CELL-SPECIFIC GENE EXPRESSION: PYLORUS MORPHOGENESIS AND HEDGEHOG-REGULATED ENHANCERS

by

Aaron Mark Udager

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cell and Developmental Biology)
in the University of Michigan
2010

Doctoral Committee:

Professor Deborah L. Gumucio, Chair
Professor Gregory R. Dressler
Professor Ronald J. Koenig
Professor Linda C. Samuelson
Associate Professor Deneen M. Wellik
Assistant Professor Scott E. Barolo

2010

To my wife Kara and my parents Mark and Anita

ACKNOWLEDGEMENTS

First, I'd like to thank my mentor Deb Gumucio. Over the past four years, you have given me freedom to explore and opportunities to succeed (as well as fail). Without a doubt, these experiences have made me a better scientist. I'll always admire your love of science and the dedication you show to teaching students, supporting your employees, and maintaining a family. I know we'll remain good friends in the future.

Second, I'd like to thank the members of my thesis committee. You challenged me when I needed to be and provided critical feedback at several key junctures in this project. Our meetings over the past several years have been an important part of my growth as a scientist.

Third, I need to thank the extensive contributions made to this project by our collaborators, both here at the University and abroad. I'd like to thank Drs. Doug Engel and Kim-Chew Lim in this department for providing *Gata3*$^{lacZ/+}$ mice and sharing a number of other resources. I'd also like to thank Dr. Richard Harvey of the Victor Chang Cardiac Research Institute in Australia for generously providing *Nkx2-5*$^{lacZ/+}$ mice. A special thanks to Dr. Scott Barolo and his post-doc Dr. Dave Parker, of this department, for their critical work on the Topo-TX project (see Chapter IV).

I'd like to thank the administrative staffs of the Department of Cell and Developmental Biology, the Center for Organogenesis, and the Medical Scientist Training Program (MSTP). In particular, I'd like to recognize Ellen Elkin, Penny Morris, Kristen Hug, and Becky Pintar. I'd also like to thank the technical staffs of the Microscopy and Image Analysis Laboratory and the Morphology Core. In particular, I'd like to recognize Shelley Almburg, Marta Dzaman, and Maria Ripberger.

I'd like to thank past and present members of the Gumucio lab: Åsa Kolterud for your help with X-gal staining; Xing Li for teaching me everything I know about microarrays; Chunbo Hu for your help with *in situ* hybridizations; Tracy Xiao for your help with the border project; and Neil Richards for your help with RT-PCR and all things cloning. I'd also like to recognize Katherine Gurdziel and Mridula Jayaraman for your past (and future!) contributions to the Topo-TX project. And thank you to Jierong Lang for your tremendous work in maintaining the mouse room; the lab would not be the same without you. I'd like to thank all of our lab managers over the past four years: Kathleen Portman, Wendy McKimpson, Xueming Tang, and in particular, Mike Czerwinski. I'd also like to acknowledge the undergraduate students that have worked with me: Mikey Thomas and Vivek Mendiratta. Thank you to my fellow graduate students (and/or post-docs), Anne Grosse, Kate Walton, Andrea Waite, and Will Zacharias, for personal as well as scientific discussions. A special thanks to Ajay Prakash for all of your help with the Gata3 project. It's been a blast. I couldn't imagine being happier anywhere else.

Lastly, I need to thank my family and friends. A special thank you to my MSTP classmates, Jae Lee, Steve Philips, and Dan Wahl, for our all too infrequent happy hour trips. To my sister Andrea, her husband Scott, and their two little girls, Annaleigh and Charlotte: I love you guys! To my in-laws, Gene and Lisa: thank you for your love and support and for being good role models for an academic career! To my grandparents, Gen, Hal, Wilbur, and Gladys: thank you for your love and support. I wish you were all around to share these important life moments with me! To my parents, Mark and Anita: thank you for recognizing what I needed to succeed and doing everything you could to provide it. To my mom: thank you for instilling in me the idea that nothing but my best is good enough. To my dad: thank you for reminding me the importance of taking time to enjoy life. Finally, to my wife Kara, who was born with Infantile Hypertrophic Pyloric Stenosis: Thank you. Thank you for listening to my rambling science discussions on our long runs. Thank you for helping me to be a better person. Thank you for making the last three and a half years the best of my life. I wouldn't be the same without you and neither would this thesis!

