotham, J.N., Juchheim, A., Prasad, N., Levy, S.E.*, et al.* (2012). The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. Cell *149*, 146-158.

Ray, R., Cabal-Manzano, R., Moser, A.R., Waldman, T., Zipper, L.M., Aigner, A., Byers, S.W., Riegel, A.T., and Wellstein, A. (2003). Up-regulation of fibroblast growth factor-binding protein, by beta-catenin during colon carcinogenesis. Cancer Res *63*, 8085-8089.

Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet *40*, 915-920.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J.*, et al.* (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262-265.

Schulze, D., Plohmann, P., Hobel, S., and Aigner, A. (2011). Anti-tumor effects of fibroblast growth factor-binding protein (FGF-BP) knockdown in colon carcinoma. Mol Cancer *10*, 144.

Sheaffer, K.L., Kim, R., Aoki, R., Elliott, E.N., Schug, J., Burger, L., Schubeler, D., and Kaestner, K.H. (2014). DNA methylation is required for the control of stem cell differentiation in the small intestine. Genes Dev *28*, 652-664.

Srinivas, S., Watanabe, T., Lin, C.S., William, C.M., Tanabe, Y., Jessell, T.M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC developmental biology *1*, 4.

Stringer, E.J., Duluc, I., Saandi, T., Davidson, I., Bialecka, M., Sato, T., Barker, N., Clevers, H., Pritchard, C.A., Winton, D.J.*, et al.* (2012). Cdx2 determines the fate of postnatal intestinal endoderm. Development *139*, 465-474.

Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion Between Intestinal Stem Cell Populations in Distinct Niches. Science (New York, NY) *334*, 1420-1424.

Tassi, E., Henke, R.T., Bowden, E.T., Swift, M.R., Kodack, D.P., Kuo, A.H., Maitra, A., and Wellstein, A. (2006). Expression of a fibroblast growth factor-binding protein during the development of adenocarcinoma of the pancreas and colon. Cancer Res *66*, 1191-1198.

Tian, H., Biehs, B., Warming, S., Leong, K.G., Rangell, L., Klein, O.D., and de Sauvage, F.J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. Nature.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols *7*, 562-578.

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A.P.*, et al.* (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241-250.

van der Flier, L.G.v.d., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Beghtel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.

Verzi, M.P., Hatzis, P., Sulahian, R., Philips, J., Schuijers, J., Shin, H., Freed, E., Lynch, J.P., Dang, D.T., Brown, M.*, et al.* (2010a). TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. Proc Natl Acad Sci U S A *107*, 15157-15162.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010b). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell *19*, 713-726.

Verzi, M.P., Shin, H., San Roman, A.K., Liu, X.S., and Shivdasani, R.A. (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol Cell Biol *33*, 281-292.

Wu, D.Q., Kan, M.K., Sato, G.H., Okamoto, T., and Sato, J.D. (1991). Characterization and molecular cloning of a putative binding protein for heparin-binding growth factors. J Biol Chem *266*, 16778-16785.

Yan, K.S., Chia, L.A., Li, X., Ootani, A., Su, J., Lee, J.Y., Su, N., Luo, Y., Heilshorn, S.C., Amieva, M.R.*, et al.* (2012). The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. Proc Natl Acad Sci U S A *109*, 466-471.

Yui, S., Nakamura, T., Sato, T., Nemoto, Y., Mizutani, T., Zheng, X., Ichinose, S., Nagaishi, T., Okamoto, R., Tsuchiya, K.*, et al.* (2012). Functional engraftment of colon epithelium expanded in vitro from a single adult Lgr5(+) stem cell. Nat Med *18*, 618-623.

Zhang, N., Yantiss, R.K., Nam, H.S., Chin, Y., Zhou, X.K., Scherl, E.J., Bosworth, B.P., Subbaramaiah, K., Dannenberg, A.J., and Benezra, R. (2014). ID1 Is a Functional Marker for Intestinal Stem and Progenitor Cells Required for Normal Response to Injury. Stem cell reports *3*, 716-724.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Zheng, H.Q., Zhou, Z., Huang, J., Chaudhury, L., Dong, J.T., and Chen, C. (2009). Kruppel-like factor 5 promotes breast cell proliferation partially through upregulating the transcription of fibroblast growth factor binding protein 1. Oncogene *28*, 3702-3713.

Chapter 4: Transcription factors GATA4 and HNF4A control distinct aspects of intestinal homeostasis in conjunction with the transcription factor CDX2

Adrianna K. San Roman, Boaz E. Aronson, Stephen D. Krasinski,
Ramesh A. Shivdasani, and Michael P. Verzi

This work was published in *The Journal of Biological Chemistry (2015) 290:1850-60.*

Contributions:

A. K. San Roman, B. E. Aronson, S. D. Krasinski, R. A. Shivdasani and M. P. Verzi worked collaboratively to develop the concept and design of the study. A. K. San Roman performed experiments on *Hnf4a/Cdx2* compound mutants, while B. E. Aronson performed the *Gata4/Cdx2* experiments. A. K. San Roman, B. A. Aronson, R. A. Shivdasani and M. P. Verzi drafted the manuscript, with input from S. D. Krasinski. All work was conducted under the direction of R. A. Shivdasani and M. P. Verzi.

Transcription factors GATA4 and HNF4A control distinct aspects of intestinal homeostasis in conjunction with the transcription factor CDX2

Adrianna K. San Roman[‡§], Boaz E. Aronson[¶#], Stephen D. Krasinski[¶], Ramesh A. Shivdasani[‡‖1], and Michael P. Verzi[‡⌘1]

From the [‡]Department of Medical Oncology and Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA; [§]Graduate Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; [¶]Division of Pediatric Gastroenterology and Nutrition, Department of Medicine, Children's Hospital, Boston, MA; [#]Academic Medical Center Amsterdam, Emma Children's Hospital, Amsterdam, the Netherlands; [‖]Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; [⌘]Department of Genetics, Rutgers – The State University of New Jersey, Piscataway, NJ

[1]To whom correspondence may be addressed: Either Ramesh A. Shivdasani, Dana-Farber Cancer Institute, 450 Brookline Ave., Boston, MA, 02215, Tel.: 617-632-5746, Fax: 617-582-7198, E-mail: ramesh_shivdasani@dfci.harvard.edu, or Michael P. Verzi, Rutgers, the State University of New Jersey, 145 Bevier Rd, Piscataway, NJ 08854, Tel.: 848-445-9578, Fax: 732-445-1147, E-mail: verzi@biology.rutgers.edu.

Summary

Distinct groups of transcription factors (TFs) assemble at tissue-specific *cis*-regulatory sites, implying that different TF combinations may control different genes and cellular functions. Within such combinations, TFs that specify or maintain a lineage, and are therefore considered master regulators, may play a key role. Gene enhancers often attract these tissue-restricted TFs as well as TFs that are expressed more broadly. However, the contributions of the individual TFs toward combinatorial regulatory activity have not been examined critically in many cases *in vivo*. We address this question using a genetic approach in mice to inactivate the intestine-specifying and intestine-restricted factor CDX2, alone or in combination with its more broadly expressed partner factors, GATA4 or HNF4A. Compared to single mutants, each combination produced significantly greater defects and rapid lethality, through distinct anomalies. Intestines lacking *Gata4* and *Cdx2* were deficient in crypt cell replication, whereas combined loss of *Hnf4a* and *Cdx2* specifically impaired viability and maturation of villus enterocytes. Integrated analysis of TF binding and of transcripts affected in *Cdx2;Hnf4a* compound mutant intestines indicated that this TF pair controls genes required to construct the apical brush border and absorb nutrients, including dietary lipids. This study thus defines combinatorial TF activities, their specific requirements during tissue homeostasis, and modules of transcriptional targets in intestinal epithelial cells *in vivo*.

Introduction

Tissue-specific gene expression reflects the coordinated activities of transcription factors (TFs) that are restricted to one or a few cell types and others that are expressed more broadly. In tissues of endodermal origin, TFs such as CDX2, PTF1 and PDX1 are restricted to individual

organs (James et al., 1994; Krapp et al., 1996; Ohlsson et al., 1993), whereas others such as Hepatocyte Nuclear Factor 4 Alpha (HNF4A) and GATA are expressed in many endoderm-derived tissues. It is unclear if the latter TFs control separate and distinct programs and cellular function or merely support the activity of master regulators, such as the intestine-restricted CDX2.

In adult mammals, the small intestine contains abundant proliferative cells in sub-mucosal crypts, and mature, post-mitotic cells along the villi. The homeodomain TF CDX2 is expressed exclusively in the epithelium of the small intestine and colon, both in replicating crypt cells and differentiated villus cells (German et al., 1992; James and Kazenwadel, 1991; Jin and Drucker, 1996; Suh et al., 1994). CDX2 is required for proper specification of the intestine during development (Gao et al., 2009) and is considered a master regulator of intestinal identity because its ectopic expression in the stomach or esophagus activates intestine-restricted genes (Liu et al., 2007; Silberg et al., 2002). In adult mice, absence of CDX2 dysregulates genes involved in terminal cell differentiation and fatal malnutrition ensues over ~3 weeks (Hryniuk et al., 2012; Stringer et al., 2012; Verzi et al., 2010). Simultaneous inactivation of its homolog CDX1 resulted in nearly complete arrest of intestinal crypt cell proliferation (Verzi et al., 2011). Together, these findings demonstrate diverse requirements for CDX2 in embryonic tissue specification, cell differentiation, and maintenance of the adult gut epithelium. It is unclear how CDX2 partners with various TFs to mediate these diverse functions.

One way to identify the TFs important in any tissue is through DNA sequence motifs that are highly enriched among active cis-regulatory sites. Enhancers active in the intestinal epithelium show few recurring sequence motifs, including the one preferred by CDX2; the GATA motif, especially in replicating cells; and the consensus motif for HNF4A, mainly in

differentiated villus cells (Verzi et al., 2010). This select group of TFs thus seems particularly important in intestinal gene regulation. Indeed, knockout mice lacking single factors show diverse, subtle, and non-lethal defects. GATA4 and GATA6 show regional (proximal 4/5) and global intestinal expression, respectively, and the corresponding mutant mice have subtle defects in crypt cell replication, secretory cell differentiation, and control of selected enterocyte genes (Beuling et al., 2012; Beuling et al., 2011; Bosse et al., 2006). Loss of HNF4A perturbs colon development (Garrison et al., 2006), but adult *Hnf4a*$^{-/-}$ mouse intestines are overtly normal, with modestly perturbed gene expression (Babeu et al., 2009). Coupled with the frequent co-occurrence of their specific sequence motifs near CDX2 binding sites, the limited defects in *Gata* and *Hnf4a* mutant mice led us to postulate that they may regulate intestinal genes in combination with CDX2. These TFs might act in several different ways: (i) simply support CDX2 activity at CDX2-dependent genes, (ii) partner with CDX2 in distinct combinations to regulate different cellular functions, or (iii) to serve additional, CDX2-independent functions.

To evaluate these possibilities, we generated compound inducible mutant mice that lack *Cdx2* and either *Gata4/6* or *Hnf4a* in the adult intestine. Distinct defects in each compound mutant strain revealed unambiguous joint requirements for CDX2 and GATA4 in crypt cell proliferation and for CDX2 and HNF4A in differentiated villus enterocytes. CDX2 and HNF4A, in particular, co-regulate genes necessary to absorb dietary lipids. Thus, the lineage-restricted factor CDX2 functions in obligate partnerships with different broadly expressed factors to regulate distinct aspects of intestinal epithelial structure and function.

Experimental Procedures

Mice

Gata4$^{Fl/Fl}$, *Gata6*$^{Fl/Fl}$, *Cdx2*$^{Fl/Fl}$, *Hnf4a*$^{Fl/Fl}$, and transgenic *VillinCre*^{ERT2} mice were described previously (el Marjou et al., 2004; Hayhurst et al., 2001; Pu et al., 2004; Sodhi et al., 2006; Verzi et al., 2010). *Gata4*$^{Fl/Fl}$, *Gata6*$^{Fl/Fl}$, *Cdx2*$^{Fl/Fl}$ and *VillinCre*^{ERT2} mice were crossed to generate compound conditional-mutant *Gata4*$^{Fl/Fl}$;*Gata6*$^{Fl/Fl}$;*Cdx2*$^{Fl/Fl}$;*VillinCre*^{ERT2} mice. *Cdx2*$^{Fl/Fl}$, *Hnf4a*$^{Fl/Fl}$, and *VillinCre*^{ERT2} mice were mated to generate *Hnf4a*$^{Fl/Fl}$;*Cdx2*$^{Fl/Fl}$;*VillinCre*^{ERT2} mice. Genotypes were verified using previously published protocols for each mutant strain (el Marjou et al., 2004; Hayhurst et al., 2001; Pu et al., 2004; Sodhi et al., 2006; Verzi et al., 2010). To activate Cre, mice received intraperitoneal (IP) injections of 1 mg tamoxifen (TAM, Sigma) in sunflower oil (Sigma) daily for 4-5 days. Mice were weighed daily and euthanized when the first mouse of a particular genotype became moribund (4 days for *Gata4*^{del}*Cdx2*^{del}, 7 days for *Hnf4a*^{del}*Cdx2*^{del} mice). Controls were injected with tamoxifen, but lacked *VillinCre*^{ERT2}. Animal Care and Use Committees at our institutions approved and monitored animal use.

Motif enrichment in ChIP-seq data

CDX2 ChIP-Seq data on intestinal villus cells (Verzi et al., 2013) (GEO accession: GSM851117) was analyzed using SeqPos (Liu et al., 2011) to identify enriched TF motifs within the 5,000 most significant CDX2 peaks.

Histochemistry and immunohistochemistry

The proximal (duodenum), central (jejunum), or distal (ileum) thirds of the small intestine were fixed overnight in 4% paraformaldehyde, embedded in paraffin and sectioned at 5 μm thickness. Some mice were injected IP with BrdU (Sigma, 1 mg/ml) 1 hour before euthanasia. Hematoxylin & eosin (H&E), Alcian blue and periodic acid-Schiff staining followed standard procedures. To detect alkaline phosphatase, slides were incubated in NTM solution (2 mM NaCl, 10 mM Tris-HCl pH 9.5, 5 mM $MgCl_2$) for 5 min, followed by nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate (NBT/BCIP ready-to-use tablets; Roche) solution for 15-30 min, and washed in PBS.

Immunohistochemistry (IHC) was performed as previously described (Verzi et al., 2010) using the following primary antibodies: rabbit anti-Ki67 (Santa Cruz, 1:200), rat anti-BrdU (AbD Serotec, 1:300), mouse anti-CDX2 (Biogenex, 1:20), goat anti-HNF4A (Santa Cruz, 1:1000), rabbit anti-Cleaved Caspase 3 (Cell Signaling, 1:1000), goat anti-CRS4C-1 (1:1000; gift of Andre Ouellette, University of Southern California, Los Angeles, CA), goat anti-GATA6 (R&D Laboratories, 1:50), or goat anti-GATA4 (Santa Cruz, 1:400). Representative images of histology and IHC from at least 5 mice of each genotype were obtained using an Olympus BX40 light microscope. Addition of scale bars and adjustment for brightness and contrast were performed in Photoshop.

To quantify morphological changes, at least 50 crypts or villi from each mouse (N=3) were measured and averaged. For Ki67 and goblet cell counts, at least 10 crypts or villi from at least 3 mice were counted and averaged; for BrdU+ counts, 25 crypts were counted and averaged. Significance was determined by *t*-test using GraphPad Prism software. *P-values* <0.05 were considered significant and are indicated in each figure.

Analysis of microarray data

Raw microarray data from the *Cdx2/Hnf4a* knockout genetic series (GEO accession GSE34567 (Verzi et al., 2013)) were re-analyzed using OneChannelGUI (Sanges et al., 2007). Background correction and normalization were performed using the robust multi-array average (RMA) method (Irizarry et al., 2003). Significant differential gene expression was determined using Limma (Smyth, 2004), with P-value adjustment (Q-value) using Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995). Fold-change values of probes targeting the same gene were averaged together, so that each gene is represented in the list only once. GENE-E software (http://www.broadinstitute.org/cancer/software/GENE-E/index.html) was used to perform hierarchical clustering of samples based on Pearson correlation and to create heatmap images. The genes displayed in the heatmap (**Figure 4.5C**) were bound by CDX2 (see methods below) and significantly down-regulated in $Hnf4a^{del}Cdx2^{del}$, compared to control, intestines. Each row displays the relative expression value in that row from the minimum (blue) to maximum (red). Rows were then clustered using *k*-means clustering into 3 groups containing genes with similar expression patterns across the genotypes.

RNA expression analysis

Mouse intestinal epithelium was harvested by incubating fresh jejuna in 5 mM EDTA solution for 45 min, as described previously (Verzi et al., 2013). RNA was isolated using TRIzol reagent (Invitrogen) and the RNeasy kit (Qiagen). For quantitative, reverse transcriptase-PCR, RNA was reverse transcribed (SuperScript III, Invitrogen) and assessed using FastStart Universal SYBR Green Master Mix (Roche) and specific primers for *Cdx2* (5'-

TCACCATCAGGAGGAAAAGTG-3' and 5'-GCAAGGAGGTCACAGGACTC-3'), *Gata4* (5'- TTTGAGCGAGTTGGG -3' and 5'- GAATGCGGGTGTGC -3'), *Gata6* (5'-CAGCAAGCTGTTGTGGTC-3' and 5'-GTCTGGTACATTTCCTCCG-3'), and *Hnf4a* (5'-GGTCAAGCTACGAGGACAGC-3' and 5'-ATGTACTTGGCCCACTCGAC-3'). Data were normalized for abundance of *Gapdh* (5'-GCCTTCCGTGTTCCTACCC-3' and 5'-TGCCTGCTTCACCACCTTC-3') or *HPRT* (5'-AAGCTTGCTGGTGAAAAGGA-3' and 5'-TTGCGCTCATCTTAGGCTTT-3') mRNA and expressed relative to control tissues.

Global assessment of RNA levels was performed on isolated jejunal epithelium from two control and two $Hnf4a^{del}Cdx2^{del}$ mice. RNA was isolated using TRIzol reagent and the RNeasy kit, followed by treatment with the Turbo DNA-free kit (Ambion) to remove genomic DNA. The RNA integrity number (RIN) for each sample was ≥ 9.8. RNA-sequencing libraries were prepared with the TruSeq RNA Sample Preparation Kit (Illumina), according to the manufacturer's instructions. 75 base pair single-end reads were sequenced on an Illumina NextSeq 500 instrument. Sequence tags were aligned to the *Mus musculus* reference genome build 9 (mm9) and the Tuxedo software package was used to align reads, assemble transcripts, and determine differences in transcript levels using a false discovery rate of 0.05 (Trapnell et al., 2012). The integrative genome viewer (IGV) was used to visualize aligned reads (Robinson et al., 2011). RNA-seq data are deposited in the GEO database, with accession number GSE62633.

Association of TF binding with nearby genes

Binding sites for CDX2 and HNF4A from ChIP-Seq experiments (GEO accession GSE34568 (Verzi et al., 2013)) were associated with the nearest gene within 30 kb using GREAT software (McLean et al., 2010). Genes with at least one binding site for each TF within

this range were considered in our further analysis. Biovenn was used to generate Venn diagrams (Hulsen et al., 2008).

Gene Ontology analysis

DAVID functional annotation clustering analysis was performed using medium classification stringency and default options (Huang da et al., 2009). Clusters with significant enrichment scores (>1.3) were considered (Huang da et al., 2009). When similar annotation clusters recurred in the list, we selected a representative GO term from the cluster and listed it with the cluster enrichment score for that group of terms.

Electron microscopy

Mouse ilea were flushed with PBS, fixed overnight or longer in EM fixative (2% formaldehyde, 2.5 % glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.4), and embedded in Taab 812 Resin (Marivac Ltd., Nova Scotia, Canada). 80 nm sections were cut, stained with 0.2% Lead Citrate, viewed, and imaged with a Philips Technai BioTwin Spirit Electron Microscope at an accelerating voltage of 80 kV.

Analysis of intestinal lipid absorption

Mice were placed on a diet (D12331 Research Diets) containing 58% calories from fat (the regular diet – Prolab Isopro RMH 3000 – contains 14% calories from fat), starting 5 days before the 1st dose of TAM until euthanasia. Intestines were fixed overnight at 4°C, equilibrated in 20% sucrose overnight at 4°C, embedded in O.C.T. compound (Tissue-Tek) and frozen on dry

ice. Lipid accumulation was visualized in 10 μm tissue sections incubated in 0.6% Oil Red O (Sigma) in propylene glycol at 60°C for 8 min.

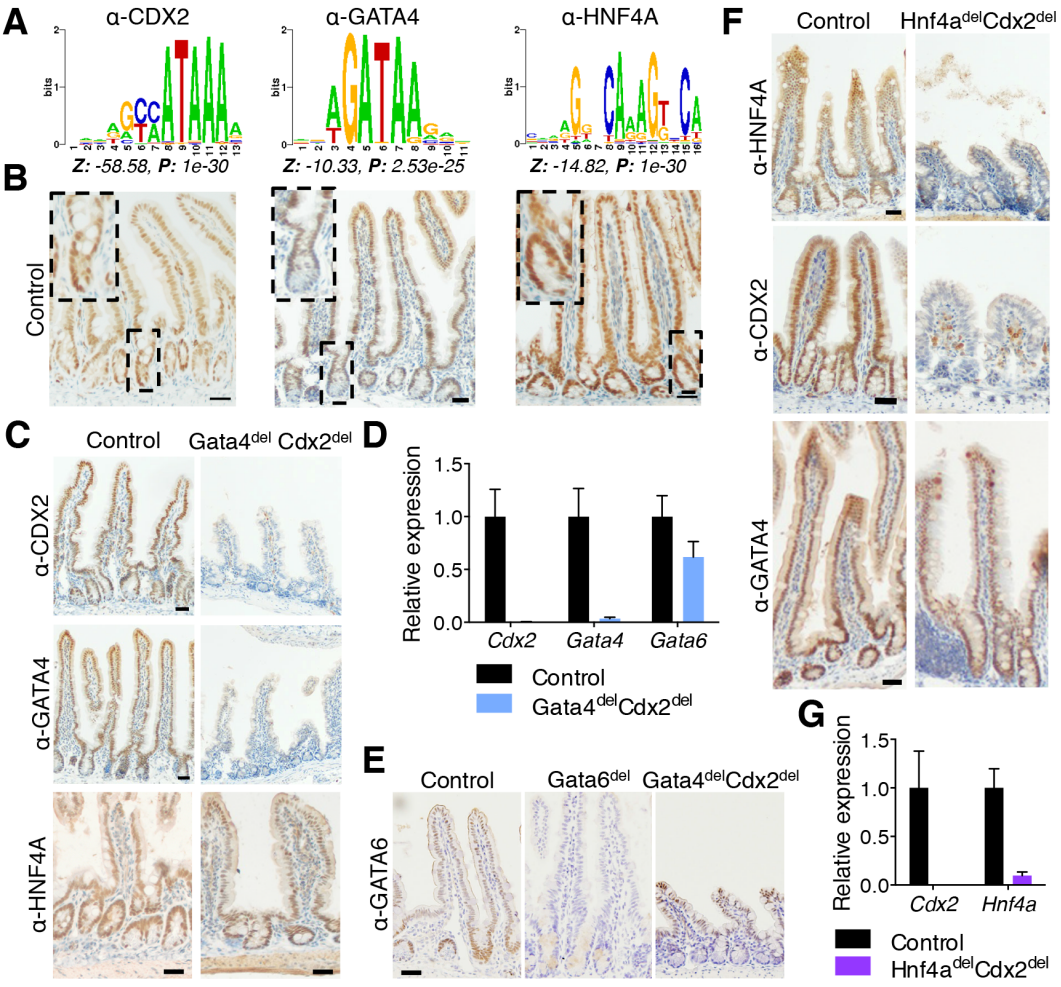
Results

Cis-element features identify candidate partner TFs in gut epithelial cells

To assess the potential for coordinate regulation among intestine-specific and pan-endodermal TFs, we searched for overrepresented sequence motifs in CDX2 chromatin immunoprecipitation (ChIP)-seq data from mouse intestinal villus cells (Verzi et al., 2013). Sequences corresponding to GATA and HNF4A were significantly enriched near CDX2-bound sites (**Figure 4.1A**), indicating that these endoderm-restricted TFs may recurrently join CDX2 at intestinal enhancers *in vivo*. Indeed, ChIP-seq for GATA4 and HNF4A in intestinal villus indicate co-occupancy with CDX2 at thousands of sites (Aronson et al., 2014; Verzi et al., 2013). Interestingly, all three TFs are expressed in crypt and villus epithelial cells (**Figure 4.1B**), suggesting the possibility of overlapping requirements that might be revealed in compound mutant mice. If CDX2 and these partner TFs largely overlap in function, then the compound mutant mice should phenocopy one another. Alternatively, distinct phenotypes would indicate that TF pairs control distinct programs. Moreover, transcripts perturbed upon combined TF loss should identify genes that require more than one TF for optimal expression.