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER II

CHAPTER III

CHAPTER IV

LIST OF TABLES

CHAPTER II

CHAPTER III

CHAPTER IV

ABSTRACT

CELL-SPECIFIC GENE EXPRESSION: PYLORUS MORPHOGENESIS AND HEDGEHOG-REGULATED ENHANCERS

by

Aaron Mark Udager

Chair: Deborah L. Gumucio

The precise spatiotemporal control of gene expression is integral to the survival of all organisms. Inappropriate gene expression can lead to developmental defects in newborns, such as Infantile Hypertrophic Pyloric Stenosis, in which hypertrophy of pyloric sphincter smooth muscle leads to gastric outlet obstruction. This thesis work analyzes the mechanisms and consequences of cell-specific gene expression in three systems: establishment of the epithelial gastro-duodenal (pyloric) border, development of smooth muscle structures at the pylorus, and transcriptional response to Hedgehog (Hh) signaling in *Drosophila*.

Microarray is used to characterize the antral, pyloric, and duodenal transcriptomes at embryonic days (E) 14.5 and 16.5. At E16.5, hundreds of genes are upregulated specifically in duodenal epithelium. This event is termed *intestinalization* because the activated genes are associated with intestinal function. Several transcription factors (i.e., *Tcfec*, *Creb3l3*, and *Hnf4γ*) are upregulated in duodenal epithelium and levels of Hh signaling are downregulated in duodenal mesenchyme. In addition, novel pyloric genes are identified, including *Gata3*, which encodes a zinc finger transcription factor.

A role for *Gata3* during pylorus development is elucidated using a genetic model of *Gata3* insufficiency. Gata3 and the homeodomain transcription factor Nkx2-5 co-localize with molecular markers of pyloric smooth muscle and are expressed in novel bilateral smooth muscle structures at the pylorus (i.e., the ventral pyloric cords). Loss of *Gata3* alters the shape of the pylorus and attenuates the pyloric constriction. The ventral pyloric cords and outer longitudinal smooth muscle at the pylorus are absent in *Gata3* null embryos. Gata3 does not control *Nkx2-5* expression at the pylorus.

An *in silico* approach identifies Hh-regulated enhancers in *Drosophila*. Binding sites for the Hh transcriptional effector cubitus interruptus (Ci) are significantly clustered in the genomes of two divergent *Drosophila* species, but mutant Ci sites are not. Putative Hh-regulated enhancers are identified by the comparison of orthologous regions of significant Ci clustering. Two of these enhancers (*inv* and *rdx*) are active in Hh-responsive cells of the *Drosophila* larval imaginal wing disc.

These studies reveal novel gene expression patterns during pylorus morphogenesis and suggest an approach to identifying direct transcriptional targets of signaling pathways.

CHAPTER I

INTRODUCTION

Cell-specific gene regulation is essential for embryonic development and adult homeostasis: combinatorial patterns of gene expression establish and maintain cellular identity; unique spatiotemporal expression domains position genes to influence morphogenesis and function; and distinct transcriptional responses to cell-cell signaling integrate spatiotemporal position and cellular identity. This thesis explores the mechanisms and consequences of cell-specific gene regulation.

In the first half of my thesis work I examined the patterning of the pylorus, the site of muscular sphincter that marks the boundary between the stomach and intestine (see Chapters II and III). To provide context for this part of the work, I will review the cellular and molecular basis of gut tube development and then focus specifically on what was known regarding the patterning of the pylorus at the beginning of my studies.

The second half of my thesis work centers on one important cell-cell signaling program, the Hedgehog (Hh) pathway. To provide background for this part of the work, I will review the multiple patterning events that are controlled by Hh signals in the gastrointestinal (GI) tract and discuss the cellular targets of Hh signaling in this tissue.

Further mechanistic understanding of the developmental events that are controlled by Hh signaling, including patterning of the pylorus, requires identification of the molecular targets of Hh signals, but to date, few targets are known. I have developed a novel computational method for the prediction of cell-specific transcriptional targets of Hh signaling (see Chapter IV). To put this work in perspective, I will review existing tools for the recognition of signaling-regulated enhancers in genomic DNA and consider their strengths and weaknesses.