Figure 4.1. Recurrence of DNA sequence motifs at CDX2 binding sites in mouse intestinal villus cells and efficient gene deletion in conditional mutant mice. (A) Motifs for CDX2, GATA and HNF4A are enriched near the 5,000 strongest CDX2 binding sites in wild-type mouse intestinal villus cells. Position weight matrices, with the corresponding Z-score and *P*-value, are indicated for each motif. **(B)** Immunohistochemistry (IHC) for each TF shows nuclear staining throughout the crypt-villus axis in wild-type mice (dashed boxes and insets show crypt details). **(C)** CDX2 (top) and GATA4 (middle) IHC in control and *Gata4*^{del}*Cdx2*^{del} jejunum confirms absence of the targeted TF proteins. IHC for HNF4A (bottom) reveals similar levels in *Gata4*^{del}*Cdx2*^{del} and control mice. **(D)** Quantitative RT-PCR for *Cdx2* and *Gata4* reveals complete loss of these mRNAs in the *Gata4*^{del}*Cdx2*^{del} jejunum, compared to controls (N=2 each). However, *Gata6* mRNA levels are only mildly reduced (N≥4 each). **(E)** GATA6 IHC shows persistent protein in *Gata4*^{del}*Cdx2*^{del} jejunum after 4 days of tamoxifen. At the same time, mice with *Gata6*^{del} alone recombined the allele efficiently and lost GATA6. **(F)** IHC for HNF4A (top) and CDX2 (middle) verifies absence of both proteins in the ileum of compound mutant intestines. Staining for GATA4 (bottom) in the duodenum of *Hnf4a*^{del}*Cdx2*^{del} mice (GATA4 is not expressed in the ileum) is similar to the control. **(G)** Quantitative RT-PCR for *Cdx2* and *Hnf4a* shows complete and nearly complete loss of mRNA levels, respectively (N=3 each). All scale bars, 30 μm. All graphs show means +/- standard error of the mean (SEM).

Figure 4.1 (Continued)



Efficient depletion of TFs in mouse intestinal epithelium

As GATA4 and CDX2 are implicated in control of selected villus cell genes in vivo (Benoit et al., 2010; Bosse et al., 2006) and in replication of cultured progenitor cells (Verzi et al., 2010), we first considered the intestinal GATA factors, and crossed mice to obtain conditional *Gata4*$^{Fl/Fl}$*;Gata6*$^{Fl/Fl}$*;Cdx2*$^{Fl/Fl}$*;Villin-Cre*^{ERT2} mutants. Tamoxifen treatment activated Cre recombinase and achieved intestinal loss of GATA4 and CDX2 mRNA and protein (**Figures 4.1C,D**), though *Gata6* loss was incomplete owing to inefficient recombination at the *Gata6*^{Fl} allele (Beuling et al., 2012) (**Figures 4.1D,E**). Thus, we refer to these mice as *Gata4*^{del}*Cdx2*^{del} and note that HNF4A expression was preserved (**Figure 4.1C**).

CDX2 and HNF4A co-occupy thousands of intestinal enhancers, where loss of CDX2 perturbs chromatin structure, resulting in decreases or loss of HNF4A binding (Verzi et al., 2013). Absence of HNF4A alone has little consequence on intestinal function (Babeu et al., 2009) or chromatin structure, but combined absence of CDX2 and HNF4A affects more transcripts than loss of either TF alone (Verzi et al., 2013). To determine the functional consequences of combined TF loss, we produced *Hnf4a*^{del}*Cdx2*^{del} intestines, which showed total or near-total loss of CDX2 and HNF4A mRNA and protein throughout the intestine (**Figures 4.1F,G**), without affecting GATA4 (**Figure 4.1F**).

Loss of Cdx2 and Gata4 impairs intestinal crypt cell replication

Gata4^{del}*Cdx2*^{del} mice lost weight rapidly (**Figure 4.2A**), became moribund, and required euthanasia within days of induced gene recombination. GATA4 is not expressed in the distal ileum, where CDX2 levels are the highest and the defects in *Cdx2*^{del} intestines are the most severe (van Wering et al., 2004). In the duodenum and jejunum, where GATA4 is abundant in wild-type mice, villi are slightly stunted in *Gata4*^{del} mutants. In contrast, *Gata4*^{del}*Cdx2*^{del} mice

showed shallow crypts and short villi throughout the small intestine (**Figures 4.2B-D**). As the effects of the *Cdx2*^{del} mice are most severe distally, we focused analysis on the most distal portion of the intestine that normally expresses *Gata4*, i.e., the jejunum. Here, significantly fewer crypt cells expressed the proliferation marker Ki67 (**Figures 4.3A,B**). In addition, the number of cells in S-phase, as marked by incorporation of BrdU during a 1 hour pulse, was also reduced in *Gata4*^{del}*Cdx2*^{del} mice (**Figure 4.3A,B**). The reduced crypt and villus heights did not reflect increased apoptosis in addition to the proliferation deficit (**Figure 4.3C**), but ectopic alkaline phosphatase expression (enterocytes) and Alcian blue staining (goblet cells) in *Gata4*^{del}*Cdx2*^{del} crypts indicated precocious cell maturation (**Figure 4.3D**), probably reflecting the cell cycle arrest. Villus alkaline phosphatase expression and Alcian blue staining verified the presence of mature cells (**Figure 4.3D**), and although cells retained a columnar morphology, they varied in shape, size, and nuclear morphology. The fraction of goblet cells, but not of enterocytes or Paneth cells, was modestly increased over intestines lacking only GATA4 or CDX2 (**Figures 4.3D,E**). These data show that intestinal crypt cell replication is a prominent shared function for CDX2 and GATA4, with a significantly larger defect in the compound mutant than either single mutant mouse. The villus defects may reflect this poor crypt cell turnover or indicate additional joint functions in cell maturation.

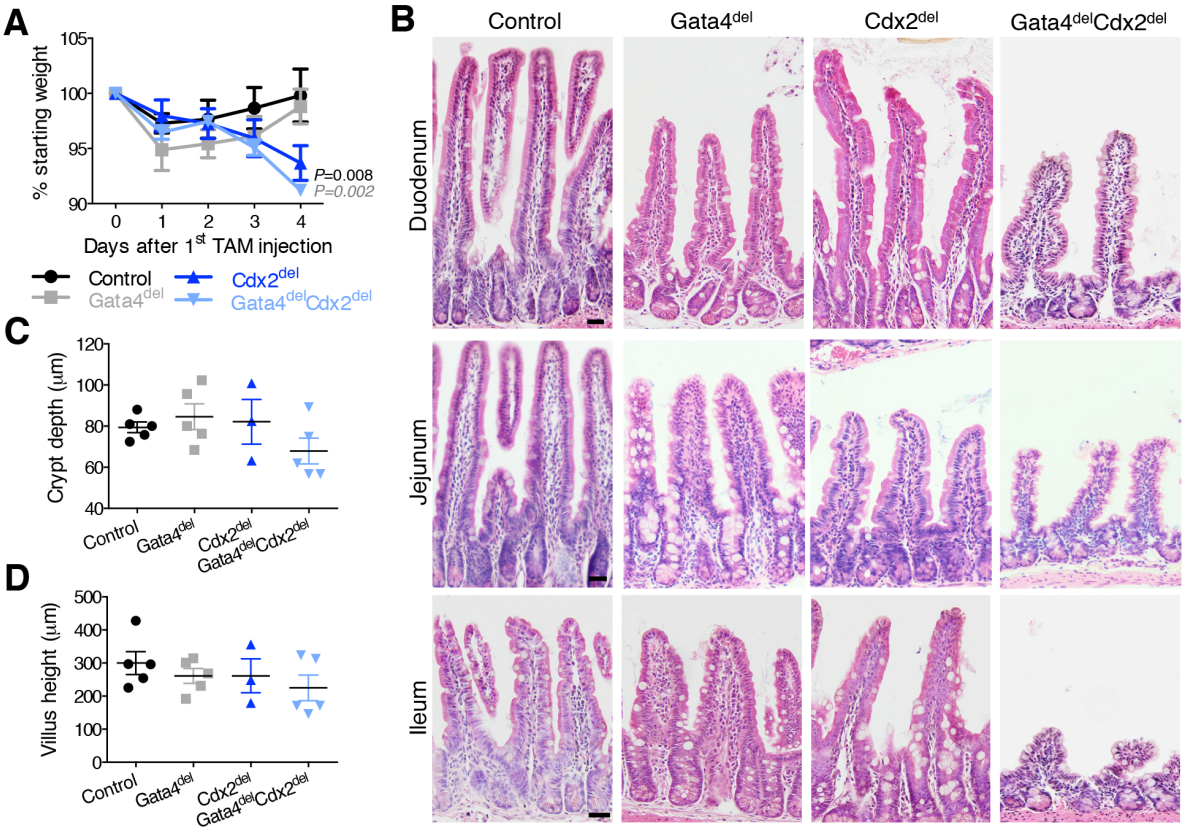
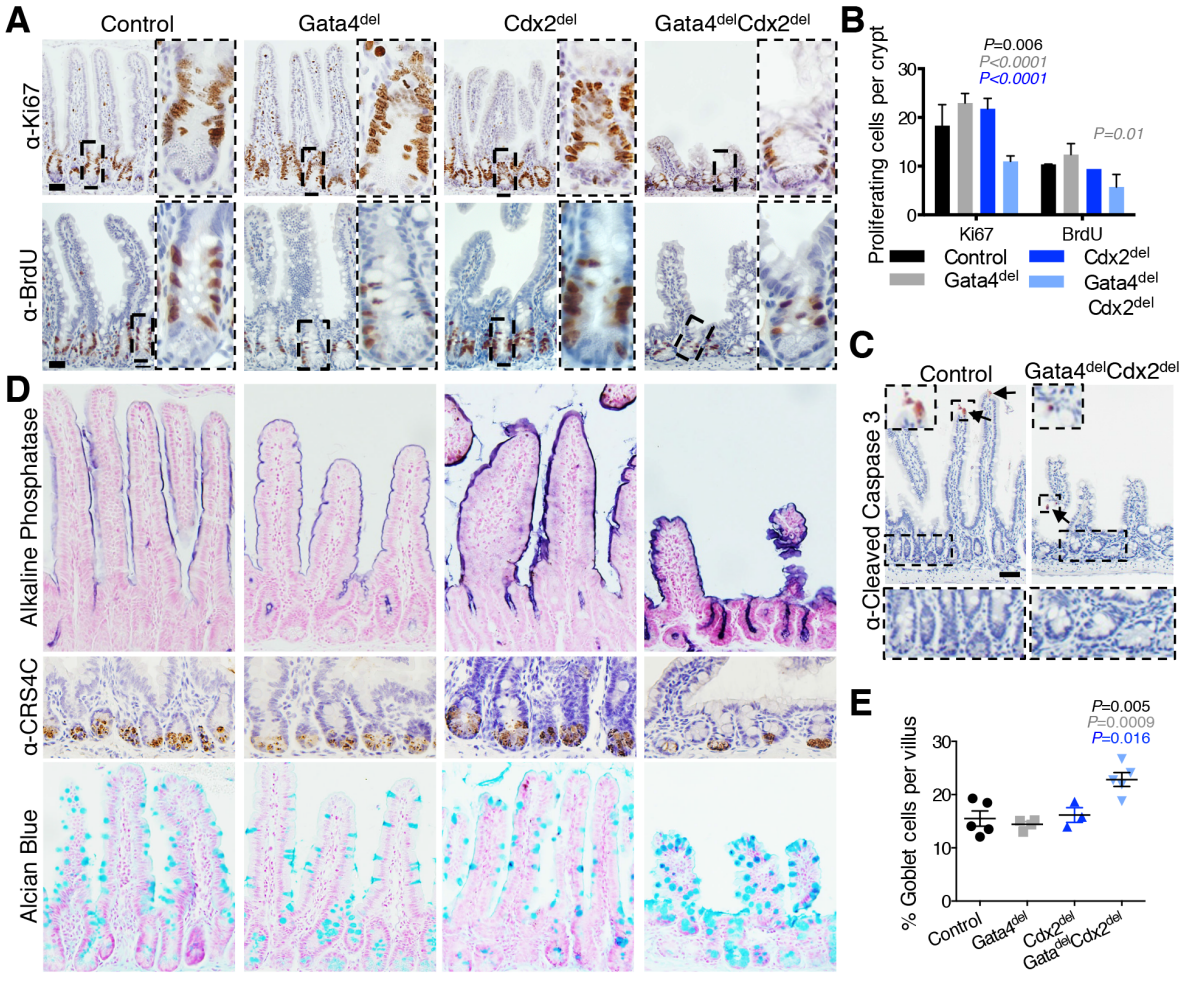


Figure 4.2. Morbidity and morphologic defects with combined loss of *Gata4* and *Cdx2*. (A) Mouse body weights each day after the start of tamoxifen injection reveal weight loss in *Cdx2*^{del} and *Gata4*^{del}*Cdx2*^{del} mice. Significance was calculated between *Gata4*^{del}*Cdx2*^{del} mice and other genotypes on day 4 (N≥5). **(B)** Duodenum (top), jejunum (middle) and ileum (bottom) tissue sections stained with hematoxylin and eosin (H&E) show reduced villus height and crypt depth in *Gata4*^{del}*Cdx2*^{del} intestines, compared to loss of either TF alone. Scale bars, 30 μm. **(C-D)** Quantitation of crypt (C) and villus (D) length in *Gata4*^{del}*Cdx2*^{del} jejunum, compared to single mutants and controls (N≥3). All graphs represent the means +/- SEM. Statistical significance was assessed using Student's *t*-test; P-values are indicated when significant and color-coded to match the sample to which *Gata4*^{del}*Cdx2*^{del} samples were compared.

Figure 4.3. Effects of combined loss of *Gata4* and *Cdx2* on proliferating and differentiated intestinal cells. (A) IHC of proliferative markers Ki67 and BrdU (administered 1 hour before euthanasia) confirms significantly reduced proliferation in $Gata4^{del}Cdx2^{del}$ jejuni, compared to single mutant and control littermates, quantified in **(B)**. Dashed boxes outline the areas magnified to the right. Scale bars, 50 μm. **(C)** Cleaved caspase 3 staining reveals that apoptosis is not increased in $Gata4^{del}Cdx2^{del}$ intestines compared to control villi or crypts (area in dotted region magnified below). Arrows, Caspase 3-positive cells (an example is magnified in the dotted region to the top left. **(D)** Histochemistry and IHC of jejunum for alkaline phosphatase and CRS4C in $Gata4^{del}Cdx2^{del}$ and control mice show no differences in mature enterocytes and Paneth cells, respectively. Alcian Blue staining reveals an increased fraction of villus goblet cells in $Gata4^{del}Cdx2^{del}$ intestines, quantified in **(E)**. Of note, $Gata4^{del}Cdx2^{del}$ mice have ectopic alkaline phosphatase and Alcian blue staining in crypts. All graphs show means +/- SEM. Significant changes between $Gata4^{del}Cdx2^{del}$ and other genotypes were calculated using Student's *t*-test. Results are color-coded by genotype; insignificant differences are not marked. Scale bars in **(A)** are 50 μm, all others 30 μm.

Figure 4.3 (Continued)

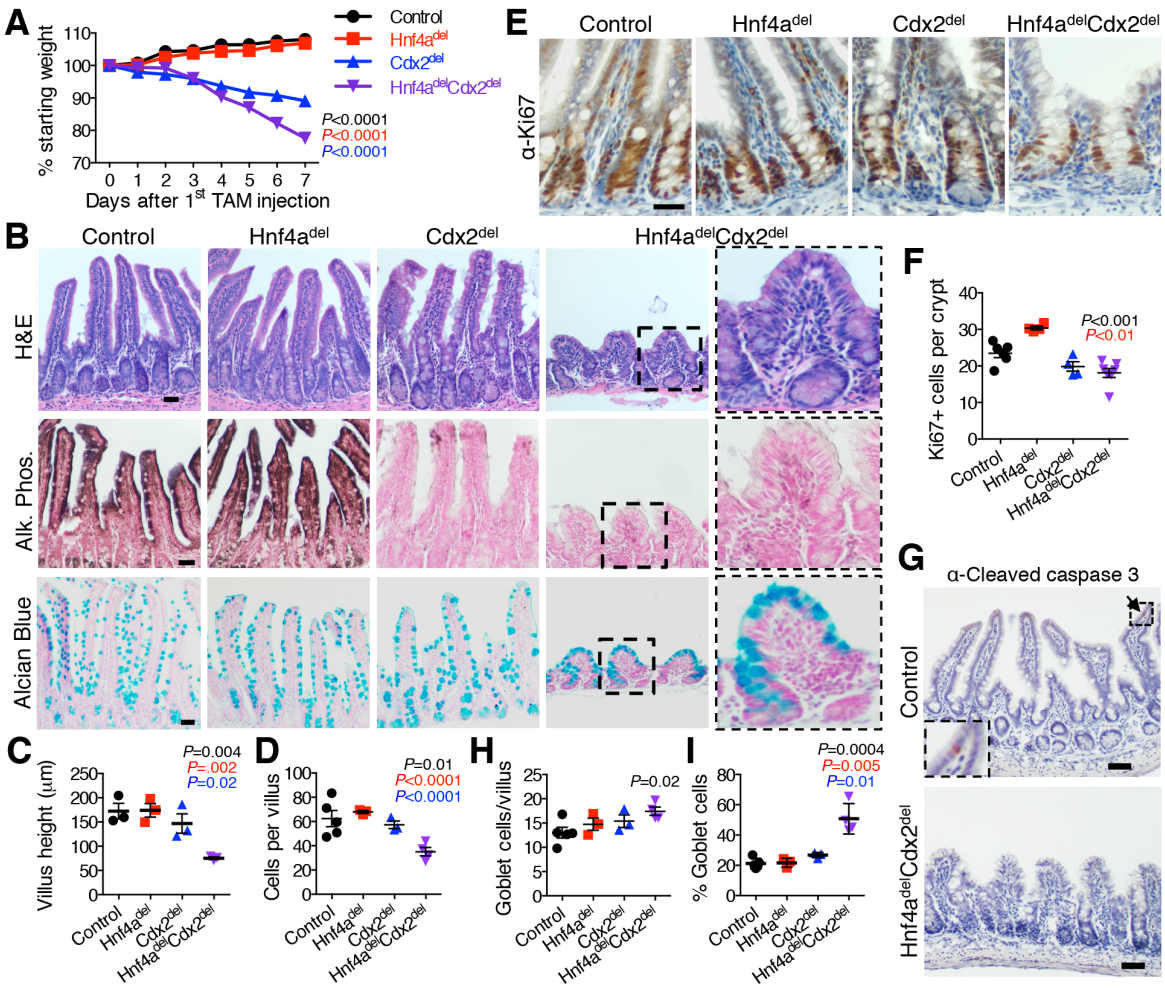


Combined loss of CDX2 and HNF4A compromises enterocyte differentiation without affecting crypt cell replication

Swift weight loss in *Hnf4a*^{del}*Cdx2*^{del} mice (**Figure 4.4A**) warranted euthanasia two weeks earlier than in *Cdx2*^{del} littermates, with severe diarrhea and steatorrhea occurring during the course of tamoxifen administration. Duodenal and jejunal villus heights were similar to those in *Cdx2*^{del} intestines, and enterocytes in these regions showed little cellular atypia (data not shown). By contrast, villi in the ileum were dysplastic and significantly stunted (**Figures 4.4B,C**), with many cells showing pyknosis, and total villus cell numbers were significantly lower than in *Hnf4a*^{del} or *Cdx2*^{del} villi (**Figures 4.4D**). Importantly, and in contrast to *Gata4*^{del}*Cdx2*^{del} mice, this was not a result of reduced crypt cell proliferation (**Figures 4.4E,F**) or of increased apoptosis (**Figure 4.4G**) but rather of aberrant epithelial cell differentiation. A total absence of alkaline phosphatase in *Cdx2*^{del} and *Hnf4a*^{del}*Cdx2*^{del} villi in the ileum (**Figure 4.4B**) implicated the enterocyte compartment. Indeed, Alcian blue staining revealed a predominance of goblet cells (**Figure 4.4B**), whereas total goblet cell numbers were similar to control mice (**Figure 4.4H**). Thus, the fraction of goblet cells per villus was significantly increased in the *Hnf4a*^{del}*Cdx2*^{del} ileum (**Figure 4.4I**), indicating a lack of mature enterocytes. Weight loss, leading to rapid demise, is likely due to this enterocyte deficit, which was greatest in the ileum, distinct from the absent (*Hnf4a*^{del}) or subtle (*Cdx2*^{del}) defects observed in single mutant mice. The rapid and dramatic consequences of combined loss of CDX2 and HNF4A indicate that the two TFs control, cooperatively or in parallel, transcriptional programs specific to mature intestinal villus cells.

Figure 4.4. Combined loss of *Hnf4a* and *Cdx2* accelerates death and results in reduction of absorptive enterocytes. (A) Weight loss in $Hnf4^{del}Cdx2^{del}$ mice is accelerated compared to single mutants or controls, requiring euthanasia within 7 days of induced gene deletion. Significance was calculated between $Hnf4a^{del}Cdx2^{del}$ mice and other genotypes on day 7 (N=6). **(B)** H&E staining of ileum in the *Hnf4a/Cdx2* genetic series shows significant shortening of $Hnf4a^{del}Cdx2^{del}$ villi, quantified in **(C)**. Both $Cdx2^{del}$ and $Hnf4a^{del}Cdx2^{del}$ mice lack alkaline phosphatase in ileal villus cells, indicating defective enterocyte maturation. Alcian blue staining reveals goblet cell abundance on $Hnf4a^{del}Cdx2^{del}$ villi. Dashed boxes outline areas magnified to the right. **(D)** Total numbers of ileal villus cells are reduced in $Hnf4a^{del}Cdx2^{del}$ mice. **(E)** Ki67 IHC shows that combined loss of *Hnf4a* and *Cdx2* affects proliferation no more than loss of *Cdx2* alone, quantified in **(F). (G)** Cleaved caspase 3 staining reveals no significant apoptosis in $Hnf4a^{del}Cdx2^{del}$ ileal epithelium; a few apoptotic cells were present at the villus tips in control mice (arrow). Caspase 3-positive cell magnified in dotted area at bottom left. **(H)** Quantitation of total goblet cell numbers per villus in the ileum shows no increase in $Hnf4a^{del}Cdx2^{del}$, compared to single mutant, intestines. **(I)** $Hnf4a^{del}Cdx2^{del}$ mice exhibit an increased proportion of ileal goblet cells. Scale bars in **(B)** and **(E)**, 30 μm; in **(G),** 50 μm. All graphs show means +/- SEM; significant changes between $Hnf4a^{del}Cdx2^{del}$ and other genotypes were calculated using Student's t-test. Results are color coded by genotype as indicated; insignificant results not indicated.

Figure 4.4 (Continued)



Identification of genes co-regulated by CDX2 and HNF4A

To investigate transcriptional programs that depend on CDX2 and/or HNF4A in maturing cells, we first compared transcriptomes in mature villi 7 days after tamoxifen-induced deletion of *Hnf4a*, *Cdx2*, or both genes (Verzi et al., 2013). Because abnormalities in the ileum could represent direct consequences of TF loss or the indirect effects of epithelial dysfunction, we examined jejunal transcriptomes, reasoning that the absence of a strong phenotype in this region would highlight the primary effects, free of confounding considerations. Compound mutant $Hnf4a^{del}Cdx2^{del}$ samples clustered separately from single $Hnf4a^{del}$ or $Cdx2^{del}$ mutants, indicating increased severity of global changes in gene expression (**Figure 4.5A**). Compared to the controls, 573 transcripts were significantly increased and 586 transcripts were reduced in $Hnf4a^{del}Cdx2^{del}$ intestines ($Q<0.05$).

To identify likely direct transcriptional targets, we selected genes showing CDX2 and HNF4A co-occupancy in jejunal villus epithelial cells (<30 kb from the nearest gene). More than half (52.9%) of the genes reduced in $Hnf4a^{del}Cdx2^{del}$ intestines are bound by both TFs, compared to 16.4% of genes with increased levels (**Figure 4.5B**), consistent with a role for these factors primarily in gene activation. Among the 310 genes down-regulated in $Hnf4a^{del}Cdx2^{del}$ mice, compared to controls, and showing TF co-occupancy, *k*-means clustering identified three prominent patterns with roughly equal frequency (**Figure 4.5C**): Genes equally reduced compared to controls in $Cdx2^{del}$ and $Hnf4a^{del}Cdx2^{del}$, but not in $Hnf4a^{del}$ intestines (cluster C); genes modestly perturbed in each single mutant and further dysregulated in the compound mutant (cluster B); and transcripts that barely change in either single mutant, compared to controls, but are significantly reduced in $Hnf4a^{del}Cdx2^{del}$ intestines (Cluster A). Selected examples of genes from clusters A (*Synpo*) and C (*Abp1*/*Aoc1*) (**Figure 4.5D**) show binding of

both CDX2 and HNF4A. This analysis indicates that CDX2 alone is required for proper expression of genes in cluster C, whereas additional loss of HNF4A was necessary to affect genes in clusters A and B.