I.1. Embryonic development of the early GI tract

The GI tract is the primary site for digestion of food particles, absorption of nutrients, and excretion of solid waste. It extends from the mouth to the anus and includes both the tubular gut (i.e., esophagus, stomach, small intestine, and colon) and accessory digestive organs (i.e., pancreas and liver). The tubular gut is comprised of two concentric tissue layers, epithelium and mesenchyme, which are derived from endoderm and mesoderm, respectively. Epithelium lines the lumen of the gut tube and is responsible for its core digestive functions. Mesenchyme surrounds the epithelium, providing structural and vascular support and facilitating the flux of nutrients (i.e., motility) within the GI tract. Gut motility is partly controlled by innervations within the mesenchyme from the enteric nervous system, which is established via the migration of ectoderm-derived, neural crest cells. Here, I will outline the major events in early GI tract development and describe the expression of specific transcription factors and roles for signaling pathways.

Formation of the primitive gut

The three primary germ layers (i.e., ectoderm, mesoderm, and endoderm) are established during gastrulation. In the mouse, beginning at embryonic day (E) 6.5 epiblast cells migrate through the primitive streak and, under the influence of cell-cell signaling, become specified as mesoderm or endoderm [1]. The endodermal cells intercalate into the extraembryonic visceral endoderm on the ventral-most surface of the embryo, forming the definitive endoderm. By E9.5 coordinated morphogenic movements convert this sheet of endoderm into the primitive gut tube. The anterior endoderm folds posteriorly and laterally to form the anterior intestinal portal (AIP), and in an analogous manner, the posterior endoderm folds anteriorly and laterally to establish the caudal intestinal portal (CIP). These folds create the primitive foregut and hindgut pockets, respectively. Lateral folding of the endoderm continues posteriorly from the AIP and anteriorly from the CIP until the endoderm forms a closed tube.

Patterning of the primitive gut

Closure of the gut tube is accompanied by the migration of its associated mesoderm. By E9.5 the primitive gut exists as a single-layered tube of endoderm surrounded by mesoderm. Elegant lineage tracing and microarray experiments, however, indicate that patterning of the gut is initiated well prior to gut tube formation [2, 3]. This partitioning of the GI tract into broad organ domains involves precise spatiotemporal control of transcription factor expression and signaling pathway activity.

Transcription factors

A number of transcription factors show restricted expression patterns in primitive gut endoderm and mesoderm. Hox genes are particularly notable because their expression extends along the entire anteroposterior (A-P) axis of the gut [4, 5]. Overlapping, combinatorial expression of Homeobox (Hox) transcription factors delineates specific GI regions (i.e., foregut) or organ territories (i.e., stomach), while tight expression boundaries between Hox genes often correspond to transitions between adjacent organs. This "gastrointestinal Hox code" is one of the earliest apparent patterns in the GI tract and influences subsequent organogenesis of the gut.

As early as E8, there is clear A-P patterning of transcription factors in the endoderm. The forkhead box A (FoxA) subfamily and *Sry-box containing gene 2* (*Sox2*) are expressed broadly in foregut endoderm, while *caudal type homeobox 2* (*Cdx2*) expression demarcates midgut and hindgut endoderm [1]. Expression of *pancreatic and duodenal homeobox 1* (*Pdx1*) spans the foregut-midgut boundary and delineates a unique domain from which a trio of accessory digestive organs (liver, gall bladder, and pancreas) later emerge. The *Pdx1*-positive area of the tubular gut gives rise to the pylorus as well as the surrounding distal antrum and proximal duodenum.

The smooth muscle sphincters of the GI tract are critical structures during digestion and are associated with unique transcription factor expression. For example, *NK2 transcription factor related, locus 5 (Drosophila)* (*Nkx2-5*) is expressed specifically in a narrow domain of mesoderm at the future site of the pyloric sphincter (see Chapter I.3) [6].

Signaling pathways

There are at least seven major developmental signaling pathways including: Bone morphogenetic protein (Bmp), Hh, Jak/Stat, Notch, nuclear receptor, receptor tyrosine kinase, and Wnt [7]. These pathways interact with organ-specific transcription factors to pattern the primitive gut and direct organogenesis within the GI tract. In the early gut tube, fibroblast growth factor (Fgf; a receptor tyrosine kinase ligand), Wnt, and retinoic acid (RA; a nuclear receptor ligand) facilitate the establishment of the A-P axis [1]. Fgf and Wnt directly activate Cdx genes in the hindgut, while RA signaling and active inhibition of Wnt limits their anterior expression boundaries in the foregut. Subsequent Fgf, Bmp, and RA signals subdivide the anterior gut into specific organ domains, while FGF and Wnt promote the elaboration of an intestinal domain in the posterior gut.