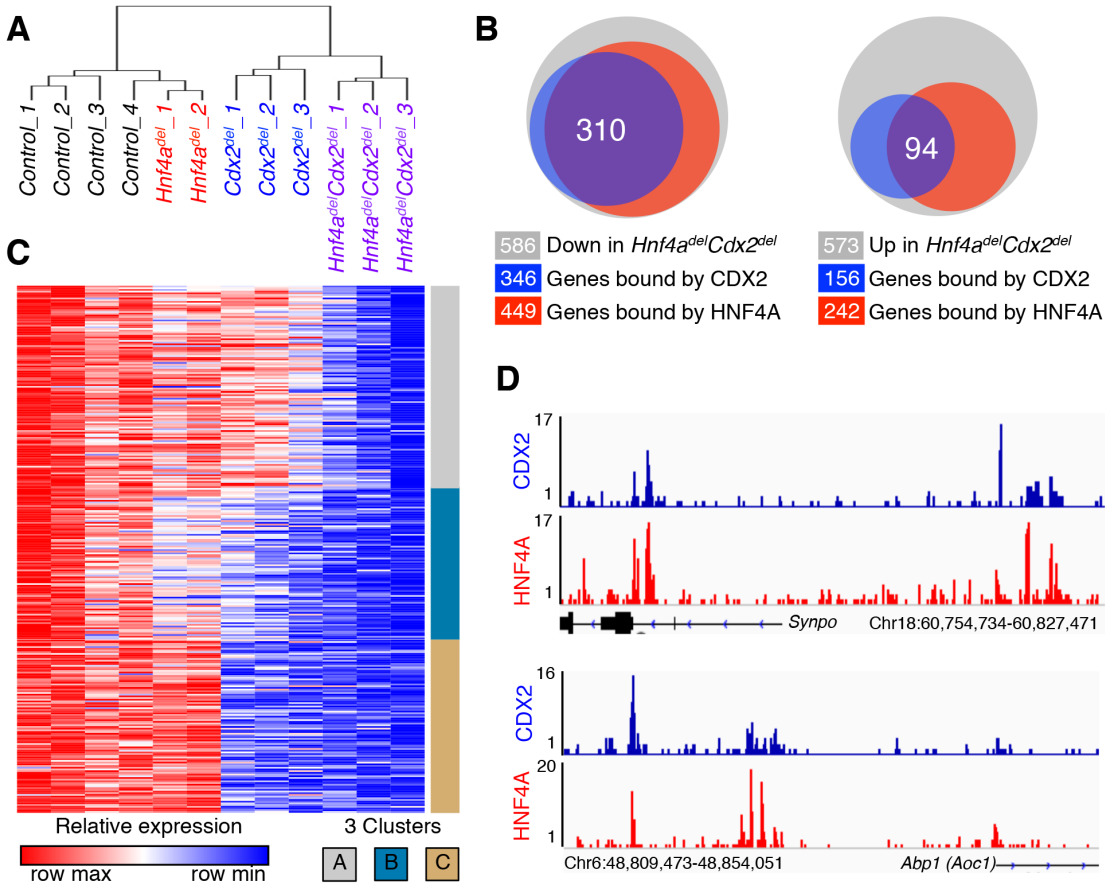


Figure 4.5. mRNA analysis of single and compound mutant intestinal epithelia reveals dependencies on single or multiple TFs. (A) Hierarchical cluster analysis of global mRNA expression in a series of *Hnf4a;Cdx2* mutant intestinal epithelia reveals that transcripts are more profoundly affected in *Hnf4a*^{del}*Cdx2*^{del} than in either single mutant. **(B)** Venn diagram depicting dysregulated genes in *Hnf4a*^{del}*Cdx2*^{del} mice (grey) and their overlap with CDX2 (blue) and HNF4A (red) binding. Gene numbers represented in each circle are indicated below. **(C)** Heatmap showing expression levels in each genotype of the 310 down-regulated and co-occupied genes. The patterns of change fall into 3 *k*-means clusters A-C. **(D)** Data traces for CDX2 and HNF4A ChIP-seq at the *Abp1* (*Aoc1*) and *Synpo* loci, demonstrating TF co-occupancy at selected sites. The Y-axis indicates the density of mapped sequence tags and genome coordinates are shown on the X-axis; Chr, chromosome.

To examine these changes in gene expression with greater confidence and quantitative accuracy, we performed RNA-seq on epithelial cells isolated from control and *Hnf4a*^{del}*Cdx2*^{del} jejuni. We confirmed disruption of *Cdx2* and *Hnf4a* by RNA-seq, noting an absence of transcript reads from exon 2 in *Cdx2* and nearly complete absence of transcript reads from exons 4 and 5 in *Hnf4a* (data not shown). RNA-seq results corroborated our microarray data and expanded the number of candidate direct transcriptional targets, identifying 840 down-regulated and 642 up-regulated genes ($Q<0.05$) in *Hnf4a*^{del}*Cdx2*^{del} intestines. Integration of RNA-seq and ChIP-seq data indicated that both CDX2 and HNF4A were bound within 30 kb of 44% (368) of the down-regulated genes, 17% of up-regulated or unaffected genes, and 3% of genes not expressed in intestinal villi (**Figures 4.6A,B**). These findings further support the idea that CDX2 and HNF4A function mainly as activators. Although the two TFs could, in principle, affect intestinal genes independently, their binding near genes that respond to loss of both TFs is a good indication of co-regulation.

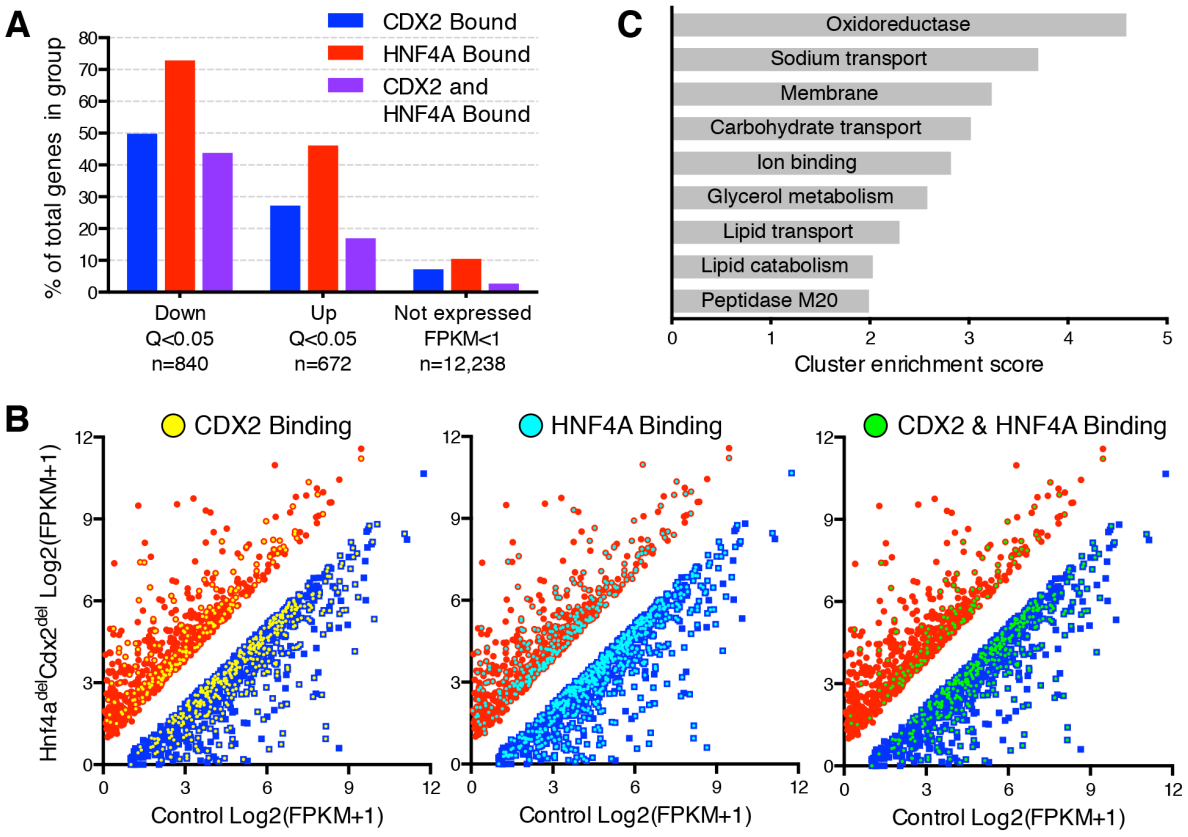


Figure 4.6. RNA-seq analysis further identifies CDX2 and HNF4A co-regulated intestinal genes. (A,B) Integration of CDX2 and HNF4A binding data from ChIP-seq with mRNA expression data from RNA-seq analysis of control and *Hnf4a*^{del}*Cdx2*^{del} intestines. (A) Genes significantly down- or up-regulated (Q<0.05) are shown in relation to those not expressed (FPKM<1) in either sample. The fraction of genes in each group bound by CDX2 (blue), HNF4A (red), and both (purple) TFs indicates again that these TFs mainly activate target genes. **(B)** FPKM graphs showing RNA-seq expression of significantly up and down-regulated genes are overlaid with binding for CDX2 (left, yellow), HNF4A (center, aqua) and both (right, green), showing a higher distribution of bound down-regulated genes. **(C)** Gene Ontology (GO) annotation clusters from DAVID analysis of the 362 genes significantly down-regulated in *Hnf4a*^{del}*Cdx2*^{del} intestines and bound by both TFs.

CDX2 and HNF4A co-regulate genes for specific aspects of digestive physiology

To determine the physiologic consequences of gene co-regulation, we performed Gene Ontology (GO) analysis on the 368 genes that RNA-seq identified as reduced in *Hnf4a*^{del}*Cdx2*^{del} intestines and where ChIP-seq showed binding of both TFs. Clustering of significantly enriched GO terms highlighted genes central to enterocyte functions, including oxidation-reduction, transport of ions, carbohydrates and lipids, and components of the plasma membrane (**Figure 4.6C**). Thus, combined loss of CDX2 and HNF4A in vivo produces attrition of enterocytes in the ileum and significantly altered gene expression in jejunal enterocytes, albeit with few overt histologic manifestations. Together, these anomalies can account for the rapid, uniform lethality in *Hnf4a*^{del}*Cdx2*^{del} mice.

Because defects in cell membranes and in nutrient absorption should be objectively observable, we focused efforts on identifying these deficits in *Hnf4a*^{del}*Cdx2*^{del} mice. Apical microvilli in *Hnf4a*^{del} mice are grossly intact (Babeu et al., 2009). In contrast, microvilli on the apical membranes of *Cdx2*^{del} enterocytes were moderately scant and stunted, whereas those in *Hnf4a*^{del}*Cdx2*^{del} cells were sparse, shortened, and disarrayed (**Figure 4.7A**). Beyond this significant brush border defect, the *Hnf4a*^{del}*Cdx2*^{del} enterocytes showed additional intracellular anomalies, including an accumulation of light droplets or vacuoles near the cell apex and dark, possibly membrane-bound vesicles within mitochondria. The contribution of these defects toward cell attrition is not presently clear. However, together with reduced expression of peptidases, other hydrolases, and genes involved in symporter activity and intestinal absorption, the brush border anomalies undoubtedly contribute to the severe malnutrition.

In addition, we made particular note of lipid metabolism because intestinal absorption of dietary lipids is highly relevant to human health and HNF4A is known to regulate hepatic lipid

metabolism (Hayhurst et al., 2001). To test the hypothesis that co-regulation of genes related to lipid absorption might also underlie the profound, rapid malnutrition in *Cdx2*^{del}*Hnf4a*^{del} mice, we placed mice on a high-fat diet starting 5 days before TF gene disruption (**Figure 4.7B**) and stained intestines 12 days later with Oil Red O to detect lipid deposits. Enterocytes accumulate lipids as a result of fatty acid and cholesterol transport as well as *de novo* biosynthesis (Sturley and Hussain, 2012). In the presence of bile salts and pancreatic lipase, enterocytes in the proximal gut (duodenum) absorb dietary lipids efficiently, leaving no luminal lipid residue for absorption by distal (ileal) enterocytes (Yamada and Alpers, 2009). Accordingly, Cre^{-} and *Hnf4a*^{del} control mice showed prominent Oil Red O staining in duodenal cells and virtually none in the ileum (**Figure 4.7C**). In contrast, *Cdx2*^{del} single-mutant mice showed reduced lipid in duodenal enterocytes and significant levels in the ileum, indicating that CDX2-null cells absorb dietary lipids inefficiently, leaving a residual amount for ileal absorption. *Hnf4a*^{del}*Cdx2*^{del} intestines showed minimal Oil Red O staining in any region (**Figure 4.7C**), revealing a global and profound defect in lipid absorption. Importantly, this was not a trivial consequence of enterocyte depletion, because the duodenum, where enterocytes were plentiful, showed no lipid uptake. Moreover, we observed no lipid uptake in ileal villi that contained both goblet cells and a few enterocytes. Taken together, these findings reveal that CDX2 and HNF4A co-regulate genes necessary for dietary lipid metabolism, which is more severely impaired when both TFs are absent than when either TF alone is lost. This striking defect likely also contributes toward rapid weight loss in *Hnf4a*^{del}*Cdx2*^{del} mice.

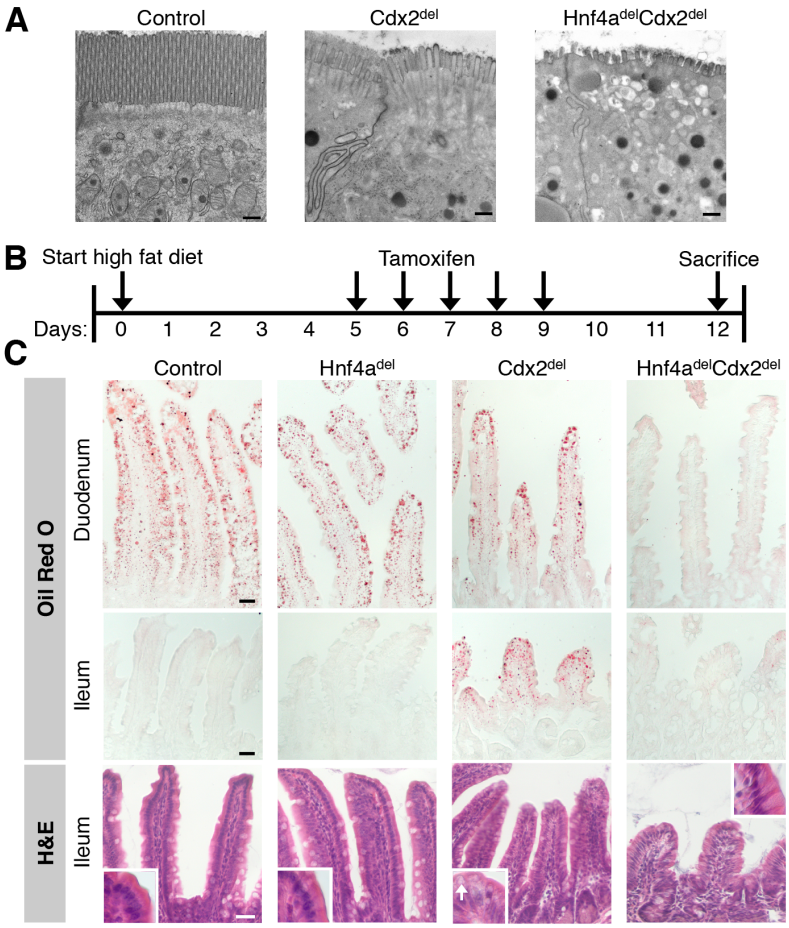


Figure 4.7. HNF4A and CDX2 co-regulate enterocyte apical microvilli and absorption of dietary lipids. (A) Electron micrographs of ileal enterocyte apices showing an intact brush border in control mice, shortened microvilli in *Cdx2*^{del} mice, and severely depleted and stunted microvilli in *Hnf4a*^{del}*Cdx2*^{del}, intestines. Scale bars, 500 nm. **(B)** Experimental schema for maintaining mice on a high-fat diet before, during, and after inducing gene deletion. **(C)** Oil Red O staining for neutral lipids revealed ectopic lipid absorption in *Cdx2*^{del} ileal enterocytes and total lack of lipid absorption in *Hnf4*^{del}*Cdx2*^{del} intestines. H&E staining of ileum reveals light lipid droplets in *Cdx2*^{del} epithelium (inset, white arrow), but not in other genotypes. Scale bars, 30 μm.

Discussion

Selected lineage-restricted TFs exert considerable control over each cell type's unique transcriptional program. For example, GATA1, TAL1, EKLF and NF-E2 together regulate most erythroid blood cell genes (Orkin and Zon, 2008) and a few basic-helix-loop-helix TFs control much of the muscle cell-specific transcriptome (Tapscott, 2005). Several lines of evidence implicate CDX2, HNF4A, and GATA4/6 in control of intestinal genes. First, functional *cis*-elements for individual intestinal genes repeatedly reveal these and few other TFs' activities (Gregory et al., 2004; Mitchelmore et al., 2000). Second, the corresponding sequence motifs recur frequently at intestine-active enhancers identified by histone marks and nucleosome depletion (Verzi et al., 2010; Zhu et al., 2013) or differential DNA methylation (Sheaffer et al., 2014). Here we show that additional loss of CDX2 dramatically unmasks GATA4 and HNF4A requirements in knockout mice. Defects in both compound mutant intestines are rapidly lethal, include distinct components, and help delineate functional TF hierarchies in the intestine.

Arrested crypt cell replication in $Gata4^{del}Cdx2^{del}$ intestines indicates collaboration of these TFs in intestinal crypts and contrasts sharply with the lack of replication deficits in $Cdx2^{del}Hnf4a^{del}$ intestines. Although our data do not address whether CDX2 and GATA control the same genes additively or different genes, the GATA sequence motif and binding of a closely related factor, GATA6, is highly enriched near CDX2 binding sites in replicating human intestinal cells in culture (Verzi et al., 2010), suggesting possible regulation of the same genes. CDX2 and GATA proteins likely have additional, joint roles in villus cell maturation, as evidenced by a large overlap of their respective binding sites in intestinal villus (Aronson et al., 2014). It was difficult to address this point unequivocally in our analysis because the deficit in crypt cell replication may explain the reduced villus cell height whether cell maturation is intact

or defective. In the future, identification of GATA4 binding sites and dependent genes in intestinal crypt cells may uncover co-regulated genes in this compartment, much as this study uncovered CDX2-HNF4A co-regulated genes in villus cells.

Suppression of crypt cell replication and the increase in goblet cell numbers in *Cdx2*^{del}*Gata4*^{del} intestines may reflect reduced Notch signaling or an effect on some other shared pathway. Whether the phenotype is caused by defects in stem cells or transit-amplifying progenitors also warrants further investigation. By contrast, because *Cdx2*^{del}*Hnf4a*^{del} crypt cells proliferate normally, we can categorically attribute these phenotypes to failures in cell maturation, and in enterocytes in particular.

CDX2 loss disrupts chromatin structure and binding of other TFs, such as HNF4A (Verzi et al., 2013), implying that CDX2 maintains chromatin access in intestinal cells. In a simple hierarchy, where other TFs depend on CDX2 wholly, compound mutant and *Cdx2*^{del} intestines should largely phenocopy one another, but our studies on HNF4A indicate otherwise. *Cdx2*^{del}*Hnf4a*^{del} mice fared profoundly worse than *Cdx2*^{del} littermates, with accelerated demise, near absence of ileal enterocytes, significant gene dysregulation in jejunal villus cells, and lack of dietary fat absorption. Many more genes are dysregulated in *Cdx2*^{del}*Hnf4a*^{del} than in either single mutant. One likely reason is that HNF4A regulates genes in addition to those where CDX2 enables chromatin access, but CDX2 loss is necessary to expose that dependency. A second reason is that HNF4A provides necessary additive activity at regulatory sites that both TFs co-occupy. Indeed, the genes most affected in *Cdx2*^{del}*Hnf4a*^{del} intestines are enriched for such co-occupancy, indicating that loss of single TFs may preserve some enhancer function but absence of both TFs is necessary to abrogate transcription completely.

In summary, combinatorial control allows TFs to elicit diverse transcriptional outcomes, in both embryos and adult tissues, and our data highlight this aspect of gene control *in vivo*. The intestinal "master regulator" CDX2 participates in various TF combinations. Specific pairings with partners such as GATA proteins drive cell replication, whereas separate pairings with HNF4A and other TFs regulate genes in terminally differentiated cells. The contributions of each TF no doubt vary at different *cis*-regulatory sites. At some sites, absence of a single factor has profound effects, as occurs commonly with CDX2, but not with GATA4 or HNF4A, loss. At other sites, gene expression suffers only when more than one TF is absent. Our studies reveal many examples of such cooperativity and the scope of gene dysregulation in $Cdx2^{del}Hnf4a^{del}$ (compared to $Cdx2^{del}$) intestines matched the greater severity of tissue defects and malnutrition. Particularly at enterocyte genes necessary to construct the microvillus brush border or absorb nutrients, including dietary fat, both CDX2 and HNF4A are necessary. These TFs co-occupy the corresponding enhancers and loss of both TFs affects mRNA levels more than the absence of either factor alone. These findings illuminate the transcriptional basis for vital intestinal functions.

Acknowledgements

We thank Sylvie Robine and William Pu for sharing *Villin-*Cre^{ER-T2} and conditional *Gata* mutant mice, respectively; Justina Chen and Kelly Stapleton for assistance with mouse dissections and immunostaining; Marian Neutra for insights into intestinal phenotypes; Andre Ouellette for the gift of CRS4C1 antibody; Pere Puigserver for providing high-fat mouse chow; the electron microscopy core facility at Harvard Medical School; the Center for Functional Cancer

Epigenetics for preparing RNA-seq libraries; and the Molecular Biology Core Facilities at Dana-Farber Cancer Institute for next-generation sequencing.

References

Aronson, B.E., Rabello Aronson, S., Berkhout, R.P., Chavoushi, S.F., He, A., Pu, W.T., Verzi, M.P., and Krasinski, S.D. (2014). GATA4 represses an ileal program of gene expression in the proximal small intestine by inhibiting the acetylation of histone H3, lysine 27. Biochim Biophys Acta *1839*, 1273-1282.

Babeu, J.-P., Darsigny, M., Lussier, C.R., and Boudreau, F. (2009). 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 *297*, G124-134.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol *57*, 289-300.

Benoit, Y.D., Pare, F., Francoeur, C., Jean, D., Tremblay, E., Boudreau, F., Escaffit, F., and Beaulieu, J.F. (2010). Cooperation between HNF-1alpha, Cdx2, and GATA-4 in initiating an enterocytic differentiation program in a normal human intestinal epithelial progenitor cell line. Am J Physiol Gastrointest Liver Physiol *298*, G504-517.

Beuling, E., Aronson, B.E., Tran, L.M., Stapleton, K.A., ter Horst, E.N., Vissers, L.A., Verzi, M.P., and Krasinski, S.D. (2012). GATA6 is required for proliferation, migration, secretory cell maturation, and gene expression in the mature mouse colon. Mol Cell Biol *32*, 3392-3402.

Beuling, E., Baffour-Awuah, N.Y.A., Stapleton, K.A., Aronson, B.E., Noah, T.K., Shroyer, N.F., Duncan, S.A., Fleet, J.C., and Krasinski, S.D. (2011). GATA Factors Regulate Proliferation, Differentiation, and Gene Expression in Small Intestine of Mature Mice. Gastroenterology *140*, 1219-1229.e1212.

Bosse, T., Piaseckyj, C.M., Burghard, E., Fialkovich, J.J., Rajagopal, S., Pu, W.T., and Krasinski, S.D. (2006). Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine. Mol Cell Biol *26*, 9060-9070.

el Marjou, F., Janssen, K.-P., Chang, B.H.-J., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186-193.

Gao, N., White, P., and Kaestner, K.H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev Cell *16*, 588-599.

Garrison, W.D., Battle, M.A., Yang, C., Kaestner, K.H., Sladek, F.M., and Duncan, S.A. (2006). Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. Gastroenterology *130*, 1207-1220.

German, M.S., Wang, J., Chadwick, R.B., and Rutter, W.J. (1992). Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. Genes & development *6*, 2165-2176.

Gregory, P.A., Lewinsky, R.H., Gardner-Stephen, D.A., and Mackenzie, P.I. (2004). Coordinate regulation of the human UDP-glucuronosyltransferase 1A8, 1A9, and 1A10 genes by hepatocyte nuclear factor 1alpha and the caudal-related homeodomain protein 2. Mol Pharmacol *65*, 953-963.

Hayhurst, G.P., Lee, Y.H., Lambert, G., Ward, J.M., and Gonzalez, F.J. (2001). Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. Mol Cell Biol *21*, 1393-1403.

Hryniuk, A., Grainger, S., Savory, J.G.A., and Lohnes, D. (2012). Cdx function is required for maintenance of intestinal identity in the adult. Dev Biol *363*, 426-437.

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

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics *9*, 488.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

James, R., Erler, T., and Kazenwadel, J. (1994). Structure of the murine homeobox gene cdx-2. Expression in embryonic and adult intestinal epithelium. J Biol Chem *269*, 15229-15237.

James, R., and Kazenwadel, J. (1991). Homeobox gene expression in the intestinal epithelium of adult mice. The Journal of biological chemistry *266*, 3246-3251.

Jin, T., and Drucker, D.J. (1996). Activation of proglucagon gene transcription through a novel promoter element by the caudal-related homeodomain protein cdx-2/3. Molecular and cellular biology *16*, 19-28.

Krapp, A., Knofler, M., Frutiger, S., Hughes, G.J., Hagenbuchle, O., and Wellauer, P.K. (1996). The p48 DNA-binding subunit of transcription factor PTF1 is a new exocrine pancreas-specific basic helix-loop-helix protein. EMBO J *15*, 4317-4329.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Liu, T., Zhang, X., So, C.K., Wang, S., Wang, P., Yan, L., Myers, R., Chen, Z., Patterson, A.P., Yang, C.S.*, et al.* (2007). Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. Carcinogenesis *28*, 488-496.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Mitchelmore, C., Troelsen, J.T., Spodsberg, N., Sjostrom, H., and Noren, O. (2000). Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. Biochem J *346 Pt 2*, 529-535.

Ohlsson, H., Karlsson, K., and Edlund, T. (1993). IPF1, a homeodomain-containing transactivator of the insulin gene. EMBO J *12*, 4251-4259.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Pu, W.T., Ishiwata, T., Juraszek, A.L., Ma, Q., and Izumo, S. (2004). GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. Dev Biol *275*, 235-244.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Sanges, R., Cordero, F., and Calogero, R.A. (2007). oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language. Bioinformatics *23*, 3406-3408.