Hh signaling acts at several points during patterning of the primitive gut, and loss of either *sonic hedgehog* (*Shh*) or *Indian hedgehog* (*Ihh*), genes encoding two of the three mouse Hh ligands, leads to GI defects [8]. *Shh* is required for proper dorsoventral (D-V) patterning of the anterior foregut. Loss of *Shh* disrupts separation of the trachea and esophagus, leading to trachea-esophageal fistulas, and causes alterations in antral as well as anal patterning [8, 9]. In contrast, specific downregulation of *Shh* is required for formation of the dorsal pancreatic bud [10]. In the intestine, Hh signaling controls Wnt and Bmp expression, in part via activation of FoxF and FoxL transcription factors, and is essential for epithelial and mesenchymal patterning later in gut development (see Chapter I.4) [11].

I.2. Establishing epithelial identity and morphology in the foregut

The foregut is comprised of a variety of organs with vastly different roles in digestion: the esophagus facilitates the transport of food and water from the mouth; the stomach breaks down food particles for digestion; and the duodenum begins the absorption of nutrients. These functions are carried out by specialized epithelia, each with unique tissue morphology, specific cell types, and distinct gene expression patterns. All of these regional or organ-specific attributes are initially patterned during primitive gut development. Here I will describe the role of specific transcription factors in establishing domains of organ-specific foregut epithelia. I will then summarize the morphological and temporal events accompanying differentiation of intestinal epithelium.

Transcription factors pre-pattern foregut endoderm

FoxA subfamily: endodermal competence

Vertebrate Fox genes encode transcription factors that are related to the *Drosophila melanogaster* (*D. melanogaster*) gene *fork head,* and Fox proteins have a characteristic winged helix ("forkhead box") DNA binding domain [12]. The FoxA subfamily members Foxa1, Foxa2, and Foxa3 were initially identified as liver transcription factors and termed hepatocyte nuclear factor-3 (Hnf3) alpha, beta, and gamma, respectively. The expression of FoxA genes, however, is not localized to the liver. *Foxa1* and *Foxa2* are expressed in the primitive streak, notochord, and endoderm throughout the primitive gut, while the expression of *Foxa3* is restricted to a domain of endoderm from the posterior foregut to

the hindgut. The overlapping expression patterns of FoxA genes indicate the possibility of genetic redundancy or compensation, and neither *Foxa1* nor *Foxa3* is uniquely required during embryogenesis [13, 14]. *Foxa2* null embryos, on the other hand, die by E11 due to loss of *Shh* expression in the notochord and significant non-endodermal defects [15].

Sophisticated genetic approaches have highlighted an important role for FoxA factors in the foregut. Endodermal loss of both *Foxa1* and *Foxa2* completely inhibits liver development, impairs differentiation of respiratory epithelium in the lung, and eliminates *Pdx1* expression in pancreatic endoderm [16-18]. In a site-specific manner FoxA factors bind directly to histones and reduce local chromatin compaction near organ-specific genes [19]. This makes the endoderm poised (or "competent") to respond to differentially expressed transcription factors. Thus, FoxA proteins act as "pioneer factors" for nuclear hormone receptors by facilitating their recruitment to cis-regulatory sequence of organ-specific genes [20].

Sox2: anterior foregut cell fates

Sox2 is a member of a large family of high mobility group (HMG) box transcription factors. There are twenty mouse Sox genes, which have been classified into nine subgroups [21]. Along with Sox1 and Sox3, Sox2 is a part of the SoxB1 subgroup of transcriptional activators. *Sox3* is expressed in foregut endoderm as early as E9.5 and is known to genetically interact with *Sox2* during formation of the second pharyngeal arch [22].