Sheaffer, K.L., Kim, R., Aoki, R., Elliott, E.N., Schug, J., Burger, L., Schubeler, D., and Kaestner, K.H. (2014). DNA methylation is required for the control of stem cell differentiation in the small intestine. Genes Dev *28*, 652-664.

Silberg, D.G., Sullivan, J., Kang, E., Swain, G.P., Moffett, J., Sund, N.J., Sackett, S.D., and Kaestner, K.H. (2002). Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. Gastroenterology *122*, 689-696.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

Sodhi, C.P., Li, J., and Duncan, S.A. (2006). Generation of mice harbouring a conditional loss-of-function allele of Gata6. BMC Dev Biol *6*, 19.

Stringer, E.J., Duluc, I., Saandi, T., Davidson, I., Bialecka, M., Sato, T., Barker, N., Clevers, H., Pritchard, C.A., Winton, D.J.*, et al.* (2012). Cdx2 determines the fate of postnatal intestinal endoderm. Development *139*, 465-474.

Sturley, S.L., and Hussain, M.M. (2012). Lipid droplet formation on opposing sides of the endoplasmic reticulum. J Lipid Res *53*, 1800-1810.

Suh, E., Chen, L., Taylor, J., and Traber, P.G. (1994). A homeodomain protein related to caudal regulates intestine-specific gene transcription. Molecular and cellular biology *14*, 7340-7351.

Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development *132*, 2685-2695.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

van Wering, H.M., Bosse, T., Musters, A., de Jong, E., de Jong, N., Hogen Esch, C.E., Boudreau, F., Swain, G.P., Dowling, L.N., Montgomery, R.K.*, et al.* (2004). Complex regulation of the lactase-phlorizin hydrolase promoter by GATA-4. Am J Physiol Gastrointest Liver Physiol *287*, G899-909.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell *19*, 713-726.

Verzi, M.P., Shin, H., Ho, L.L., Liu, X.S., and Shivdasani, R.A. (2011). Essential and redundant functions of caudal family proteins in activating adult intestinal genes. Mol Cell Biol *31*, 2026-2039.

Verzi, M.P., Shin, H., San Roman, A.K., Liu, X.S., and Shivdasani, R.A. (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol Cell Biol *33*, 281-292.

Yamada, T., and Alpers, D.H. (2009). Textbook of gastroenterology, 5th edn (Chichester, West Sussex ; Hoboken, NJ: Blackwell Pub.).

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L.*, et al.* (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642-654.

Chapter 5: HNF1A binds widely and partners with CDX2 to control cell differentiation in the intestinal epithelium

Contributions: A. K. San Roman and R. A. Shivdasani conceived of the project. R. Sulahian performed HNF1A ChIP-seq in mouse villus cells and collaborated on analysis. A. K. San Roman conducted all other experiments and data analysis. A. K. San Roman wrote this chapter.

Summary

Transcription factor complexes are required for proper gene expression during cell differentiation, a process critical to the rapid turnover of the intestinal epithelium. Evidence from *in vitro* models of cell differentiation suggests that the transcription factor HNF1A may be important during this process. Here, we interrogate the role of HNF1A in cell differentiation of the intestinal epithelium *in vivo*. Chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-seq) revealed a large number of HNF1A binding sites throughout the genome, mostly at promoters and enhancers containing marks of active chromatin. Paradoxically, mice without *Hnf1a* have minimal gene expression changes compared to the loss of the master regulator CDX2, which occupies fewer genomic regions and overlaps with HNF1A at thousands of sites. To understand their cooperation in controlling cell differentiation, we mated *Cdx2* conditional knockout and *Hnf1a* knockout mice and found that they work together to maintain the structural integrity of the villus epithelium and signaling with the lamina propria. These results add another layer of complexity to the growing characterization of intestinal enhanceosomes.

Introduction

The coordinate regulation of gene expression changes during cell differentiation is carried out on a global level by the action of chromatin remodeling and sequence-specific transcription factors (TF), which may act in various complexes that define transcriptional modules. The adult mouse intestinal epithelium serves as a good model to study

mechanisms of cell differentiation, because of its rapid renewal of physiologically important differentiated cell populations. Previous work in our lab identified several TFs that might play a role in the transition from progenitor to differentiated cells by looking for DNA binding motifs enriched in enhancers that become active in differentiated cells (Verzi et al., 2010). Amongst the top three TF binding motifs were CDX2, a homeodomain TF specifically expressed in the intestinal epithelium in the adult (Silberg et al., 2000), and HNF4A, a nuclear hormone receptor superfamily zinc-finger TF expressed throughout the endoderm (Sladek et al., 1990). The binding of these two TFs were subsequently characterized and identified at thousands of H3K4me2-marked active enhancer regions, many of which contain binding of both factors (Verzi et al., 2013). Additionally, the combined loss of *Cdx2* and *Hnf4a* in the intestine causes large gene expression changes and rapid loss of mature absorptive enterocytes in contrast to mild phenotypes when either factor is removed individually (**Chapter 4**). The final and most significant motif identified in the *in vitro* differentiated cells was HNF1 (Verzi et al., 2010), which we did not initially pursue due to difficulties performing ChIP in primary mouse intestinal cells, despite using a previously validated antibody that performed well in human tissues (Odom et al., 2004). Now, thanks to improved reagents and fixation methods, we seek to understand the contribution of this factor to the regulation of gene expression in differentiated cells of the mouse intestinal villus and its interaction with the TFs we have already examined.

The canonical HNF1 motif is recognized by two closely related proteins, HNF1A and HNF1B. The structure of the proteins is similar; both contain N-terminal activation domains, a POU box-specific domain which confers sequence specificity on binding, and

a variant homeodomain containing an unusual linker region between conserved α-helices (Cereghini, 1996). HNF1 proteins can bind DNA as either homo- or heterodimers and have been shown to interact with components of the enhanceosome including chromatin remodelers through their activation domains (Cereghini, 1996; Soutoglou et al., 2000; Soutoglou et al., 2001). HNF1 is known to regulate some intestinal genes (Gautier-Stein et al., 2006; Gregory et al., 2004) and though loss of either subunit, HNF1A or HNF1B, alone causes few intestinal defects, *Hnf1a;Hnf1b* compound mutant mice die of defects in enterocyte maturation and water absorption (D'Angelo et al., 2010). There is some evidence indicating that HNF1 cooperates with CDX2 to control a few specific intestinal loci (Gregory et al., 2004). In addition, studies of HNF1 and HNF4A have shown co-regulation of each other and of gene expression in several tissues (Boj et al., 2010; Eeckhoute et al., 2004; Odom et al., 2004; Rowley et al., 2006). Thus, we hypothesize that HNF1 plays a role in the regulation of gene expression in differentiated cells of the intestinal epithelium and that it interacts with other transcription factors already characterized, including CDX2 and HNF4A.

To evaluate these possibilities, we profiled the binding of HNF1A genome wide using ChIP-seq. We find that HNF1A binds at functional genomic regions marked by histone modifications correlating with active chromatin. Moreover, many HNF1A sites overlap with CDX2 as well as HNF4A sites throughout the genome. To understand combinatorial regulation, we generated compound inducible mutant mice that lack *Cdx2* and *Hnf1a* in the adult intestine, uncovering a requirement for interactions between CDX2 and HNF1A in controlling the structural integrity of the villus epithelium.

Experimental Procedures

Mice

Cdx2$^{Fl/Fl}$, *Hnf1a*$^{-/-}$, *Hnf4a*$^{Fl/Fl}$, and transgenic *VillinCre*^{ERT2} mice were described previously (el Marjou et al., 2004; Hayhurst et al., 2001; Lee et al., 1998; Pu et al., 2004; Sodhi et al., 2006; Verzi et al., 2010). *Cdx2*$^{Fl/Fl}$, *Hnf1a*$^{-/-}$ and *VillinCre*^{ERT2} mice were mated to generate compound *Cdx2*$^{Fl/Fl}$*;Hnf1a*$^{-/-}$*;VillinCre*^{ERT2} mice. To activate Cre recombinase, mice received intraperitoneal (IP) injections of 1 mg tamoxifen (TAM, Sigma) in sunflower oil (Sigma) daily for 4-5 days. To account for dwarfism, *Hnf1a*$^{-/-}$ and *Hnf1a*$^{-/-}$*;Cdx2*^{del} mice received 0.2 mg TAM per 5 g of body weight, to a maximum dose of 1 mg. Mice were weighed daily and euthanized when the first *Hnf1a*$^{-/-}$*Cdx2*^{del} mice became moribund, at 10 days after the first TAM injection. Controls were homozygous for conditional alleles, but wild-type for *Hnf1a* and lacking *VillinCre*^{ERT2}. Animal Care and Use Committees at our institutions approved and monitored animal use.

Histochemistry and immunohistochemistry

The proximal (duodenum) or distal (ileum) thirds of the small intestine were fixed overnight in 4% paraformaldehyde, embedded in paraffin and sectioned at 5 μm thickness. Hematoxylin & eosin (H&E), Alcian blue and periodic acid-Schiff staining followed standard procedures. To detect alkaline phosphatase, slides were incubated in NTM solution (2 mM NaCl, 10 mM Tris-HCl pH 9.5, 5 mM $MgCl_2$) for 5 min, followed by nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate (NBT/BCIP ready-to-use tablets; Roche) solution for 15-30 min, and washed in PBS. Immunohistochemistry

was performed as previously described (Verzi et al., 2010) using the following primary antibodies: rabbit anti-Ki67 (Santa Cruz, 1:200), mouse anti-BrdU (AbD Serotec, 1:300), mouse anti-CDX2 (Biogenex, 1:20), rabbit anti-Cleaved Caspase 3 (Cell Signaling, 1:1000), rabbit anti-Lysozyme (Invitrogen, 1:50), or goat anti-HNF1A (Santa Cruz, 1:100). Representative images of histology and IHC from n≥5 mice of each genotype were obtained using an Olympus BX40 light microscope. Addition of scale bars and adjustment for brightness and contrast were performed in Photoshop. To quantify proliferating or differentiated cells, 10 crypts or villi from at least 3 mice were counted and averaged; significance was determined by *t*-test using GraphPad Prism Software with a *P-value* cutoff of 0.05 (*P*-values are indicated in figures).

Villus epithelium harvests for ChIP-seq and RNA-seq

Mouse intestinal epithelium was harvested by incubating fresh jejuna in 5 mM EDTA solution for 45 min, as described previously(Verzi et al., 2013). Villi were separated from crypts by collecting the cells captured on top of a 70 μm filter when the whole epithelium solution is passed through.

ChIP-seq

For HNF1A ChIP, isolated jejunal villus epithelium was cross-linked with 2 mM disuccinimidyl glutarate (DSG, Pierce, catalog no. 20593) in PBS for 45 min, followed by 10 min with 1% formaldehyde (Pierce) in PBS, at room temperature. Cells were lysed and sonicated to obtain fragments between 200-500 bp as previously described (Verzi et al., 2010). ChIP was performed using a 30 μL slurry of equal amounts Protein A and

Protein G magnetic beads in a 1:1 ratio (Dynabeads, Life Technologies, catalog no. 1002D) coupled to one of the following primary antibodies: HNF1A (Cell Signaling, Cat. #12425). ChIP-Seq libraries were prepared using the Rubicon Genomics ThruPLEX-FD Prep Kit (Cat. #R40048-08, QAM-094-002) and sequenced on an Illumina NextSeq 500 (75 bp single end reads).

Sequences were aligned to the *Mus musculus* reference genome build 9 (mm9) using bowtie and binding peaks for HNF1A were identified using MACS (Cistrome, version 1.4.2) with a p-value cutoff of 10^{-5} (Zhang et al., 2008). We used a sequenced input DNA (DSG/Formaldehyde cross-linked and sonicated) library as background when calling the MACS peaks. Wiggle traces for HNF1A were normalized to the sample with the largest number of uniquely mapped reads (the $Cdx2^{del}$ sample) for comparison between samples. Significantly enriched TF motifs around 600 bp of the HNF1A ChIP-Seq peak summits were analyzed using SeqPos (Liu et al., 2011). Peaks were associated with the nearest gene at various distances up to 30 kb using GREAT software (McLean et al., 2010). H3K4me2 ChIP seq data (Verzi et al., 2013) was re-analyzed using NPS and wiggles were normalized based on the average signal of the top 1% of regions in the sample with the most reads.

RNA analysis

RNA was isolated from jejunal villus using TRIzol reagent (Invitrogen) and the RNeasy kit (Qiagen), followed by treatment with the Turbo DNA-free kit (Ambion) to remove genomic DNA. For global gene expression analysis, total RNA from 3

independent biological replicates per genotype was used to make RNA-seq libraries using the TruSeq RNA Sample Preparation Kit v2 (Illumina), following the manufacturers instructions. An Illumina NextSeq 500 was used to sequence 75 base pair single end reads, which were aligned to the *Mus musculus* reference genome build 9 (mm9). We used the Tuxedo software package to align reads, assemble transcripts and perform differential expression analysis using an FDR of 0.05. Biovenn was used to generate proportional Venn diagrams(Hulsen et al., 2008). GENE-E (Broad Institute) was used for hierarchical clustering and K-means clustering of RNA-Seq samples, using Pearson correlation.

Gene ontology analysis

DAVID functional annotation clustering for gene ontology analysis was performed using medium classification stringency and default options (Huang da et al., 2009). Clusters with significant enrichment scores (>1.3) were considered in the analysis(Huang da et al., 2009). To list the clusters, we selected a representative GO term from the cluster, and when similar annotation clusters recurred in the list, we represent only the one with the highest score.

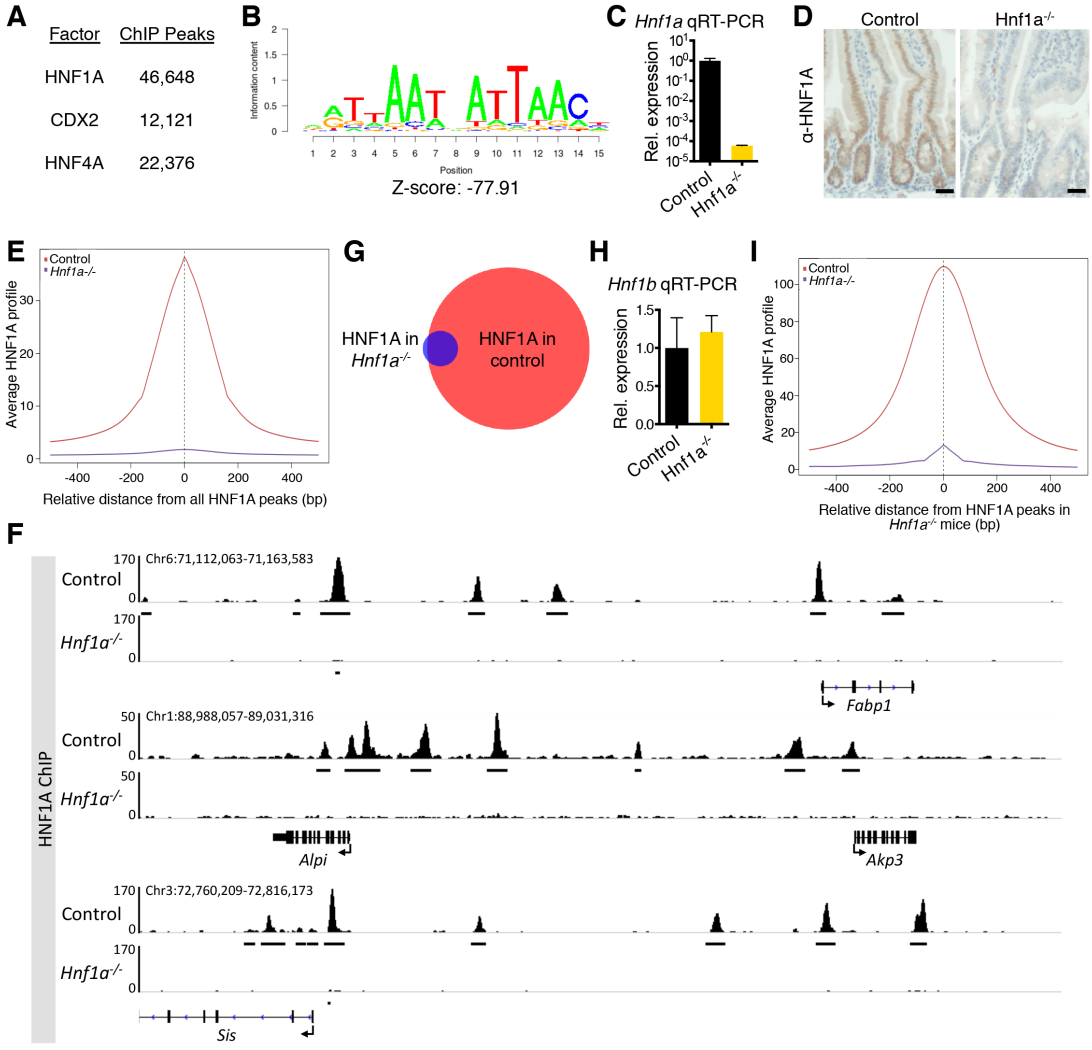
Results

HNF1A binds widely in intestinal villus cells

To understand the role of HNF1A in the intestinal epithelium, we profiled its binding genome-wide in jejunal villus, where we had previously performed CDX2 and HNF4A ChIP-seq. Following sequencing, 46,648 high-confidence HNF1A binding peaks were identified using the MACS algorithm (Zhang et al., 2008). Interestingly, this number was much larger than the number of peaks we had previously identified for CDX2 (12,121) or HNF4A (22,376) (**Figure 5.1A**). With such a large number of peaks we questioned antibody specificity, which we addressed in two ways. First, we analyzed the motifs around the summits of the 5,000 tallest peaks, which returned the HNF1A motif as the top hit with a very significant Z-score of -77.91 (**Figure 5.1B**). Second, we performed an HNF1A ChIP-seq experiment in jejunal villus isolated from *Hnf1a*$^{-/-}$ mice, which completely lacks HNF1A mRNA and protein (**Figures 5.1C,D**). In the *Hnf1a*$^{-/-}$ sample, there was an almost complete reduction of sequence tags in the regions identified as HNF1A peaks in wild-type mice, as seen by aggregate plots (**Figure 5.1E**) and in individual regions (**Figure 5.1F**). When we analyzed the HNF1A ChIP-seq data from *Hnf1a*$^{-/-}$ mice using MACS software, 2,328 weak peaks were identified, which almost entirely overlap with the peaks identified in wild-type mice (**Figure 5.1G**). These residual peaks could be due to the antibody recognizing HNF1B, which binds to the same motif and remains expressed in the *Hnf1a*$^{-/-}$ epithelium (**Figure 5.1H**). However, we note that the average intensity of these peaks was very low compared to the wild-type signal in these regions (**Figure 5.1I**), indicating relative specificity of the antibody.

Figure 5.1. HNF1A binds to many sites throughout the genome of villus cells. (A) Table of number of binding peaks called using the MACS algorithm from HNF1A ChIP-Seq, compared to CDX2 and HNF4A. HNF1A peaks outnumber both CDX2 and HNF4A. **(B)** The HNF1A motif is the highest scoring at the HNF1A ChIP-seq peak summits, with a very significant Z-score. **(C)** Quantitative RT-PCR for *Hnf1a* confirms absence of *Hnf1a* mRNA expression in the *Hnf1a*$^{-/-}$ mice. **(D)** IHC of control and *Hnf1a*$^{-/-}$ intestine for HNF1A demonstrates loss of protein. **(E)** Aggregate profile of HNF1A binding in control (red line) and *Hnf1a*$^{-/-}$ (blue line) mice, centered on HNF1A peak summits from control mice. **(F)** Wiggle traces from control mice reveal many strong HNF1A binding peaks in intergenic, intronic and promoter regions. Traces from *Hnf1a*$^{-/-}$ mice reveal that the signal is almost completely reduced, indicating the antibody is specific. The Y-axis represents number of sequencing tag counts; samples were normalized for sequencing depth. Arrows at genes represent the direction of transcription and the genomic region displayed is noted in the upper-left of each region; Chr, chromosome. **(G)** Venn diagram of HNF1A peaks from control mice (red; 46,648 peaks) and *Hnf1a*$^{-/-}$ mice (blue; 2,328 peaks) shows that the peaks remaining in the knockout almost entirely overlap with those from the control. **(H)** Quantitative RT-PCR for *Hnf1b* reveals that its expression remains unchanged in *Hnf1a*$^{-/-}$ mice compared to controls. **(I)** Aggregate plot centered on the peaks called in *Hnf1a*$^{-/-}$ mice, with signal from control (red) and *Hnf1a*$^{-/-}$ (blue) mice indicate that even at the few peaks called in the *Hnf1a*$^{-/-}$ mice, the peak height is <10% of the control.

Figure 5.1 (Continued)



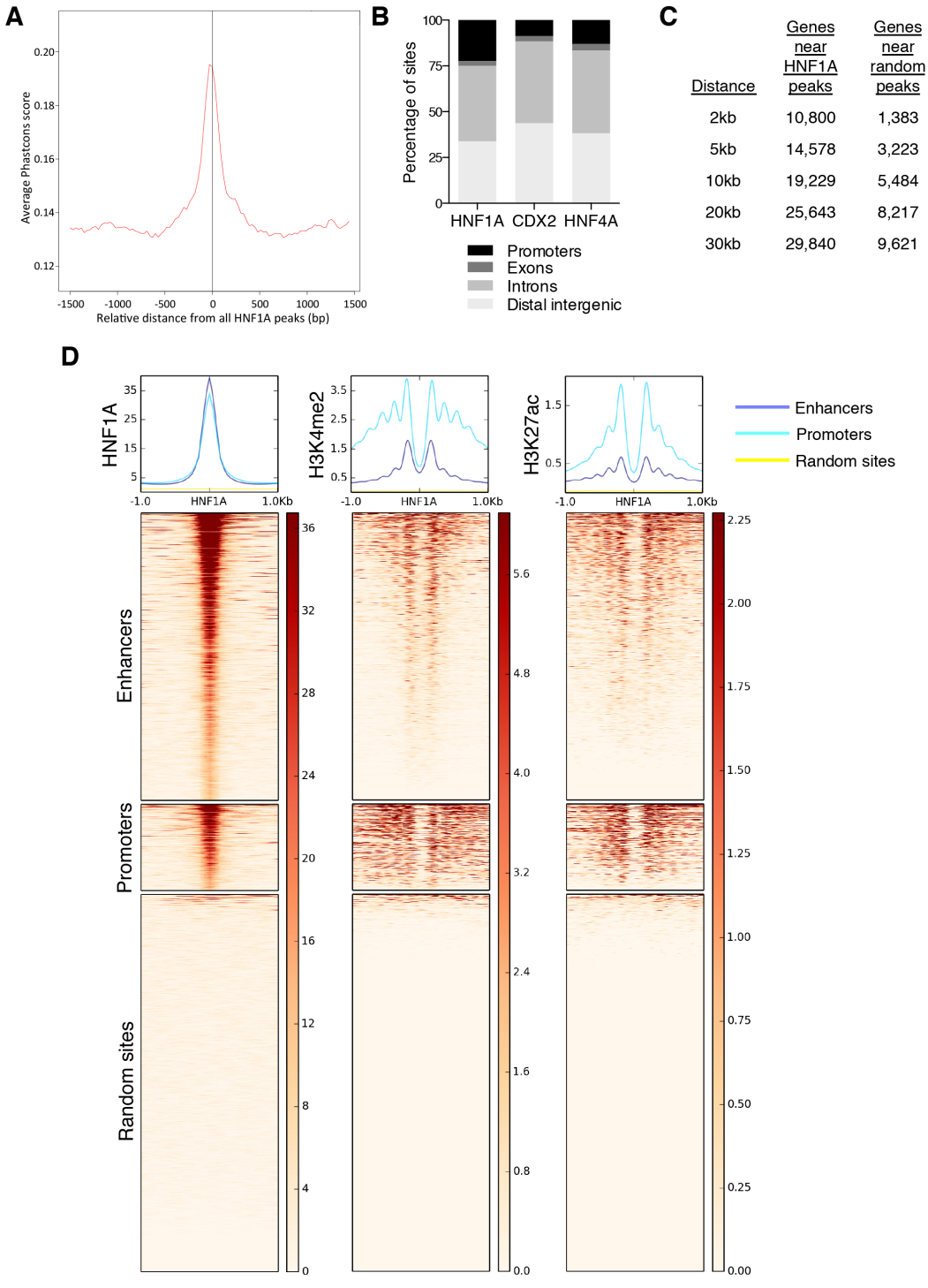
HNF1A binds at functional genomic regions

Once we established confidence in the HNF1A peaks, we wanted to know whether they were in functional genomic regions, which we addressed in several ways. First, we looked at whether regions of HNF1A binding are evolutionarily conserved, which is a characteristic of many functional cis-regulatory elements (Loots et al., 2000; Siepel et al., 2005). We profiled the average PhastCons score at HNF1A peaks, which is a measure of the probability that the region is evolutionarily conserved (Siepel et al., 2005), and found high conservation at these sites (**Figure 5.2A**). Next, we analyzed the distribution of HNF1A peaks to promoters, exons, introns and distal intergenic regions. Most HNF1A binding sites occur in distal intergenic and intronic regions, although there was also a significant fraction of sites found in promoters. Interestingly, this fraction was larger than what we previously obtained for CDX2 and HNF4A (**Figure 5.2B**), which suggests additional functional roles for HNF1A outside of interactions with these TFs. We were then interested in understanding whether the peaks clustered around several genes or were widely distributed throughout the genome. To do this, we used GREAT software (McLean et al., 2010) to associate each HNF1A peak with the single nearest gene within various distances (2kb (promoter), 5kb, 10kb, 20kb and 30kb). We found that the peaks were widely distributed throughout the genome and by the 30kb distance, most annotated genes were associated with an HNF1A peak (**Figure 5.2C**). This association was not simply due to the number of HNF1A peaks, as a matched set of randomly distributed peaks were associated with significantly fewer genes at all distances (**Figure 5.2C**; $P<0.0001$; Chi-square test).