Sox2 is expressed in both A-P and D-V gradients. The highest expression is in the esophagus, anteriorly in the A-P gradient and dorsally in the D-V gradient [23]. Sox2 levels also correlate with epithelial morphology: stratified squamous (high) and simple columnar (low). In the early gut tube, the posterior boundary of Sox2 expression occurs at the pylorus, where it abuts the anterior border of intestinal Cdx2 expression until at least E14.5 [3]. This boundary may be maintained by mutual transcriptional repression interactions between Sox2 and Cdx2 (see below). While *Sox2* null mice die before gastrulation, mice carrying hypomorphic *Sox2* alleles survive to birth and show significant foregut defects [23, 24]. Genetic reduction of *Sox2* posteriorizes anterior foregut cell fates. For example, intestinal genes are misexpressed in the esophagus, and the anterior stomach displays the morphological characteristics of the posterior stomach [23]. Thus, while the FoxA family establishes a competence for subsequent organ-specific gene expression in foregut endoderm, Sox2 induces specific foregut cell fates in a dosage-dependent manner.

Cdx2: intestinal cell fate

The vertebrate Cdx genes *Cdx1*, *Cdx2*, and *Cdx4* encode transcription factors that are related to the *D. melanogaster* gene *caudal*, an essential gene for proper A-P patterning in flies [25]. *Cdx1* and *Cdx2* are expressed in overlapping patterns in the developing intestine. The anterior boundary of Cdx2 occurs at the pylorus and is maintained by mutual repressive interaction with Sox2 [3]. *Cdx1* null mice have no apparent gut phenotype and *Cdx2* null embryos die prior to gastrulation [26, 27]. Endoderm-specific

deletion of *Cdx2* causes a conversion of the simple columnar intestinal epithelium of the ileum to esophagus-like stratified squamous epithelium [28]. Loss of *Cdx2* is accompanied by upregulation of Sox2 and downregulation of Ihh. Conversely ectopic expression of *Cdx2* in the stomach leads to intestinal metaplasia, wherein patches of intestine-like epithelium are found within a field of normal stomach epithelium [29]. Thus Cdx2 is a fundamental regulator of intestinal cell identity: it is required not only to activate intestinal gene expression, but repress anterior Sox2-dependent cell fates.

Pdx1: pancreatic cell fate

Pdx1 is expressed in a narrow domain of posterior foregut endoderm, beginning at E8.5 [3]. Its expression extends from the distal stomach, across the pylorus, and into the proximal small intestine, and it does not adhere to the boundary formed by mutual Cdx2 and Sox2 repression. The dorsal and ventral pancreatic buds emerge from the gut tube within this *Pdx1* expression domain [1]. *Pdx1* null embryos have a small rudimentary pancreas that lacks insulin-secreting β-cells [30]. Loss of *Pdx1* also disrupts the morphogenesis of the proximal small intestine: there is loss of villi, absence of Brunner's glands, and conversion of intestinal to bile duct epithelium. Thus, Pdx1 specifies a unique domain of pancreatic progenitors and may contribute to the development of other specialized structures (i.e., Brunner's glands) at the pylorus.

Epithelial morphogenesis during villus emergence

By E9.5 broad organ domains have been established in the primitive gut. From E9.5-E14.5, the pancreatic, hepatic, and splenic anlagen emerge from a narrow region in the

distal foregut and undergo organ-specific morphogenic programs, while the stomach acquires its characteristic saclike shape [1]. At the same time, however, relatively minor changes occur in the epithelium of the tubular gut.

After gut tube formation the number of cells in the endodermal layer increases dramatically, and the sheet of intestinal endoderm becomes either a multi-layer stratified or pseudo-stratified epithelium [31]. (Unpublished preliminary data from our lab suggest that it is pseudo-stratified.) This transformation nearly fills the gut tube lumen with endodermal cells and increases epithelial girth. Then over the course of two days, from E14-E16, this thick tube of intestinal epithelial cells is dramatically remodeled. While lumenal surface slowly expands throughout the intestinal epithelium, mesenchymal condensations ("clusters") form at regularly spaced intervals beneath it [32]. As the epithelium is converted to a simple columnar epithelium, generating fingerlike projections known as villi, the clusters are enveloped by the forming villi and become the mesenchymal cores. Between these emerging villi is the proliferative intervillus epithelium, which will be remodeled postnatally to form the intestinal crypt. Thus concomitant with these changes in epithelial morphology is the primary establishment of the intestinal crypt-villus axis. The emergence of villi begins in the anterior intestine and continues posteriorly in a "wave" along the A-P axis [31]. By E16.5 the villi of the proximal duodenum have been formed and many of the differentiated cells types of the intestine are present: absorptive enterocytes, mucous-secreting goblet cells, and hormone-secreting enteroendocrine cells.