Another way we assessed the functional relevance was by associating HNF1A binding sites with chromatin marks that correspond to active enhancers and promoters. Of these, H3K4me2 was of particular interest because the HNF1A motif was highly enriched in regions that gain H3K4me2 signal during a cell culture model of intestinal differentiation (Verzi et al., 2010). To investigate whether HNF1A is found in H3K4me2-marked regions we profiled the H3K4me2 ChIP-seq signal (Verzi et al., 2013) surrounding HNF1A peak summits in promoters and enhancers. We found that there were high average levels of H3K4me2 surrounding most HNF1A summits at promoters and many of the enhancers (**Figure 5.2D**). The lack of H3K4me2 signal in the center of the graph (at the HNF1A summits) is consistent with a labile central nucleosome displaced by the binding of HNF1A, possibly as a complex with other TFs. We performed the same analysis using the H3K27ac mark, which also correlates with active transcription and saw similar results (**Figure 5.2D**). As a control, we included a matched set of summits that were shuffled randomly throughout the genome and observe no H3K4me2 or H3K27ac signal (**Figure 5.2D**). Together, this data suggests that HNF1A may regulate a large number of genes through its binding.

Figure 5.2. HNF1A binds at conserved, functional genomic elements, many of which are surrounded by active chromatin marks. (A) HNF1A peaks in control mice are found in highly conserved regions of the genome, as measured by the average PhastCons score. **(B)** Distribution of peak summits amongst genomic elements including promoters (-2 kb and +1 kb from TSS), exons, introns and distal intergenic regions in HNF1A, CDX2 and HNF4A. In all the TFs, most of the binding sites were found in introns and distal intergenic sites, but a higher percentage of HNF1A peaks are found in promoters compared to CDX2 or HNF4A, with minimal binding in exons. **(C)** Associations of HNF1A peaks from control mice or a matched set of randomly distributed peaks with the single nearest gene within the indicated distances. Fisher's exact test between the two samples at each distance indicates the difference is significant ($P<0.0001$). **(D)** Heatmaps centered on HNF1A binding sites in enhancers, promoters and random sites (matched to the total number of HNF1A peaks) with signal of HNF1A (left), H3K4me2 (center), and H3K27ac (right) ChIP-seq shows well-positioned marked nucleosomes surrounding the HNF1A binding sites in both promoters (aqua), and enhancers (purple) but not around random sites (yellow), which is also visualized in the aggregate plots above.

Figure 5.2 (Continued)



HNF1A dysregulates expression of a surprisingly small number of genes

The ChIP-seq data implies that HNF1A could be regulating many genes due to its binding at functional elements containing marks of active transcription, however a previous gene expression analysis only identified a modest number of significantly changed genes upon loss of *Hnf1a* (Lussier et al., 2010). We decided to repeat this gene expression analysis using RNA-seq from control and *Hnf1a*$^{-/-}$ mice. In our hands, loss of *Hnf1a* caused a significant ($Q < 0.05$) up-regulation of 444 genes and down-regulation of 560 genes (**Figure 5.3A**). To ascertain which of these are direct targets of HNF1A, we combined ChIP-seq (using a distance cut-off of 30 kb) and RNA-seq data. We found that 462 of the down-regulated and 338 of the up-regulated genes had at least 1 HNF1A binding site nearby, and many had several, which were abolished in *Hnf1a*$^{-/-}$ mice (see examples in **Figure 5.3B**). Next, we analyzed the gene ontology terms for the up and down regulated genes upon *Hnf1a* loss. Interestingly, amongst the up-regulated genes we see terms characteristic of ISCs (**see Figure 3.1C**), including mitosis and control of the microtubule cytoskeleton (**Figure 5.3C**). The down-regulated genes are involved in the endoplasmic reticulum, cell membrane and oxidation-reduction.

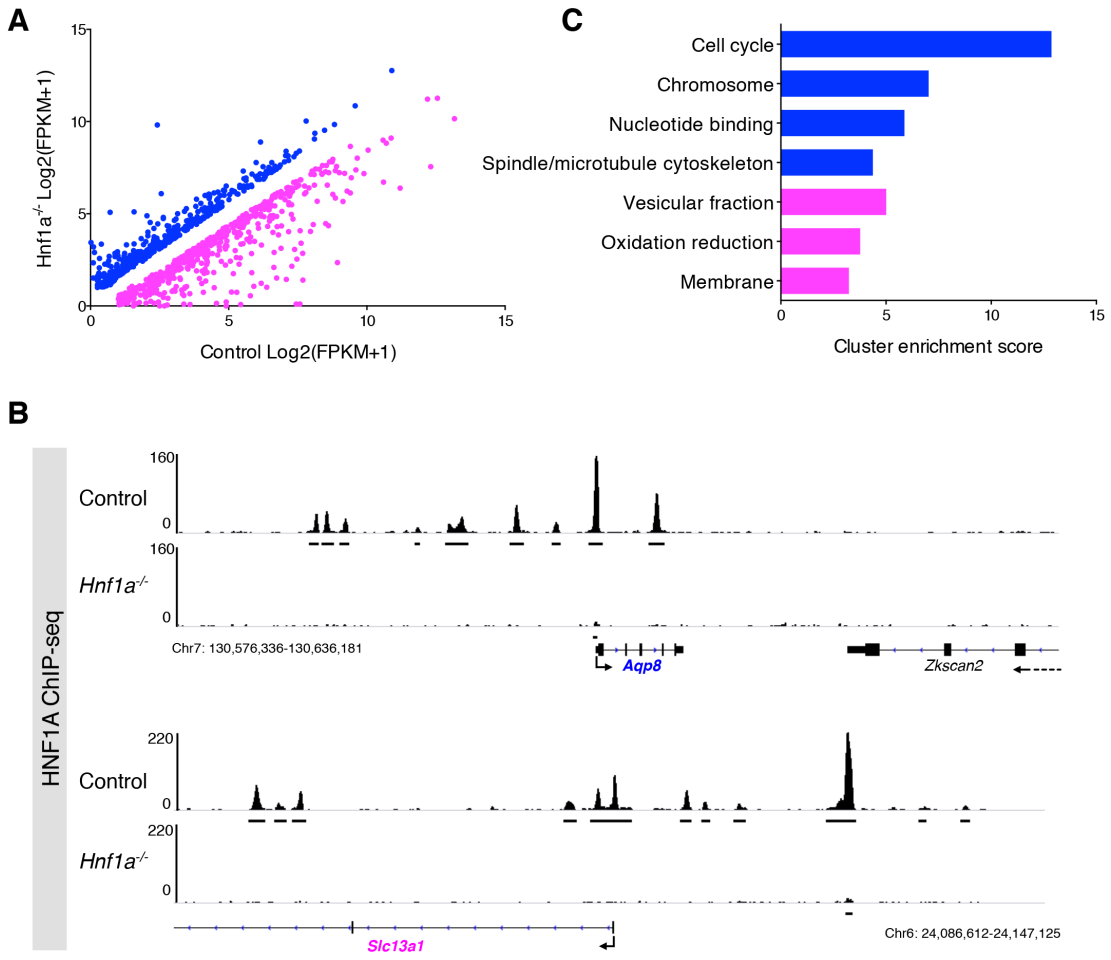


Figure 5.3. HNF1A regulates the expression of fewer than expected genes. (A) RNA-seq data comparing villus gene expression in *Hnf1a*$^{-/-}$ mice to controls. Blue dots indicate genes up-regulated and pink dots indicate genes down-regulated in *Hnf1a*$^{-/-}$ mice. **(B)** HNF1A ChIP-seq signals from control and *Hnf1a*$^{-/-}$ mice around *Aqp8*, which is significantly up-regulated and *Slc13a1,* which is significantly down-regulated in *Hnf1a*$^{-/-}$ mice. The Y-axis represents number of sequencing tag counts; samples were normalized for sequencing depth. Arrows at genes represent the direction of transcription and the genomic region displayed is for each region; Chr, chromosome. **(C)** Gene ontology analysis shows terms up (blue) and down (pink) regulated upon loss of *Hnf1a.*

HNF1A binding sites significantly overlap with CDX2 and HNF4A

As many fewer genes were dysregulated upon *Hnf1a* loss than would have been predicted by its large number of binding sites, we sought to understand whether HNF1A might partner with other TFs that have been characterized in the intestine, including CDX2 and HNF4A. To understand whether HNF1A binding sites overlap with CDX2 and HNF4A, we calculated the number of CDX2 and HNF4A peaks that overlap within 300 bp of the summit of HNF1A peaks (**Figure 5.4A**). Almost all of the sites we had previously identified as being shared between HNF4A and CDX2 were also shared by HNF1A (**Figure 5.4B**). More than half of the remaining HNF4A sites overlapped with HNF1A without CDX2. In addition, 2,822 CDX2 sites overlapped with HNF1A without HNF4A. A very large number of peaks (30,654) do not have a nearby HNF4A or CDX2 binding site, suggesting that there may be other TF partners for HNF1A (**Figure 5.4B**). By computing the overlaps with a set of matched random sites instead of HNF1A and comparing with our observed results, we found that the co-bound sites are unlikely to occur by chance ($P<0.0001$ for both HNF1A with CDX2 and with HNF4A; Fisher's exact test; **Figures S5.1A,B**).

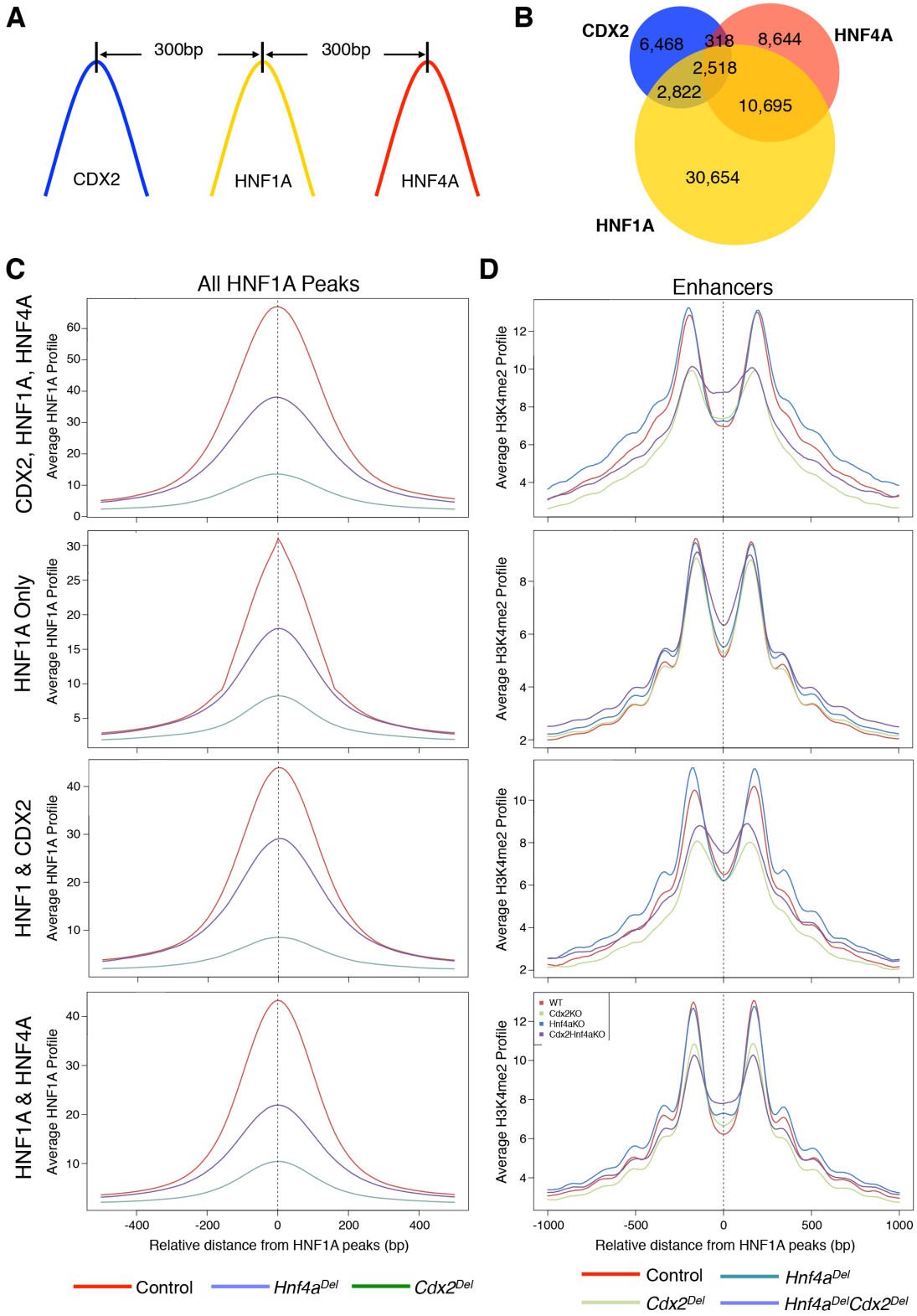
Previously, we found that HNF4A depends on the presence of CDX2 at co-bound sites indicating that there is a hierarchy amongst TFs (Verzi et al., 2013). Thus, we wondered if HNF1A relies on the presence of HNF4A or CDX2 at co-bound regions. To test this, we performed ChIP-seq for HNF1A in *Cdx2*^{del} *and Hnf4a*^{del} jejunal villi and analyzed the signal within 4 categories of sites: those shared by all 3 factors, shared by CDX2 and HNF1A, shared by HNF4A and HNF1A, and HNF1A only. Interestingly, we find that the average HNF1A binding signal is highly enriched in the regions where all 3

factors bind together (**Figure 5.4C**, top). Notably, throughout all 4 categories of sites, the HNF1A signal is reduced by ~2-fold in the *Hnf4a*^{del} mice, while in the *Cdx2*^{del} mice it is reduced ~4-6 fold depending on the region (**Figures 5.4C**; **S5.1C**). This analysis demonstrates that loss of CDX2 and HNF4A have a global effect on the binding of HNF1A, even at sites where it binds without the presence of either TF.

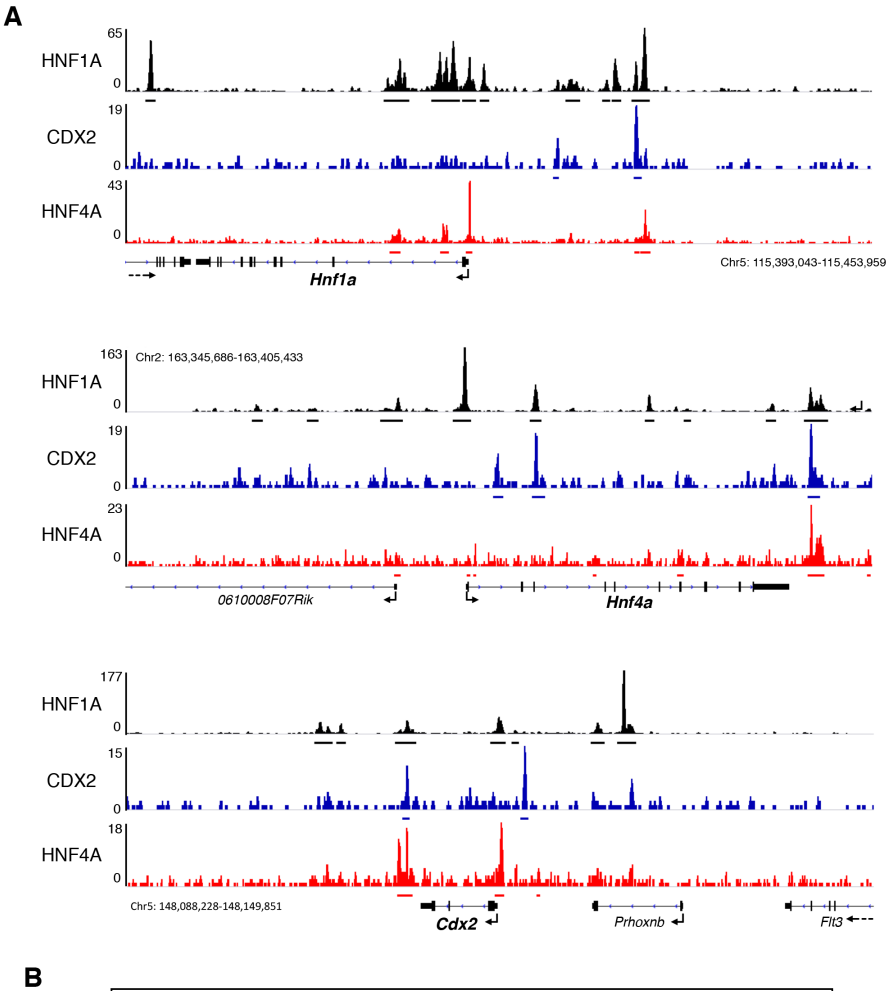
Reduction of HNF1A binding throughout the genome could be due to CDX2 or HNF4A regulating the accessibility of these sites to HNF1A, and in their absence these sites are now "closed". Alternatively, the loss of *Cdx2* or *Hnf4a* could reduce overall HNF1A levels leading to this difference. To test the first hypothesis, we examined H3K4me2 positioning in the *Cdx2*^{del}*, Hnf4a*^{del}*, and Hnf4a*^{del}*Cdx2*^{del} mice at the same 4 categories of sites as above, focusing on enhancers, since the majority of HNF4A and CDX2 binding occurs in these regions. This analysis reveals that H3K4me2 signals are affected by the loss of CDX2, and even more by the simultaneous loss of Cdx2 and HNf4a together at the sites where all 3 factors bind (**Figure 5.4D**). In regions where HNF4A binds with HNF1A there are mild effects on H3K4me2, while it is more severe in regions with CDX2 and HNF1A co-binding, although neither are as severe as in the sites bound by all 3 factors, where the loss of *Cdx2* alone and simultaneous loss of *Hnf4a* and *Cdx2* disrupts positioning of the nucleosomes and the magnitude of the signal. Interestingly, the peaks where HNF1A binds alone are minimally affected in all of the knockouts. Thus, some of the HNF1A binding loss at the commonly bound sites could be explained by the remodeling of chromatin, but it does not explain the loss at sites occupied by HNF1A only.

Figure 5.4. HNF1A binding sites overlap with CDX2 and HNF4A and are reduced upon knockout of either factor. (A) Significant overlaps were measured by assessing TF co-binding within 300 bp of peak summits. **(B)** Venn diagram showing peak overlaps between CDX2, HNF4A and HNF1A. **(C-D)** Aggregate plots centered on regions where all three factors bind, with HNF1A only, with both HNF1A and CDX2 and with HNF1A and HNF4A, from top to bottom. The aggregate HNF1A **(C)** or H3K4me2 **(D)** signal from ChIP-seq in various genetic backgrounds (legends corresponding to colored lines are represented below the figures) was plotted around the HNF1A peaks.

Figure 5.4 (Continued)



To ascertain whether HNF1A levels are affected by loss of *Cdx2* or *Hnf4a,* we looked at the binding of CDX2 and HNF4A around the HNF1A gene. We find that in addition to HNF1A, there are several sites for CDX2 and HNF4A nearby (**Figure 5.5A**). To ascertain whether *Hnf1a* gene expression is affected, we examined gene expression data in $Cdx2^{del}$ villus (by RNA-seq), or $Hnf4a^{del}$ villus (by microarray; (Verzi et al., 2013)), compared to controls. *Hnf1a* expression was significantly down-regulated by almost two-fold in $Cdx2^{del}$, which may explain the reduction in HNF1A binding (**Figure 5.5B**). Moreover, *Hnf1a* or *Hnf4a* knockout did not perturb expression of each other or *Cdx2*, despite binding near both these genes (**Figures 5.5A,B**).



B

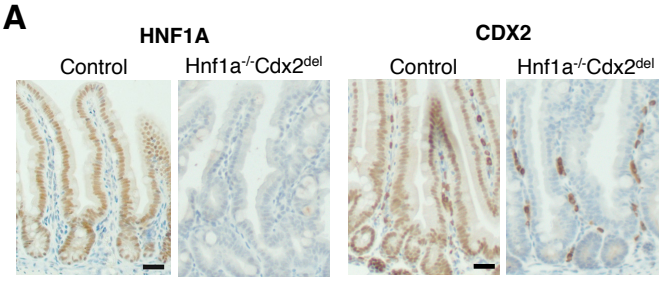
Genotype	*Hnf1a*	*Cdx2*	*Hnf4a*
Hnf1a$^{-/-}$	-4.57	NS	NS
Cdx2^{del}	-0.91	NS (exon 2 deleted)	-0.96
Hnf4a^{del}	NS	NS	-1.39

Figure 5.5. HNF1A, CDX2 and HNF4A bind near each gene and CDX2 regulates *Hnf1a* gene expression. (A) ChIP-seq traces for HNF1A (black), CDX2 (blue) and HNF4A (red) around the *Hnf1a* (top), *Cdx2* (middle), and *Hnf4a* (bottom) genes indicates that each factor binds to all regions. **(B)** Significant gene expression changes are represented relative to controls (log2 fold change) in *Hnf1a*$^{-/-}$, *Cdx2*^{del}, (RNA-seq) and *Hnf4a*^{del} (microarray; (Verzi et al., 2013)) mice. *Hnf1a* expression is significantly affected in *Cdx2*^{del} mice. NS, not significant.

Morphologic and functional deficits in Hnf1a$^{-/-}$Cdx2^{del} intestines

To examine the joint functions of HNF1A and CDX2 on gene expression *in vivo*, we crossed mice carrying a germline *Hnf1a* null allele with *Cdx2$^{Fl/Fl}$;Villin-Cre^{ERT2}* mice and compound mutant animals showed loss of both TFs throughout the small intestine (**Figure 5.6A**). RNA-seq analysis of single and compound mutants versus controls indicated that unlike *Hnf1a*, loss of *Cdx2* causes large gene expression changes, dysregulating thousands of genes (**Figure 5.6B**). Interestingly, changes in gene expression were only mildly increased in the compound *Hnf1a$^{-/-}$Cdx2^{del}* mice, compared to controls. In fact, when we compare compound mutants with *Cdx2^{del}* mice, we only find a couple hundreds of genes altered (**Figure 5.6B**). As CDX2 regulates the expression of *Hnf1a*, it is possible that some HNF1A-regulated genes are already perturbed in the *Cdx2^{del}* mice, minimizing the observed changes in compound mutants.

A



B

Genotype	Down	Up
Hnf1a$^{-/-}$	560	444
Cdx2^{del}	1,684	1,599
Hnf1a$^{-/-}$*Cdx2*^{del}	1,778	1,874
Hnf1a$^{-/-}$*Cdx2*^{del} *vs Cdx2*^{del}	271	253

Figure 5.6. Loss of *Hnf1a* and *Cdx2* perturbs more genes than *Cdx2* loss alone. (A) Immunohistochemistry for HNF1A (left) and CDX2 (right) in control and compound *Hnf1a*$^{-/-}$*Cdx2*^{del} mutant intestines. **(B)** Summary of gene expression changes in *Hnf1a*$^{-/-}$, *Cdx2*^{del}, and *Hnf1a*$^{-/-}$*Cdx2*^{del} mice compared to controls. Bottom, *Hnf1a*$^{-/-}$*Cdx2*^{del} mice are compared to *Cdx2*^{del}, indicating that a few hundred genes are more severely affected in the compound mutant mice.

Despite the modest additional gene expression changes, compound mutant mice developed severe diarrhea, lost significantly more weight than *Hnf1a*$^{-/-}$ or *Cdx2*^{del} single mutants (**Figure 5.7A**), and required euthanasia within 10 days of CDX2 loss. Crypt cell proliferation indices in *Hnf1a*$^{-/-}$*Cdx2*^{del} ileum were similar to those in both single mutants (**Figure 5.7B**), but villus abnormalities were striking and distinctive. Ileal villi in *Hnf1a*$^{-/-}$*Cdx2*^{del} mice were uniformly narrow and varied considerably in height, though average cell numbers were similar to *Cdx2*^{del} villi (**Figure 5.7C**; **S5.2A**). Epithelial cells were short and lacked the typical columnar shape; their nuclei were globally disarrayed, varied widely in size and morphology, and the intensity of hematoxylin staining indicated aberrant chromatin distribution (**Figure 5.7C**, magnified to the right). Enterocytes commonly showed apical vacuoles and, similar to *Cdx2*^{del} mice, lacked alkaline phosphatase (**Figure 5.7D**). A consistent fraction of villus goblet cells indicated that, unlike *Hnf4a*^{del}*Cdx2*^{del} mice, enterocyte numbers and proportions were intact (**Figure 5.7D**). Lysozyme, a Paneth-cell marker, was reduced or absent in many crypts (**Figure S5.2B**), as noted previously in *Hnf1a*$^{-/-}$*Hnf1b*^{del} but not in mice lacking HNF1A alone (D'Angelo et al., 2010). Finally, the *Hnf1a*$^{-/-}$*Cdx2*^{del} villus lamina propria was uniformly narrow, with a paucity of mesenchymal cells (**Figure 5.7C**, right). Apoptosis was not increased in the epithelium or lamina propria (**Figure S5.2C**, *right*); rather, the combination of a mesenchymal cell deficit and loss of epithelial cell height made every villus abnormally slender. As *Cdx2* and *Hnf1a* are expressed, and *Villin-Cre*^{ERT2} is active, only in the epithelium, the lamina propria defect is likely indirect. Thus, defects in the *Hnf1a*$^{-/-}$*Cdx2*^{del} ileum differ from those in *Cdx2*^{del} or *Hnf1a*$^{-/-}$ mice. Combined loss of

HNF1A and CDX2 affected enterocytes far more than loss of either TF alone, without a concomitant defect in crypt cell replication.

Villi in *Hnf1a*$^{-/-}$*Cdx2*^{del} duodenum were short, with a hypocellular lamina propria and gross disorganization (**Figure 5.8A**), similar to villi in the ileum (**Figure 5.7C**), but additionally showed prominent epithelial ingrowths. Superficially, this morphology resembled regions of proliferating villus cells when Bone Morphogenetic Protein signaling is inhibited (Haramis et al., 2004) or the transcription factor ASCL2 is forcibly overexpressed (van der Flier et al., 2009). In *Hnf1a*$^{-/-}$*Cdx2*^{del} duodeni, however, these structures lacked proliferating (**Figure 5.8A**), apoptotic or Paneth (**Figure S5.2D**) cells and carried mature, alkaline phosphatase expressing enterocytes and Alcian blue-avid goblet cells (**Figure 5.8A,B**). They appeared in almost every duodenal villus, with some transverse sections giving the appearance of villus fusion (**Figure 5.8B**). These unique aberrant structures were absent in the mutant jejunum or ileum, or in any part of *Cdx2*$^{-/-}$ or *Hnf1a*$^{-/-}$ intestines. Thus, HNF1A and CDX2 co-regulate genes necessary for a structurally intact epithelium.

Figure 5.7. Combined intestinal loss of *Hnf1a* and *Cdx2* is lethal, with ileal defects distinct from those in *Hnf4a*^{del}*Cdx2*^{del} mice. (A) One week after TAM injection, *Hnf1a*$^{-/-}$*Cdx2*^{del} mice weighed significantly less than single mutants or controls. **(B)** Absence of a proliferation deficit in *Hnf1a*$^{-/-}$*Cdx2*^{del} intestinal crypts, as revealed by Ki67 IHC; quantified to right. **(C)** H&E staining of ileal mucosa reveals villi of variable length in *Hnf1a*$^{-/-}$*Cdx2*^{del} mice. Dotted rectangles outline regions magnified to the right, where *Hnf1a*$^{-/-}$*Cdx2*^{del} villi show variable cell and nuclear shape and size. Arrows point to narrowing of the lamina propria space. Alkaline phosphatase staining indicates lack of proper enterocyte maturation in *Cdx2*^{del} and *Hnf1a*$^{-/-}$*Cdx2*^{del} intestines. **(D)** Alcian blue staining shows intact goblet cell differentiation, with proportion of goblet cells quantified to the right. Graphs represent mean +/- SEM. NS, not significant. All scale bars, 30 μm.

Figure 5.7 (Continued)

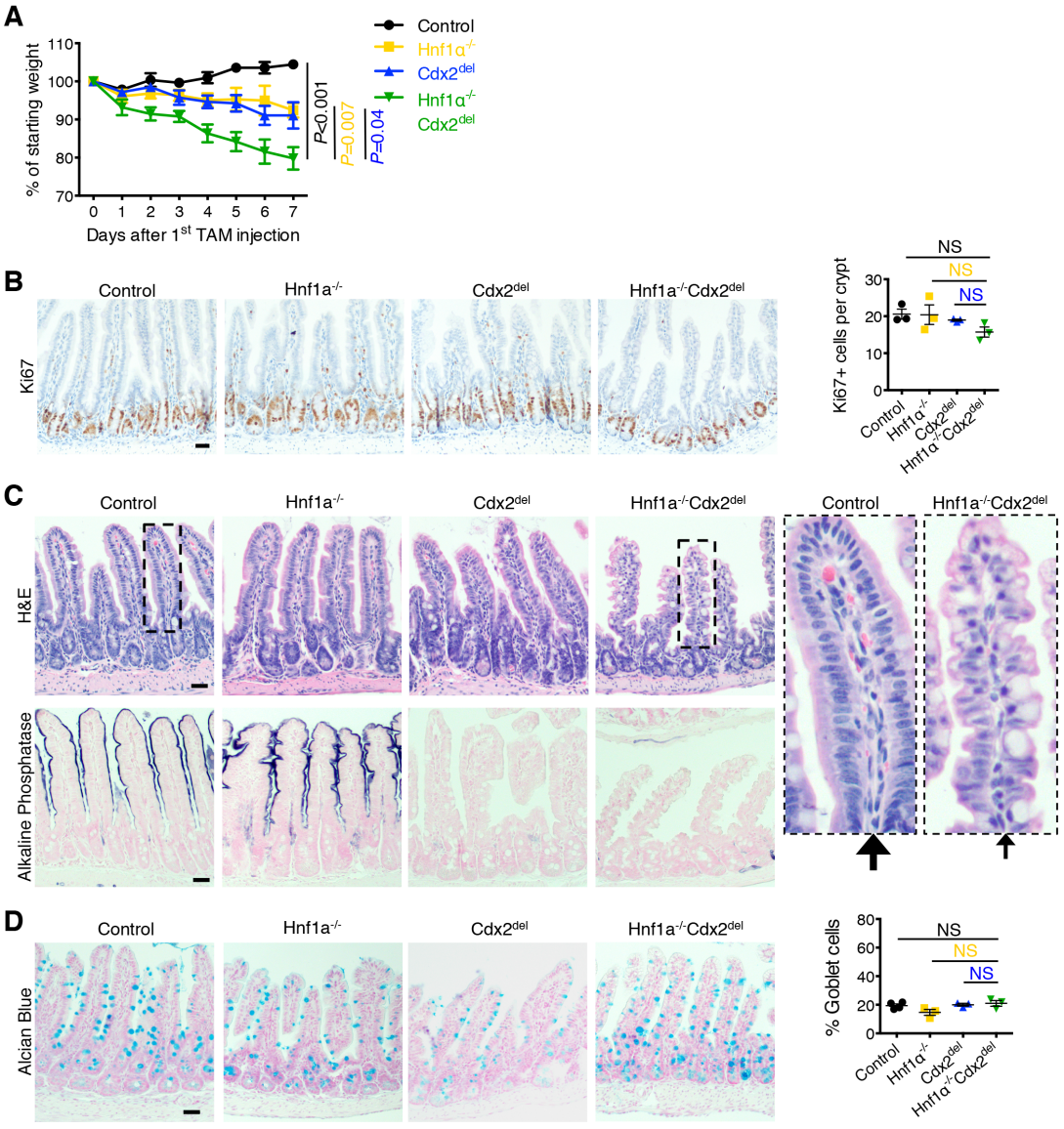
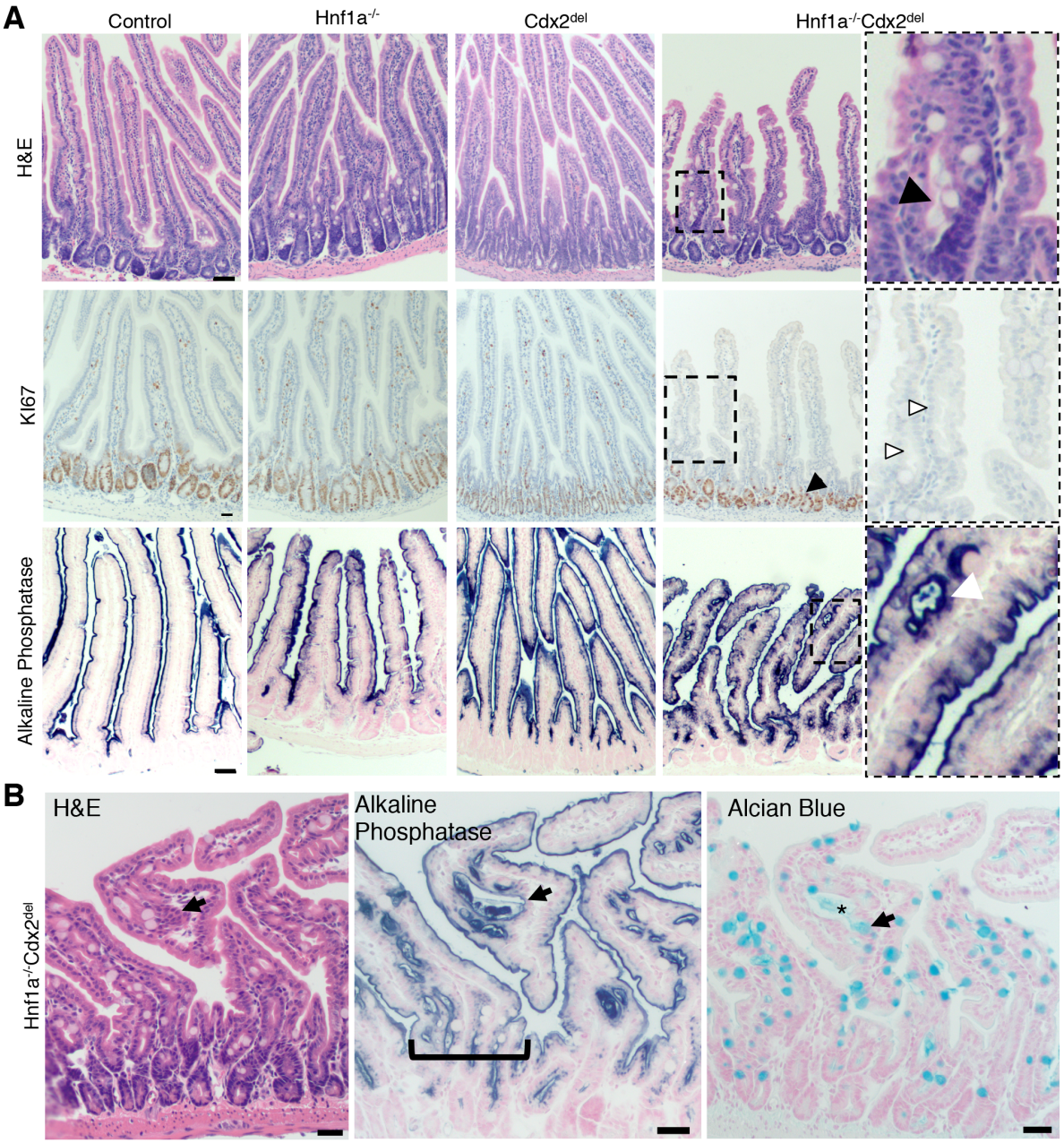


Figure 5.8. Differentiated duodenal villus ingrowths in intestines lacking *Cdx2 and Hnf1a* **(A)** H&E staining shows normal duodenal morphology in control and single mutant intestines, whereas *Hnf1a*$^{-/-}$*Cdx2*^{del} intestines have short villi, reduced lamina propria width, and abundant epithelial ingrowths (arrowhead). Ki67 immunohistochemistry demonstrates absence of proliferation in the villus ingrowths (white arrowheads), distinct from active proliferation in the crypts (black arrowhead). Alkaline phosphatase is present in each case, including within *Hnf1a*$^{-/-}$*Cdx2*^{del} ingrowths and cysts (arrowhead). **(B)** Serial sections of *Hnf1a*$^{-/-}$*Cdx2*^{del} epithelial inclusions, which are visualized best by alkaline phosphatase staining, showing that the anomaly is widespread. Villi give an appearance of fusion (bracket) and ingrowths contain differentiated alkaline phosphatase-positive enterocytes as well as Alcian blue-avid goblet cells (arrows). Center of the cysts often contain mucus stained with Alcian blue (asterisk). All scale bars, 50 μm.

Figure 5.8 (Continued)



Discussion

The results presented here indicate that HNF1A binds widely throughout the genome at functional elements, however it deregulates the expression of few genes upon its loss. One caveat of this study that may contribute to this discrepancy is that HNF1B, which is still expressed in the HNF1A knockout mice, may be able to compensate when *Hnf1a* is disrupted. Unfortunately, we were unable to obtain *Hnf1b* knockout mice for the present study, but redundancy has been supported in previous studies using compound *Hnf1a/b* mutant mice (D'Angelo et al., 2010).

Despite the persistence of HNF1B in our study, combined $Hnf1a^{-/-}Cdx2^{del}$ mice show significantly worse phenotypes compared to single knockouts, indicating that loss of *Cdx2* can unmask the deficits of *Hnf1a* loss. Indeed, combined loss of *Hnf1a* and *Cdx2* uncovered combined contributions towards enterocyte maturation in the ileum and a striking structural defect in the duodenum. It is possible that intestinal cell polarity is disturbed in these mice and leads to these defects, a facet that we would like to investigate further by investigating markers of apical and basal membranes. Interestingly deletion of laminin alpha 5, a component of the basal extracellular matrix, causes structural defects in villi that somewhat resemble the phenotype we observe in the duodenum (Mahoney et al., 2008), suggesting that this may be a useful avenue for follow-up. In addition, removal of *Hnf1a* and *Cdx2* had a prominent effect on the lamina propria throughout the intestine, which must be non-cell autonomous because neither factor is expressed in the lamina propria and the deletion of *Cdx2* only occurs in the epithelial layer. This defect suggests that *Cdx2* and *Hnf1a* may co-regulate Hedgehog signaling, which is essential for gut mesenchymal growth during embryonic development (Mao et al., 2010). Further analysis of the genes disrupted in the compound mutant mice will address this possibility.

Based upon the addition of HNF1A binding data to the growing knowledge of the components of the intestinal enhanceosome, we hope to continue to analyze this data to understand more about the relationships between these TFs *in vivo*. A particular interest in profiling the conditions in which binding of CDX2, HNF1A or HNF4A controls proximal gene expression motivates the combination of various types of data including additional ChIP-seq on other TFs, chromatin modifications, chromatin remodelers and histone modifiers. In the future, we will put significant effort into characterizing these binding sites using computational and bioinformatics methods. We would also like to better define the sets of direct target genes for each TF, with the caveat that at least CDX2 can bind to and modulate expression of both HNF1A and HNF4A.

In sum, this study provides preliminary characterization of combinatorial activities for HNF1A, HNF4A and CDX2 at the level of chromatin binding and evidence for the functional interaction between HNF1A and CDX2 to control intestinal structure and signaling to mesenchymal cells.

Acknowledgements

We thank Sylvie Robine and for sharing *Villin-Cre*$^{ER-T2}$; Marla Perez for mouse husbandry and genotyping; Marian Neutra for assistance characterizing intestinal defects; and the Center for Functional Cancer Epigenomics and the MBCF at Dana-Farber Cancer Institute.

References

Boj, S.F., Petrov, D., and Ferrer, J. (2010). Epistasis of transcriptomes reveals synergism between transcriptional activators Hnf1alpha and Hnf4alpha. PLoS Genet *6*, e1000970.

Cereghini, S. (1996). Liver-enriched transcription factors and hepatocyte differentiation. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology *10*, 267-282.

D'Angelo, A., Bluteau, O., Garcia-Gonzalez, M.A., Gresh, L., Doyen, A., Garbay, S., Robine, S., and Pontoglio, M. (2010). Hepatocyte nuclear factor 1alpha and beta control terminal differentiation and cell fate commitment in the gut epithelium. Development *137*, 1573-1582.

Eeckhoute, J., Formstecher, P., and Laine, B. (2004). Hepatocyte nuclear factor 4alpha enhances the hepatocyte nuclear factor 1alpha-mediated activation of transcription. Nucleic Acids Res *32*, 2586-2593.

el Marjou, F., Janssen, K.-P., Chang, B.H.-J., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186-193.

Gautier-Stein, A., Zitoun, C., Lalli, E., Mithieux, G., and Rajas, F. (2006). Transcriptional regulation of the glucose-6-phosphatase gene by cAMP/vasoactive intestinal peptide in the intestine. Role of HNF4alpha, CREM, HNF1alpha, and C/EBPalpha. J Biol Chem *281*, 31268-31278.

Gregory, P.A., Lewinsky, R.H., Gardner-Stephen, D.A., and Mackenzie, P.I. (2004). Coordinate regulation of the human UDP-glucuronosyltransferase 1A8, 1A9, and 1A10 genes by hepatocyte nuclear factor 1alpha and the caudal-related homeodomain protein 2. Mol Pharmacol *65*, 953-963.

Haramis, A.-P.G., Beghtel, H., van den Born, M., van Es, J., Jonkheer, S., Offerhaus, G.J.A., and Clevers, H. (2004). De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. Science *303*, 1684-1686.

Hayhurst, G.P., Lee, Y.H., Lambert, G., Ward, J.M., and Gonzalez, F.J. (2001). Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. Mol Cell Biol *21*, 1393-1403.

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

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics *9*, 488.

Lee, Y.H., Sauer, B., and Gonzalez, F.J. (1998). Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf-1alpha knockout mouse. Mol Cell Biol *18*, 3059-3068.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., and Frazer, K.A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science *288*, 136-140.

Lussier, C.R., Brial, F., Roy, S.A.B., Langlois, M.-J., Verdu, E.F., Rivard, N., Perreault, N., and Boudreau, F. (2010). Loss of hepatocyte-nuclear-factor-1alpha impacts on adult mouse intestinal epithelial cell growth and cell lineages differentiation. PLoS One *5*, e12378.

Mahoney, Z.X., Stappenbeck, T.S., and Miner, J.H. (2008). Laminin alpha 5 influences the architecture of the mouse small intestine mucosa. J Cell Sci *121*, 2493-2502.

Mao, J., Kim, B.M., Rajurkar, M., Shivdasani, R.A., and McMahon, A.P. (2010). Hedgehog signaling controls mesenchymal growth in the developing mammalian digestive tract. Development *137*, 1721-1729.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Odom, D.T., Zizlsperger, N., Gordon, D.B., Bell, G.W., Rinaldi, N.J., Murray, H.L., Volkert, T.L., Schreiber, J., Rolfe, P.A., Gifford, D.K.*, et al.* (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1378-1381.

Pu, W.T., Ishiwata, T., Juraszek, A.L., Ma, Q., and Izumo, S. (2004). GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. Dev Biol *275*, 235-244.

Rowley, C.W., Staloch, L.J., Divine, J.K., McCaul, S.P., and Simon, T.C. (2006). Mechanisms of mutual functional interactions between HNF-4alpha and HNF-1alpha revealed by mutations that cause maturity onset diabetes of the young. Am J Physiol Gastrointest Liver Physiol *290*, G466-475.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

Silberg, D.G., Swain, G.P., Suh, E.R., and Traber, P.G. (2000). Cdx1 and cdx2 expression during intestinal development. Gastroenterology *119*, 961-971.

Sladek, F.M., Zhong, W.M., Lai, E., and Darnell, J.E., Jr. (1990). Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev *4*, 2353-2365.

Sodhi, C.P., Li, J., and Duncan, S.A. (2006). Generation of mice harbouring a conditional loss-of-function allele of Gata6. BMC Dev Biol *6*, 19.

Soutoglou, E., Papafotiou, G., Katrakili, N., and Talianidis, I. (2000). Transcriptional activation by hepatocyte nuclear factor-1 requires synergism between multiple coactivator proteins. J Biol Chem *275*, 12515-12520.

Soutoglou, E., Viollet, B., Vaxillaire, M., Yaniv, M., Pontoglio, M., and Talianidis, I. (2001). Transcription factor-dependent regulation of CBP and P/CAF histone acetyltransferase activity. EMBO J *20*, 1984-1992.

van der Flier, L.G.v.d., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Beghtel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell *19*, 713-726.

Verzi, M.P., Shin, H., San Roman, A.K., Liu, X.S., and Shivdasani, R.A. (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol Cell Biol *33*, 281-292.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Chapter 6: Discussion

Wnt ligand secretion in the intestinal stem cell niche

In 2007, the field of intestinal biology changed dramatically when the first ISC marker, *Lgr5*, was identified (Barker et al., 2007). This discovery allowed for the study of an ISC population critical for constant regeneration of the intestinal epithelium. Based on data from patients with colorectal cancer, as well as genetic studies in mice, it had already been appreciated that the Wnt signaling pathway controlled functions of the putative ISC. As the Wnt pathway relies on secretion of ligands to activate receptors on ISCs, the cellular source of required Wnt ligands within the niche is of great interest. With the identification of the ISCs came the identification of the ISC niche and the proximal cells most likely to provide Wnts, namely the epithelial Paneth cells, and the mesenchymal sub-epithelial myofibroblasts.

A focus on Paneth cells drove particular debate in the field for several years. Although a few groups characterized no ISC defects using various Paneth cell ablation methods (Clarke, 2006; Gregorieff and Clevers, 2005; Harada et al., 1999), others argued that these knockouts were incomplete and that observable transient defects had been rescued by escaper crypts (Sato et al., 2011). Subsequently, it was shown that Paneth cells express many ligands that may be critical for ISC functions and that associations of ISCs with Paneth cells increase the percentage of organoids (Sato et al., 2011). Despite this report, two studies used genetic knockout of *Atoh1* to completely remove all Paneth cells and carefully showed that it has no effect on the ISCs, suggesting that they are not a required source of Wnt *in vivo* (Durand et al., 2012; Kim et al., 2012). At the culmination of these studies, a diplomatic model was proposed in which Paneth cells provide Wnts redundantly with non-epithelial sources (Farin et al., 2012), the most likely being the sub-epithelial myofibroblast population, which cups the base of the crypts (Shaker and Rubin, 2010). However, this had not been tested rigorously *in vivo*. In fact, studying the Wnt

pathway is complicated because there are many ligands, which compensate for each other in single knockout studies. Fortunately, a new mouse line enabled the elimination of all Wnt secretion simultaneously, making it possible to address the source of Wnts in the ISC niche.

The gene *Porcn* encodes an o-acyl transferase that is required for all Wnt ligand secretion in *in vitro* assays using tagged Wnt ligands (Chen et al., 2009; Najdi et al., 2012; Proffitt and Virshup, 2012). The generation of *Porcn* conditional knockout mice allows the ablation of Wnt secretion in specific cell types, based on where Cre recombinase is expressed. We used mouse lines that express tamoxifen-inducible Cre recombinase specifically in the muscle cells (including the sub-epithelial myofibroblasts) and the epithelium (including the Paneth cells). After eliminating *Porcn* in the epithelium, muscle or in both compartments, we surprisingly found no effects on ISCs. There are several interpretations for this result.

The first pair of interpretations has to do with the possibility that Wnt was not completely eliminated in our model. It is possible some Wnts do not require Porcn for their secretion *in vivo*, and that these Wnts can maintain ISC function. We believe this is unlikely due to two rigorous *in vitro* studies showing Porcn is required for palmitoylation and secretion of all Wnts (Coombs et al., 2010; Liu et al., 2013). Another possibility is that residual Wnt sequestered in the extracellular matrix can sustain ISCs when sources are extinguished. As we examined some of our mice several weeks following ablation, residual Wnt ligands are unlikely to be sufficient to sustain the constantly regenerating epithelium.

The second interpretation addresses redundancy of Wnt sources in the intestinal epithelium. As we saw no phenotype in our studies, we accept the possibility that other cell types in the niche can redundantly provide Wnts. There are many cell types in the ISC niche including leukocytes, lymphocytes, endothelial cells, non-muscle fibroblasts, and neurons that

could contribute Wnt ligands (Shaker and Rubin, 2010). The final, and possibly most interesting, interpretation is that while downstream Wnt signaling is required in ISCs (including the actions of APC, β-Catenin and TCF4; (Korinek et al., 1998; Korinek et al., 1997; Morin et al., 1997; van Es et al., 2012)), Wnt ligands are not required to activate the pathway *in vivo*. Overexpression of the antagonist Dkk1, which interacts with the Wnt co-receptor Lrp6 (Bafico et al., 2001; Mao et al., 2001; Semenov et al., 2001) cripples intestinal crypts (Kuhnert et al., 2004; Pinto et al., 2003). While this study may imply that Wnt ligands are required for signaling (as Lrp6 inhibition by Dkk1 does not allow for Wnt binding or the formation of the Wnt-Frizzled-Lrp6 signaling complex), it only proves that downstream signaling from Lrp6 has important functions in ISCs. Interestingly, several other ligands including Norrin and R-spondins have been shown to activate the Wnt signaling pathway through various mechanisms (Niehrs, 2012; Xu et al., 2004), although the latter usually occurs in conjunction with Wnt ligand binding. In addition, it is possible that other signaling pathways may be able to cross-talk with the downstream Wnt effectors, thus bypassing the requirement for Wnt ligands, such as the FGF and Notch pathways (Fre et al., 2009; Katoh and Katoh, 2006).

Following our report (San Roman et al., 2014), another group found a similar phenotype using *Porcn* and *Villin-Cre* to eliminate Wnt secretion from the epithelium (Kabiri et al., 2014). Instead of using another Cre line to inactivate *Porcn* in the mesenchyme, however, they used a pharmacologic inhibitor of PORCN, C59, and saw effects only at very high doses. This result could reflect drug toxicity or off-target effects. If the drug is indeed specific at these high levels, this lends support to the hypothesis that there are redundant cellular sources of Wnts in the ISC niche, although this was not pursued further.

To make progress in this area, new tools and careful studies of purified ISC niche populations would be helpful. Antibodies that recognize Wnts or the development of transgenic animals with functional tagged Wnt proteins will be critical for future experiments. If many cell populations can redundantly provide Wnts, more work must be done to survey the populations of cells within the ISC niche, as well as to characterize them in greater detail. If these populations can be isolated, co-culture with *Wnt3a*$^{-/-}$ organoids, which do not survive without Wnt supplementation, would be a good way to show that they are capable of secreting Wnts to functionally maintain the ISCs. As Porcupine inhibitors are in clinical trials to treat Wnt-driven breast cancer (Lum and Clevers, 2012), it will be crucial to understand how this may or may not affect ISC function.

Role of CDX2 in intestinal stem cells

In addition to the important role of extrinsic niche factors on ISC function, intrinsic control is also critical. Recently, several TFs have been implicated in control of ISCs, including ASCL2, KLF5, ID1, YY1 and VDR (Bell and Shroyer, 2014; Nakaya et al., 2014; Peregrina et al., 2014; Perekatt et al., 2014; van der Flier et al., 2009; Zhang et al., 2014), suggesting that other factors controlling gene expression may also play a role in this cell population. Fortuitously, at the beginning of this thesis work, the Shivdasani Lab had been studying CDX2 for several years, and noted its expression throughout the intestinal epithelium, however its role was only well described in differentiated intestinal cells (Verzi et al., 2010). Following the identification of *Lgr5*$^{+}$ ISCs (Barker et al., 2007) and re-evaluation of conditional *Cdx2* knockout mice generated in the Shivdasani Lab, a very specific proliferation defect in ISCs was identified. To further study this, we took advantage of the *Lgr5*-knockin mice that enable ISC identification and isolation, ISC-specific gene knockouts, and lineage tracing (Barker et al., 2007). We found

that without *Cdx2* ISCs do not proliferate, contribute to mature villus cells *in vivo,* or generate organoids *in vitro*. These results were supported by two studies published during this thesis characterizing similar defects in ISC proliferation upon loss of *Cdx2*, however the mechanism behind this control was not addressed (Hryniuk et al., 2012; Stringer et al., 2012).

To better understand how CDX2 controls ISCs, we collected millions of ISCs from many mice, and performed CDX2 ChIP-seq and gene expression analysis on control and *Cdx2*-null ISCs. This allowed for identification of possible CDX2 target genes in ISCs and characterization of differential CDX2 binding during cell differentiation. To undercover the mechanism by which CDX2 controls ISCs, we narrowed our focus to one direct target gene, *Fgfbp1*, which promotes fibroblast growth factor signaling, a pathway identified within a network of down-regulated genes in *Cdx2*-null ISCs.

Together, these studies provide a foundation for further investigation of the role of the FGF pathway in ISCs. FGF ligands present in the ISC niche include FGF1, FGF4 and FGF7, which bind with FGFBP1 to enhance FGF signaling, have mitogenic effects in cell culture models, and increase in response to inflammation and radiation injury (Bajaj-Elliott et al., 1997; Brauchle et al., 1996; Finch et al., 1996; Sasaki et al., 2004). To follow up on this pathway, we will profile its activation status using phospho-receptor tyrosine kinase arrays, which will give us sensitive information about the activation status of FGF receptors in control and *Cdx2*-null ISCs. We will also pursue FGF pathway perturbation experiments in organoids using recombinant FGFs, recombinant FGFBP1 and a specific inhibitor of FGF receptors (BGJ398) to better understand the contribution of this pathway to ISC self-renewal and differentiation.

More broadly, this study contributes to our understanding of mechanisms for expression regulation during cell differentiation. Although we profile differential CDX2 binding in ISCs

and villus cells, how this occurs is still largely unknown, although a few mechanisms are likely. First, evidence supports the partnering of CDX2 with specific TFs to mediate gene expression in proliferating or differentiated cells (See **Chapters 4 and 5**, (Verzi et al., 2010; Verzi et al., 2013)). Second, CDX2 is phosphorylated at serine 60 by the mitogen-activated protein kinase pathway specifically in the crypts, but not villi, which may contribute to its differential activity in these compartments (Rings et al., 2001). Third, the chromatin state encountered by CDX2 in each cell state may influence its binding. Indeed, a survey of active cis-regulatory elements in ISCs, secretory or enterocyte progenitors and their mature counterparts revealed that most of the active chromatin marks remained consistent during cell differentiation, with smaller modules becoming differentially active in mature cells (Kim et al., 2014). Whether this change in chromatin is a result of or allows for TF binding is not well understood. Now that we are readily able to isolate sufficient quantities of ISCs to perform ChIP-seq analysis, we are well poised to understand more about chromatin dynamics during the cell differentiation process.

Another interesting avenue for further study is the contribution of CDX2 to intestinal cancer. Our study implies that CDX2 may function as an oncogene, since it promotes proliferation of ISCs, however previously research on this topic is confusing and contradictory. Some studies directly testing the link between CDX2 and intestinal cancer in mouse models have paradoxically concluded that loss of *Cdx2* reduces tumor formation in the small intestine and increases it in the large intestine (Aoki et al., 2011; Aoki et al., 2003). Whereas *Apc* mutations in humans induce colorectal cancer, *Apc* loss in mice leads mainly to small intestine adenomas and fewer colonic tumors (Su et al., 1992). This appears to be a species idiosyncrasy, and Apc^{Min} and other mouse models have made a great contribution to our understanding of colorectal cancer pathogenesis. However, interpretation of results using this mouse model is complicated by the

reliance on loss of heterozygosity of both *Apc* (and in some experiments also *Cdx2*), which occurs at different rates in the small and large intestine. Even within research on colon there are controversial findings; some studies show CDX2 loss in human colon cancer samples, which is supported by some mouse studies, suggesting a tumor suppressor role (Hinoi et al., 2003; Hryniuk et al., 2014); others show overexpression, suggesting oncogenic potential (Bhat et al., 2012; Witek et al., 2005). *In vitro* studies postulate that CDX2 controls the growth inhibitor p21 and that CDX2 overexpression suppresses proliferation in cancer cell lines (Bai et al., 2003; Mallo et al., 1998). Others contend that *CDX2* mutations are rare in colorectal cancer and that CDX2 loss reduces proliferation, potentially because it regulates pro-proliferative signaling pathways (Dang et al., 2006; Uesaka et al., 2002). My data shows that *Cdx2* loss specifically and significantly impairs proliferation of $Lgr5^{+}$ ISCs, supporting the hypothesis that CDX2 is an oncogene. In the future, this should be tested rigorously by eliminating both alleles of *Cdx2* and *Apc* simultaneously to separate the formation of adenomas from other CDX2 effects on DNA replication and cell cycle control, which may contribute to the frequency of loss of heterozygosity and confound interpretations.

Transcription factor interactions in cell differentiation

The continual and rapid process of cell differentiation within the adult intestinal epithelium relies on the precise execution of gene expression programs, however the factors contributing to this process have not been well understood. Since much gene regulation occurs through the action of TF binding at cell-type specific enhancers, the Shivdasani Lab profiled the transcription factor motifs within differentially active enhancer regions in proliferating and differentiated cells (Verzi et al., 2010). This proved to be a useful approach, as many TFs

expressed in the intestinal epithelium were identified, including GATA factors in proliferating cells, HNF1A and HNF4A in differentiated cells, and CDX2 in both. Subsequent studies *in vivo* demonstrated a co-occupancy of CDX2 and HNF4A at a few thousand intestinal genes, and provided some interesting insights into TF relationships (Verzi et al., 2013). Through investigation of *Cdx2* conditional knock-out mice, CDX2 was shown to be required for HNF4A binding and maintenance of active chromatin marks at sites where they are co-bound. These two pieces of information led to the conclusion that CDX2 is higher in the TF hierarchy than HNF4A, however whether these interactions lead to functional consequences *in vivo* remained unknown.

To investigate the relationship between CDX2 and HNF4A *in vivo*, we crossed *Cdx2* and *Hnf4a* conditional knock-out mice and examined their intestinal phenotypes. Because CDX2 controls HNF4A and chromatin at co-bound sites, and because mice lacking *Cdx2* alone fare worse than mice lacking *Hnf4a*, we expected that the combined knockout would resemble the *Cdx2* single knockout. However, we found that the $Hnf4a^{del}Cdx2^{del}$ mice fared significantly worse, requiring humane euthanasia within 7 days of tamoxifen injection. Analyses of intestinal tissues from these mice reveal short villi that lack mature enterocytes and cannot take up lipids. Perturbation of gene expression in these mice is significantly worse than in either single knockout, although this would not have been predicted by the analysis of chromatin alone. In parallel experiments, we found that *Gata4* partners with *Cdx2* to control crypt cell proliferation, a property that was not significantly altered in the $Hnf4a^{del}Cdx2^{del}$ knockout mice. It is likely that *Gata4* also regulates genes involved in cell maturation, however the defects in cell proliferation confounded this analysis. Together, these studies support a role for CDX2 partnering with HNF4A and GATA4 to control specific gene expression programs.

From the initial studies profiling active enhancer regions in differentiated cells, the most enriched motif was HNF1A. Previous attempts to study this factor were unsuccessful due to poor antibody quality, however new reagents enabled us to investigate this factor. Interestingly, HNF1A binds at many more sites throughout the genome than HNF4A or CDX2. As predicted, we find a high percentage of the HNF1A binding sites in putative enhancers marked by the active marks H3K4me2 and H3K27ac. These results suggest that HNF1A regulates many genes, however the phenotype of *Hnf1a*$^{-/-}$ mice is mild and we only identified the perturbation of a few hundred genes. To better understand whether HNF1A may depend on CDX2 for its intestinal functions, we overlapped their binding sites and found co-occupancy at more than 5,000 sites, which included almost all of the sites co-bound CDX2 and HNF4A. Detailing the complexities of these interactions, we find that loss of *Cdx2* reduces HNF1A levels, which may account for an observed global reduction in HNF1A binding in these knockout mice. Finally, to determine whether HNF1A functionally interacts with CDX2, we removed them simultaneously and observed the resulting intestinal phenotype. We found that loss of these two factors disrupts the structural integrity of the villus epithelium, causing abnormal invaginations to form within villi. We also observed defects in the lamina propria, likely reflecting disruption of epithelial-mesenchymal signaling.

As the binding and gene expression data was only available recently, this analysis represents our initial investigations into the role of HNF1A in the intestinal enhanceosome. There are still several hurdles to cross with respect to understanding how HNF1A relates to CDX2 and HNF4A in differentiated intestinal cells. First, it was difficult to reliably identify direct target genes of HNF1A because most of the binding sites do not correlate with perturbation of nearby gene expression upon its loss. In the future, we would like to consider

other methods of annotating these binding sites by combining them with the presence of chromatin marks, chromatin remodelers and other TFs, as well as metrics such as distance to the TSS and total number of binding sites near a given gene. This approach may lead us to understand the conditions in which HNF1A does or does not modulate expression of nearby genes. Second, we have demonstrated that CDX2 controls expression of both *Hnf1a* and *Hnf4a*, which means that studying $Cdx2^{del}$ mice may actually represent a combinatorial knockout to some degree and confound results when comparing compound knockout mice to $Cdx2^{del}$ mice alone. Third, although a simplistic picture of co-occupancy involves TFs binding their adjacent sequence motifs in enhancer regions, there are likely complex interactions between TFs on chromatin. For example, there is evidence that HNF1A and HNF4A can bind to each other and modulate gene expression even when only one factor binds DNA (Eeckhoute et al., 2004; Rowley et al., 2006). Bioinformatics and computational approaches will be valuable for understanding the intricacies of these interactions and to delineate true transcriptional targets of each TF.

Conclusions

In sum, this thesis provides many contributions to the understanding of mechanisms that maintain intestinal homeostasis throughout adult life. **Chapter 2** challenged the common belief that either Paneth cells or sub-epithelial myofibroblast are a source of required Wnt ligands for ISCs, suggesting unexpected functional redundancy or that other ligands or pathways are sufficient. **Chapter 3** identified a requirement for CDX2 in controlling ISC functions, defines differential transcriptional targets and CDX2 occupancy in ISCs compared to differentiated villus cells, and suggests that the fibroblast growth factor pathway may be important in ISCs.

Chapter 4 confirmed that CDX2 functionally interacts with GATA4 to control proliferating cells and HNF4A to regulate differentiation and maturation of mature enterocytes *in vivo*. Finally, **Chapter 5** characterized the role of HNF1A in the intestinal enhanceosome and provided evidence that interactions between CDX2 and HNF1A control villus structure and signaling to mesenchymal cells. These studies offer numerous avenues for further investigation and may aid in the understanding of cell differentiation in other tissues as well how these processes are inappropriate regulated in various intestinal pathologies including cancer.

References

Aoki, K., Kakizaki, F., Sakashita, H., Manabe, T., Aoki, M., and Taketo, M.M. (2011). Suppression of colonic polyposis by homeoprotein CDX2 through its nontranscriptional function that stabilizes p27Kip1. Cancer Res *71*, 593-602.

Aoki, K., Tamai, Y., Horiike, S., Oshima, M., and Taketo, M.M. (2003). Colonic polyposis caused by mTOR-mediated chromosomal instability in Apc+/Delta716 Cdx2+/- compound mutant mice. Nat Genet *35*, 323-330.

Bafico, A., Liu, G., Yaniv, A., Gazit, A., and Aaronson, S.A. (2001). Novel mechanism of Wnt signalling inhibition mediated by Dickkopf-1 interaction with LRP6/Arrow. Nat Cell Biol *3*, 683-686.

Bai, Y.-Q., Miyake, S., Iwai, T., and Yuasa, Y. (2003). CDX2, a homeobox transcription factor, upregulates transcription of the p21/WAF1/CIP1 gene. Oncogene *22*, 7942-7949.

Bajaj-Elliott, M., Breese, E., Poulsom, R., Fairclough, P.D., and MacDonald, T.T. (1997). Keratinocyte growth factor in inflammatory bowel disease. Increased mRNA transcripts in ulcerative colitis compared with Crohn's disease in biopsies and isolated mucosal myofibroblasts. Am J Pathol *151*, 1469-1476.

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

Bell, K.N., and Shroyer, N.F. (2014). Krupple-Like Factor 5 Is Required for Proper Maintenance of Adult Intestinal Crypt Cellular Proliferation. Dig Dis Sci.

Bhat, A.A., Sharma, A., Pope, J., Krishnan, M., Washington, M.K., Singh, A.B., and Dhawan, P. (2012). Caudal Homeobox Protein Cdx-2 Cooperates with Wnt Pathway to Regulate Claudin-1 Expression in Colon Cancer Cells. PLoS One *7*, e37174.

Brauchle, M., Madlener, M., Wagner, A.D., Angermeyer, K., Lauer, U., Hofschneider, P.H., Gregor, M., and Werner, S. (1996). Keratinocyte growth factor is highly overexpressed in inflammatory bowel disease. Am J Pathol *149*, 521-529.

Chen, B., Dodge, M.E., Tang, W., Lu, J., Ma, Z., Fan, C.-W., Wei, S., Hao, W., Kilgore, J., Williams, N.S.*, et al.* (2009). Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nat Chem Biol *5*, 100-107.

Clarke, A.R. (2006). Wnt signalling in the mouse intestine. Oncogene *25*, 7512-7521.

Coombs, G.S., Yu, J., Canning, C.A., Veltri, C.A., Covey, T.M., Cheong, J.K., Utomo, V., Banerjee, N., Zhang, Z.H., Jadulco, R.C.*, et al.* (2010). WLS-dependent secretion of WNT3A requires Ser209 acylation and vacuolar acidification. J Cell Sci *123*, 3357-3367.

Dang, L.H., Chen, F., Ying, C., Chun, S.Y., Knock, S.A., Appelman, H.D., and Dang, D.T. (2006). CDX2 has tumorigenic potential in the human colon cancer cell lines LOVO and SW48. Oncogene *25*, 2264-2272.

Durand, A., Donahue, B., Peignon, G., Letourneur, F., Cagnard, N., Slomianny, C., Perret, C., Shroyer, N.F., and Romagnolo, B. (2012). Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). Proc Natl Acad Sci U S A *109*, 8965-8970.

Eeckhoute, J., Formstecher, P., and Laine, B. (2004). Hepatocyte nuclear factor 4alpha enhances the hepatocyte nuclear factor 1alpha-mediated activation of transcription. Nucleic Acids Res *32*, 2586-2593.

Farin, H.F., Van Es, J.H., and Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. Gastroenterology *143*, 1518-1529 e1517.

Finch, P.W., Pricolo, V., Wu, A., and Finkelstein, S.D. (1996). Increased expression of keratinocyte growth factor messenger RNA associated with inflammatory bowel disease. Gastroenterology *110*, 441-451.

Fre, S., Pallavi, S.K., Huyghe, M., Lae, M., Janssen, K.P., Robine, S., Artavanis-Tsakonas, S., and Louvard, D. (2009). Notch and Wnt signals cooperatively control cell proliferation and tumorigenesis in the intestine. Proc Natl Acad Sci U S A *106*, 6309-6314.

Gregorieff, A., and Clevers, H. (2005). Wnt signaling in the intestinal epithelium: from endoderm to cancer. Genes & development *19*, 877-890.

Harada, N., Tamai, Y., Ishikawa, T., Sauer, B., Takaku, K., Oshima, M., and Taketo, M.M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. The EMBO journal *18*, 5931-5942.

Hinoi, T., Loda, M., and Fearon, E.R. (2003). Silencing of CDX2 expression in colon cancer via a dominant repression pathway. The Journal of biological chemistry *278*, 44608-44616.

Hryniuk, A., Grainger, S., Savory, J.G., and Lohnes, D. (2014). Cdx1 and cdx2 function as tumor suppressors. J Biol Chem *289*, 33343-33354.

Hryniuk, A., Grainger, S., Savory, J.G.A., and Lohnes, D. (2012). Cdx function is required for maintenance of intestinal identity in the adult. Dev Biol *363*, 426-437.

Kabiri, Z., Greicius, G., Madan, B., Biechele, S., Zhong, Z., Zaribafzadeh, H., Edison, Aliyev, J., Wu, Y., Bunte, R.*, et al.* (2014). Stroma provides an intestinal stem cell niche in the absence of epithelial Wnts. Development *141*, 2206-2215.

Katoh, M., and Katoh, M. (2006). Cross-talk of WNT and FGF signaling pathways at GSK3beta to regulate beta-catenin and SNAIL signaling cascades. Cancer Biol Ther *5*, 1059-1064.

Kim, T.-H., Escudero, S., and Shivdasani, R.A. (2012). Intact function of Lgr5 receptor-expressing intestinal stem cells in the absence of Paneth cells. Proc Natl Acad Sci U S A *109*, 3932-3937.

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

Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P.J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet *19*, 379-383.

Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science *275*, 1784-1787.

Liu, J., Pan, S., Hsieh, M.H., Ng, N., Sun, F., Wang, T., Kasibhatla, S., Schuller, A.G., Li, A.G., Cheng, D.*, et al.* (2013). Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. Proc Natl Acad Sci U S A *110*, 20224-20229.

Lum, L., and Clevers, H. (2012). The Unusual Case of Porcupine. Science (New York, NY) *337*, 922-923.

Mallo, G.V., Soubeyran, P., Lissitzky, J.C., André, F., Farnarier, C., Marvaldi, J., Dagorn, J.C., and Iovanna, J.L. (1998). Expression of the Cdx1 and Cdx2 homeotic genes leads to reduced malignancy in colon cancer-derived cells. The Journal of biological chemistry *273*, 14030-14036.

Mao, B., Wu, W., Li, Y., Hoppe, D., Stannek, P., Glinka, A., and Niehrs, C. (2001). LDL-receptor-related protein 6 is a receptor for Dickkopf proteins. Nature *411*, 321-325.

Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science *275*, 1787-1790.

Najdi, R., Proffitt, K., Sprowl, S., Kaur, S., Yu, J., Covey, T.M., Virshup, D.M., and Waterman, M.L. (2012). A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. Differentiation *84*, 203-213.

Nakaya, T., Ogawa, S., Manabe, I., Tanaka, M., Sanada, M., Sato, T., Taketo, M.M., Nakao, K., Clevers, H., Fukayama, M.*, et al.* (2014). KLF5 regulates the integrity and oncogenicity of intestinal stem cells. Cancer Res *74*, 2882-2891.

Niehrs, C. (2012). The complex world of WNT receptor signalling. Nat Rev Mol Cell Biol *13*, 767-779.

Peregrina, K., Houston, M., Daroqui, C., Dhima, E., Sellers, R.S., and Augenlicht, L.H. (2014). Vitamin D is a determinant of mouse intestinal Lgr5 stem cell functions. Carcinogenesis.

Perekatt, A.O., Valdez, M.J., Davila, M., Hoffman, A., Bonder, E.M., Gao, N., and Verzi, M.P. (2014). YY1 is indispensable for Lgr5+ intestinal stem cell renewal. Proc Natl Acad Sci U S A *111*, 7695-7700.

Proffitt, K.D., and Virshup, D.M. (2012). Precise Regulation of Porcupine Activity Is Required for Physiological Wnt Signaling. J Biol Chem *287*, 34167-34178.

Rings, E.H., Boudreau, F., Taylor, J.K., Moffett, J., Suh, E.R., and Traber, P.G. (2001). Phosphorylation of the serine 60 residue within the Cdx2 activation domain mediates its transactivation capacity. Gastroenterology *121*, 1437-1450.

Rowley, C.W., Staloch, L.J., Divine, J.K., McCaul, S.P., and Simon, T.C. (2006). Mechanisms of mutual functional interactions between HNF-4alpha and HNF-1alpha revealed by mutations that cause maturity onset diabetes of the young. Am J Physiol Gastrointest Liver Physiol *290*, G466-475.

San Roman, A.K., Jayewickreme, C.D., Murtaugh, L.C., and Shivdasani, R.A. (2014). Wnt secretion from epithelial cells and subepithelial myofibroblasts is not required in the mouse intestinal stem cell niche in vivo. Stem cell reports *2*, 127-134.

Sasaki, H., Hirai, K., Yamamoto, H., Tanooka, H., Sakamoto, H., Iwamoto, T., Takahashi, T., Terada, M., and Ochiya, T. (2004). HST-1/FGF-4 plays a critical role in crypt cell survival and facilitates epithelial cell restitution and proliferation. Oncogene *23*, 3681-3688.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature *469*, 415-418.

Semenov, M.V., Tamai, K., Brott, B.K., Kuhl, M., Sokol, S., and He, X. (2001). Head inducer Dickkopf-1 is a ligand for Wnt coreceptor LRP6. Curr Biol *11*, 951-961.

Shaker, A., and Rubin, D.C. (2010). Intestinal stem cells and epithelial-mesenchymal interactions in the crypt and stem cell niche. Translational research : the journal of laboratory and clinical medicine *156*, 180-187.

Stringer, E.J., Duluc, I., Saandi, T., Davidson, I., Bialecka, M., Sato, T., Barker, N., Clevers, H., Pritchard, C.A., Winton, D.J.*, et al.* (2012). Cdx2 determines the fate of postnatal intestinal endoderm. Development *139*, 465-474.

Su, L.K., Kinzler, K.W., Vogelstein, B., Preisinger, A.C., Moser, A.R., Luongo, C., Gould, K.A., and Dove, W.F. (1992). Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science (New York, NY) *256*, 668-670.

Uesaka, T., Lu, H., Katoh, O., and Watanabe, H. (2002). Heparin-binding EGF-like growth factor gene transcription regulated by Cdx2 in the intestinal epithelium. American Journal Of Physiology-Gastrointestinal And Liver Physiology *283*, G840-847.

van der Flier, L.G.v.d., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Beghtel, H., van den Born, M., Guryev, V., Oving, I.*, et al.* (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell *136*, 903-912.

van Es, J.H., Haegebarth, A., Kujala, P., Itzkovitz, S., Koo, B.K., Boj, S.F., Korving, J., van den Born, M., van Oudenaarden, A., Robine, S.*, et al.* (2012). A critical role for the Wnt effector Tcf4 in adult intestinal homeostatic self-renewal. Mol Cell Biol *32*, 1918-1927.

Verzi, M.P., Shin, H., He, H.H., Sulahian, R., Meyer, C.A., Montgomery, R.K., Fleet, J.C., Brown, M., Liu, X.S., and Shivdasani, R.A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev Cell *19*, 713-726.

Verzi, M.P., Shin, H., San Roman, A., Liu, X.S., and Shivdasani, R.A. (2013). Intestinal Master Transcription Factor CDX2 Controls Chromatin Access for Partner Transcription Factor Binding. Mol Cell Biol *33*, 281-292.

Witek, M.E., Nielsen, K., Walters, R., Hyslop, T., Palazzo, J., Schulz, S., and Waldman, S.A. (2005). The Putative Tumor Suppressor Cdx2 Is Overexpressed by Human Colorectal Adenocarcinomas. Clin Cancer Res *11*, 8549-8556.

Xu, Q., Wang, Y., Dabdoub, A., Smallwood, P.M., Williams, J., Woods, C., Kelley, M.W., Jiang, L., Tasman, W., Zhang, K.*, et al.* (2004). Vascular development in the retina and inner ear: control by Norrin and Frizzled-4, a high-affinity ligand-receptor pair. Cell *116*, 883-895.

Zhang, N., Yantiss, R.K., Nam, H.S., Chin, Y., Zhou, X.K., Scherl, E.J., Bosworth, B.P., Subbaramaiah, K., Dannenberg, A.J., and Benezra, R. (2014). ID1 Is a Functional Marker for Intestinal Stem and Progenitor Cells Required for Normal Response to Injury. Stem cell reports *3*, 716-724.

Appendix

In addition to the original research presented in chapters 2-5, I present here the abstracts from one published review article and one book chapter in press that I wrote and two published research articles that I collaborated on.

Appendix 1: Boundaries, junctions and transitions in the gastrointestinal tract

Adrianna K. San Roman and Ramesh A. Shivdasani

This work was published in *Exp. Cell Res. (2011)* 317:2711-2718.

Abstract

Contiguous regions along the mammalian gastrointestinal tract, from the esophagus to the rectum, serve distinct digestive functions. Some organs, such as the esophagus and glandular stomach or the small bowel and colon, are separated by sharp boundaries. The duodenal, jejunal and ileal segments of the small intestine, by contrast, have imprecise borders. Because human esophageal and gastric cancers frequently arise in a background of tissue metaplasia and some intestinal disorders are confined to discrete regions, it is useful to appreciate the molecular and cellular basis of boundary formation and preservation. Here we review the anatomy and determinants of boundaries and transitions in the alimentary canal with respect to tissue morphology, gene expression, and, especially, transcriptional control of epithelial identity. We discuss the evidence for established and candidate molecular mechanisms of boundary formation, including the solitary and combinatorial actions of tissue-restricted transcription factors. Although the understanding remains sparse, genetic studies in mice do provide insights into dominant mechanisms and point the way for future investigation.

Contributions: A. K. San Roman wrote the article with critical feedback from R. A. Shivdasani.

Appendix 2: The alimentary canal

Adrianna K. San Roman, Tae-Hee Kim, and Ramesh A. Shivdasani

This chapter is in press as part of The Supplement to Kaufman's Atlas of Mouse Development, edited by R. Baldock, J. Bard, D. Davidson, and G. Morriss-Kay, *Elsevier*.

Abstract

In progressing from a simple tube to mature digestive organs, the alimentary tract is patterned along distinct axes. Transcription factors pattern the gut tube along the rostro-caudal axis, delineating a foregut, midgut and hindgut. Reciprocal signals along the radial axis then enable differentiation of mesoderm-derived mesenchyme and muscle cells from diverse endoderm-derived epithelia. Inductive interactions further specify discrete organs, such as the stomach, liver and pancreas, which show dorso-ventral and left-right asymmetries. Finally, patterning of the radially symmetric small bowel along a luminal-mural axis separates progenitors in sub-mucosal crypts from differentiated cells restricted to the villi. Diverse forces drive morphogenesis: radial cell intercalation, cell shape changes, asymmetric tissue growth, and tension from surrounding structures. Much remains unclear about how positional and inductive cues activate particular genes to implement this developmental program.

Contributions: A. K. San Roman and T.-H. Kim wrote this chapter with feedback from R. A. Shivdasani.

Appendix 3: Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding

Michael P Verzi, Hyunjin Shin, Adrianna San Roman, X Shirley Liu, and Ramesh A Shivdasani

This article was published in *Molecular and Cellular Biology (2013) 33:281-292.*

Abstract

Tissue-specific gene expression requires modulation of nucleosomes, allowing transcription factors to occupy cis elements that are accessible only in selected tissues. Master transcription factors control cell-specific genes and define cellular identities, but it is unclear if they possess special abilities to regulate cell-specific chromatin and if such abilities might underlie lineage determination and maintenance. One prevailing view is that several transcription factors enable chromatin access in combination. The homeodomain protein CDX2 specifies the embryonic intestinal epithelium, through unknown mechanisms, and partners with transcription factors such as HNF4A in the adult intestine. We examined enhancer chromatin and gene expression following Cdx2 or Hnf4a excision in mouse intestines. HNF4A loss did not affect CDX2 binding or chromatin, whereas CDX2 depletion modified chromatin significantly at CDX2-bound enhancers, disrupted HNF4A occupancy, and abrogated expression of neighboring genes. Thus, CDX2 maintains transcription-permissive chromatin, illustrating a powerful and dominant effect on enhancer configuration in an adult tissue. Similar, hierarchical control of cell-specific chromatin states is probably a general property of master transcription factors.

Contributions: A. K. San Roman performed matings of *Cdx2* and *Hnf4a* mice and subsequent histologic analyses.

Appendix 4: Dissecting engineered cell types and enhancing cell fate conversion via CellNet

Samantha A Morris, Patrick Cahan, Hu Li, Anna M Zhao, Adrianna San Roman, Ramesh A Shivdasani, James J Collins, and George Q Daley

This article was published in *Cell (2014) 158:889-902.*

Abstract

Engineering clinically relevant cells in vitro holds promise for regenerative medicine, but most protocols fail to faithfully recapitulate target cell properties. To address this, we developed CellNet, a network biology platform that determines whether engineered cells are equivalent to their target tissues, diagnoses aberrant gene regulatory networks, and prioritizes candidate transcriptional regulators to enhance engineered conversions. Using CellNet, we improved B cell to macrophage conversion, transcriptionally and functionally, by knocking down predicted B cell regulators. Analyzing conversion of fibroblasts to induced hepatocytes (iHeps), CellNet revealed an unexpected intestinal program regulated by the master regulator Cdx2. We observed long-term functional engraftment of mouse colon by iHeps, thereby establishing their broader potential as endoderm progenitors and demonstrating direct conversion of fibroblasts into intestinal epithelium. Our studies illustrate how CellNet can be employed to improve direct conversion and to uncover unappreciated properties of engineered cells.

Contributions: A. K. San Roman performed timed matings and genotyping of *Cdx2*$^{Fl/Fl}$ embryos and histological preparations for subsequent analysis by S. A. Morris.

Supplemental Figures and Tables

Chapter 2 Supplemental Information

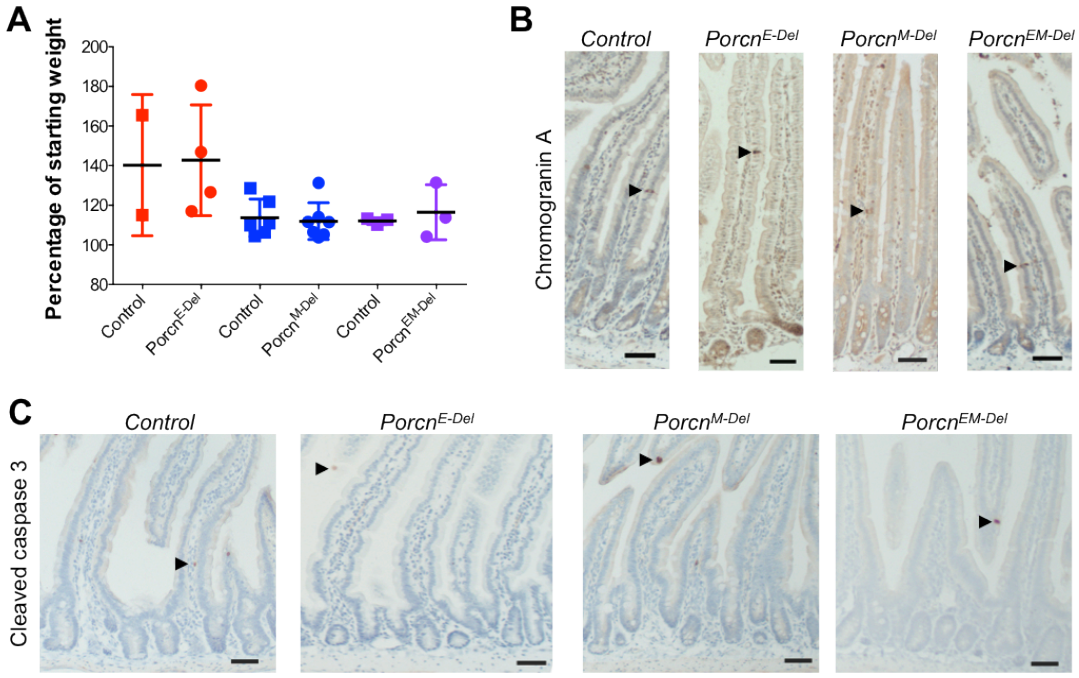


Figure S2.1. Effects of *Porcn* loss on body weight, enteroendocrine cells, and apoptosis. (A) Relative weights of Control, *Porcn*$^{E\text{-}Del}$, *Porcn*$^{M\text{-}Del}$, and *Porcn*$^{EM\text{-}Del}$ mice on the day of euthanasia, compared to the first day of TAM injection. **(B-C)** Immunohistochemistry for Chromogranin A (B) and Cleaved caspase 3 (C) reveals no enteroendocrine cell deficit or apoptosis in mice lacking *Porcn* in epithelium only, muscle only, or both compartments compared to controls (*Porcn*$^{+/Y}$*; Villin-Cre*^{ERT2}*; Myh11-Cre*^{ERT2}). Scale bars, 50 μm.

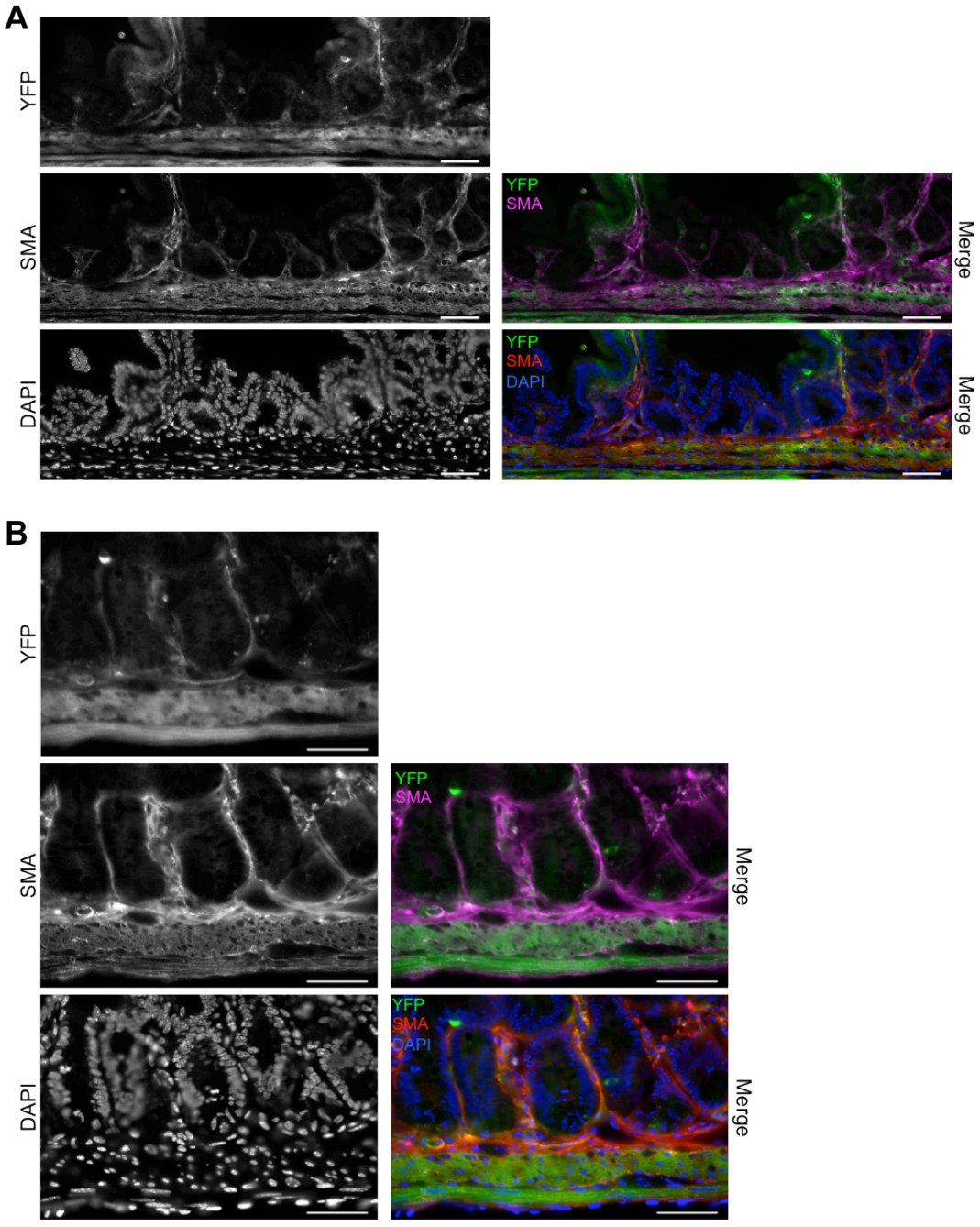


Figure S2.2. *Myh11-Cre*^{ERT2} is expressed in the same cells as smooth muscle actin. Low- (A) and high- (B) magnification images of *Myh11-Cre*$^{ER-T2}$*;Rosa26R*^{YFP} (green) stained with α-smooth muscle actin antibody (SMA; magenta/red). Nuclei were counterstained with DAPI (blue). Scale bars, 50 μm.

Table S2.1. Wnt target gene expression statistics.

Gene	*Control* Mean (N=2)	*Porcn*$^{E\text{-}Del}$ Mean (N=4)		
Myc	1	0.90		
Cyclin D1	1	0.93		
Axin2	1	0.96		
Ascl2	1	0.77		
Cd44	1	0.98		
Sox9	1	1.00		
Lgr5	1	0.56		
	***Control* Mean (N=4)**	***Porcn*$^{M\text{-}Del}$ Mean (N=4)**	**P-Value**	**Holm-Sidak Significant?**
Myc	1	1.03	0.93	No
Cyclin D1	1	1.19	0.49	No
Axin2	1	0.48	0.16	No
Ascl2	1	0.83	0.49	No
Cd44	1	0.90	0.55	No
Sox9	1	0.55	0.31	No
Lgr5	1	0.92	0.78	No
	***Control* Mean (N=3)**	***Porcn*$^{EM\text{-}Del}$ Mean (N=3)**		
Myc	1	0.60	0.071	No
Cyclin D1	1	0.70	0.20	No
Axin2	1	0.33	0.090	No
Ascl2	1	0.59	0.055	No
Cd44	1	0.57	0.022	No
Sox9	1	0.68	0.25	No
Lgr5	1	0.76	0.36	No

Supplementary Experimental Procedures

Mouse intestine epithelium harvests

The middle 1/3 of the small intestine was dissected and flushed with phosphate-buffered saline (PBS). The first 1 cm was reserved for whole intestine (unfractionated) DNA or mRNA analysis. The remainder was incubated on a shaker for 30 min in 5 mM EDTA solution at 4°C, followed by another 15 minute incubation with fresh EDTA solution at 4°C, and further shaken by hand for 2 min to extract the epithelium. Villi and crypts were separated as the residue and flow-through, respectively, after passage over 70 μm filters. Cells were washed in PBS and pelleted at

3000 rpm for 5 min at 4°C. For DNA analyses, tissue was snap-frozen in liquid nitrogen. For RNA analyses, tissue was frozen in TRIzol reagent (Invitrogen). All tissues were stored at -80°C.

Porcn genotyping

Porcn gene recombination was assessed using the following primers in PCR reactions using Platinum Blue PCR Supermix (Invitrogen): TGAGTGCTCAAATCCCAACC (common forward), CCAGCATGTGAAAATGTCAAC ($Porcn^{WT}$ 685-bp product, $Porcn^{Fl}$ 762-bp product), and GTGTCCACCATGTGCATCTC (recombined $Porcn^{Del}$ 485-bp product).

Gene expression analysis by qRT-PCR

Quantitative reverse transcriptase PCR (qRT-PCR) was performed in 20 μl reaction volumes using FastStart Universal SYBR Green Master Mix (Roche) and the following gene specific primers:

Primer Name	**Sequence 5' → 3'**
Porcn Forward	CTGCCTACTGTCCAACAGGG
Porcn Reverse	GCATGCTTCAGGTAAGACGG
Ascl2 Forward	GAGCAGGAGCTGCTTGACTT
Ascl2 Reverse	TCCGGAAGATGGAAGATGTC
Lgr5 Forward	CCTACTCGAAGACTTACCCAGT
Lgr5 Reverse	GCATTGGGGTGAATGATAGCA
Axin2 Forward	TGACTCTCCTTCCAGATCCCA
Axin2 Reverse	TGCCCACACTAGGCTGACA
Cd44 Forward	CACCATTGCCTCAACTGTGC
Cd44 Reverse	TTGTGGGCTCCTGAGTCTGA
Cyclin D1 Forward	GCGTACCCTGACACCAATCTC
Cyclin D1 Reverse	CTCCTCTTCGCACTTCTGCTC
Myc Forward	ATGCCCCTCAACGTGAACTTC
Myc Reverse	CGCAACATAGGATGGAGAGCA
Sox9 Forward	GAGCCGGATCTGAAGAGGGA
Sox9 Reverse	GCTTGACGTGTGGCTTGTTC
HPRT Forward	AAGCTTGCTGGTGAAAAGGA
HPRT Reverse	TTGCGCTCATCTTAGGCTTT

Relative mRNA levels were determined using the Δ-Δ-Ct method (Livak and Schmittgen, 2001) and normalization to HPRT transcripts. GraphPad Prism 6.0c software for Mac OS X was used for graphing and statistical analysis. Data is represented on graphs as mean ± standard error of the mean (SEM) from biological replicates. For *Porcn* mRNA expression in knockout samples, *P-values* were calculated using a one-tailed *t*-test. For Wnt target gene expression, *P-values* were calculated using a two-tailed *t*-test for each gene and significance was assessed using the Holm-Sidak method to correct for multiple comparisons.

Immunohistochemistry

Paraffin embedded sections protocol and antibodies

For paraffin sections, tissue was dehydrated in an ethanol series, embedded in paraffin, and cut in 5 μm sections. Tissue sections were stained with hematoxylin and eosin and alcian blue using standard histology protocols. For immunohistochemistry, antigens were retrieved in 10 mM sodium citrate buffer, pH 6.0, followed by incubation in methanol with 0.5% H_2O_2 to inhibit endogenous peroxidases. Tissues were blocked in 5% fetal bovine serum (FBS; Invitrogen) and incubated overnight at 4°C with one of the following antibodies:

Name	**Species**	**Dilution**	**Company**
Ki67	Mouse	1:2000	Vector Labs
BrdU	Rat	1:300	AbD Serotec
Chromogranin A	Rabbit	1:500	Immunostar
β-catenin	Rabbit	1:250	BD Biosciences
Cleaved caspase 3	Rabbit	1:1000	Cell Signaling
Lysozyme	Rabbit	1:50	Invitrogen

After washing, slides were incubated in species-specific biotin-conjugated anti-IgG (1:300, Vector Laboratories). Antigens were detected using the Vectastain Elite ABC Kit (Vector

Laboratories) with diaminobenzidine (Sigma) substrate and tissue was counterstained with Harris' Hematoxylin (Electron Microscopy Science).

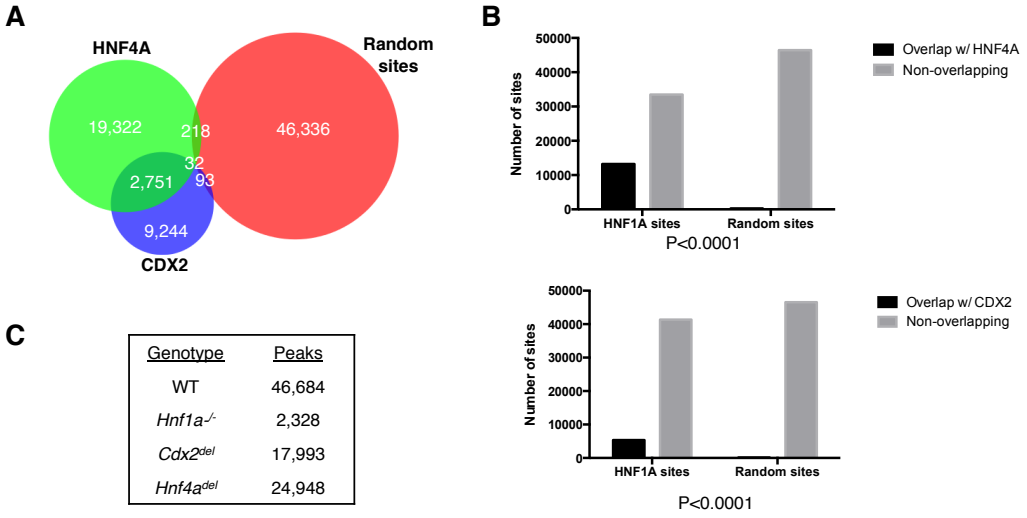
Frozen sections protocol and antibodies

For frozen sections, tissue was incubated in 30% sucrose overnight, embedded in Optimal Cutting Temperature compound (Tissue-Tek) on dry ice, and cut in 10 μm sections. Sections were blocked in 5% FBS then stained with mouse anti-αSMA (mouse, 1:300, Biogenex), followed by incubation with Alexa 546 conjugated anti-mouse IgG secondary antibody (1:500, Life Technologies). Vectashield mounting media with DAPI (Vector Laboratories) was added prior to imaging.

Supplementary References

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Chapter 5 Supplemental Information



Genotype	Peaks
WT	46,684
Hnf1a$^{-/-}$	2,328
Cdx2^{del}	17,993
Hnf4a^{del}	24,948

Figure S5.1. HNF1A significantly overlaps with CDX2 and HNF4A. (A) Venn diagram showing overlap of CDX2 and HNF4A with a random set of sites matched to the HNF1A set but randomly distributed throughout the genome. **(B)** Fisher's exact test analysis of overlaps between HNF1A and HNF4A (top) or HNF1A and CDX2 (bottom). Both overlaps are very significant compared to the random set (P<0.0001). **(C)** Number of HNF1A binding peaks called by MACs in the indicated genotypes.

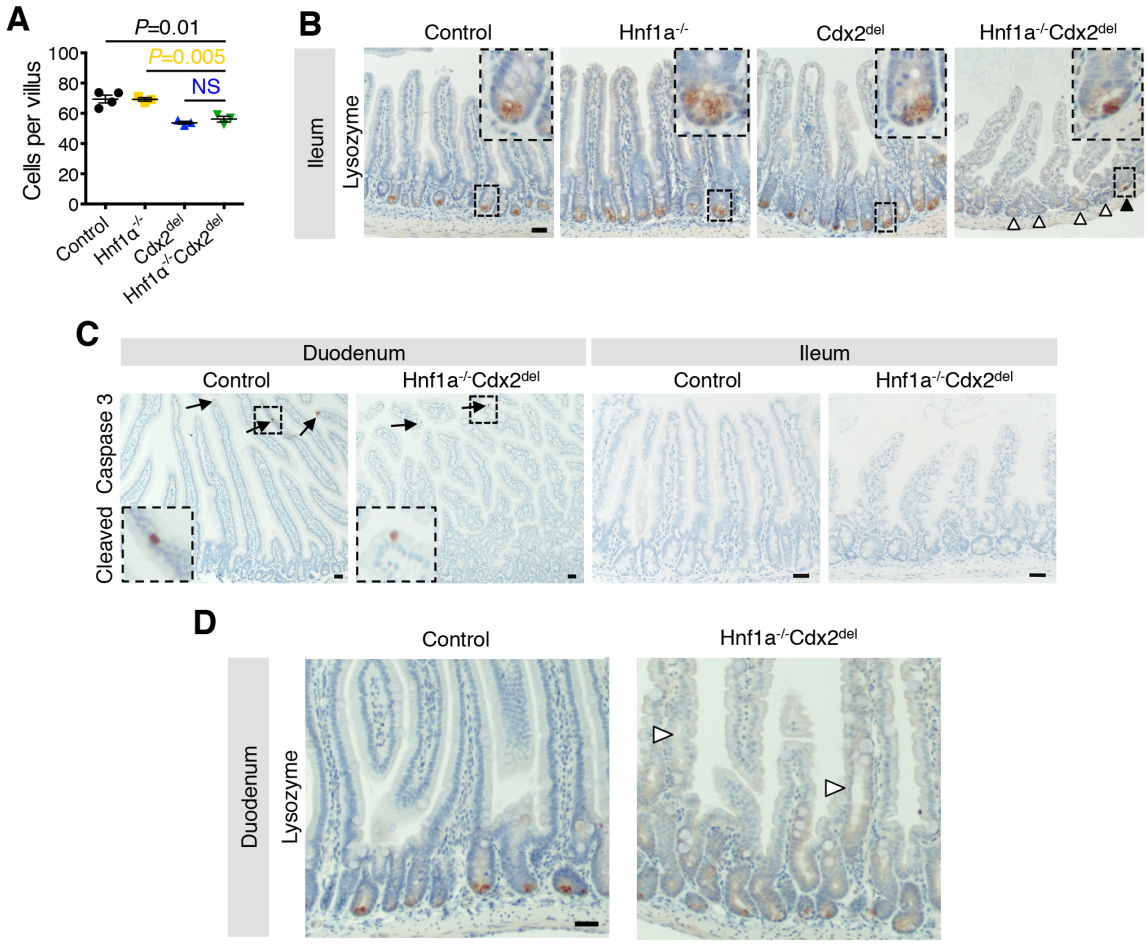


Figure S5.2. Consequences of intestinal deletion of *Hnf1a and Cdx2.*

(A) Quantitation of total number of cells per villus shows no further decrease in *Hnf1a*$^{-/-}$*Cdx2*^{del} mice compared to *Cdx2*^{del}. Bars represent mean +/- SEM. NS, not significant. **(B)** Lysozyme staining in ileal Paneth cells shows reduced levels in most *Hnf1a*$^{-/-}$*Cdx2*^{del} crypts (white arrows), with occasional strong staining (example in inset, black arrow). **(C)** Cleaved caspase 3 staining shows slightly reduced numbers of positive cells in *Hnf1a*$^{-/-}$*Cdx2*^{del} compared to control duodenum (left) and ileum (right). Examples of positive magnified in dotted region on bottom left. **(D)** Lysozyme staining indicates that *Hnf1a*$^{-/-}$*Cdx2*^{del} duodenal epithelial ingrowths lack Paneth cells (white arrowheads), which are present in the crypts but not in the abnormal villus structures. All scale bars, 30 μm.