Shortly after villus formation begins, another dramatic patterning event can be discerned: the establishment of a sharp anterior boundary for intestinal gene expression. This boundary was first noted while examining *villin* gene expression [33]. At E14.5 robust *villin* expression extends throughout intestinal endoderm and into distal stomach endoderm, where it gradually decreases in a posteroanterior gradient. Two days later after the endoderm has been remodeled into a simple columnar epithelium, a tight, one-cell-thick anterior boundary of *villin* expression is discernable. This expression boundary corresponds to the transition between differentiated intestinal and gastric epithelial cell types along the GI tract.

I.3. Regulation of nutrient transport within the GI tract

Compartmentalization of organ-specific epithelia by sphincters of the GI tract facilitates the precise regulation of individual digestive functions, and malformation or dysfunction of these sphincters leads to significant human pathologies [34]. In the tubular GI tract sphincters control the passage of nutrients during digestion by adjusting the diameter of the lumen between adjacent organs (i.e., the stomach and intestine). There are six sphincters within the tubular GI tract: upper esophageal, lower esophageal, pyloric, ileocecal, internal anal, and external anal. All but the lower esophageal are anatomic (or "true") sphincters, which have distinguishable muscular thickenings that adjust lumenal diameter in response to muscular tone. In contrast, the lower esophageal sphincter, which separates the esophagus from the acidic environment of the stomach, is "physiological". It is not associated with an obvious muscular thickening, and its constriction is regulated

in part by differences in intrathoracic and intraabdominal pressure. The muscular control of anatomic sphincters is either voluntary (i.e., upper esophageal and external anal) or involuntary (i.e., pyloric, ilio-cecal, and internal anal). Here I will describe the structure, function, and development of the pylorus and its associated sphincter.

Control of gastroduodenal flux by the pylorus

Through its sphincter the pylorus regulates the timing of gastric emptying as well as the size of food particles that pass from the stomach into the duodenum [34]. It prevents the backflow of nutrients into the stomach during peristaltic contractions in the intestine and assists in the mechanical digestion of food by strong muscular contractions of the antrum.

The pylorus, like other areas of the tubular GI tract, is fundamentally comprised of two basic tissue layers, epithelium and mesenchyme. At the pylorus there is an abrupt one-cell-thick transition between differentiated gastric and intestinal epithelial cells (i.e., the epithelial pyloric border; see Chapter II) [33]. Directly surrounding the pyloric epithelium, the sub-mucosa is a layer of loose mesenchyme that contains myofibroblasts, smooth muscle cells (i.e., muscularis mucosa), blood vessels, and neurons [34]. Peripheral to the sub-mucosa, the thick smooth muscle of the muscularis externa defines the gross morphological shape of the pylorus. The serosa is a thin layer of mesothelial cells that comprises the outer surface of the tubular gut. There are numerous innervations throughout the pyloric mesenchyme, which control smooth muscle cell contraction and regulate the tone of the pyloric sphincer.

Muscularis externa

Flanking the pylorus in the distal antrum and proximal duodenum, the muscularis externa has two discrete smooth muscle layers: inner circular and outer longitudinal [34]. The orientation of muscle fibers in the inner circular and outer longitudinal layers is perpendicular and parallel to the long axis of the gut tube, respectively. Throughout much of the tubular gut, the inner circular layer is thicker than the outer longitudinal layer. There are, however, organ-specific differences in the muscularis externa. For example, it is much thicker in the stomach than the intestine.

Whereas the muscularis externa of the distal antrum and proximal duodenum is arranged in two discrete, continuous layers, the pyloric sphincter has a more complex muscular structure [34]. The inner circular layer of the pylorus is thicker than even the neighboring antral inner circular layer, and it is organized into fascicular bundles of fibers. These bundles are separated by dense connective tissue and travel at a slightly oblique angle to the long axis of the gut tube. In contrast the outer longitudinal layer of the pylorus is mostly continuous with the outer longitudinal layers of the distal antrum and proximal duodenum. This layer, however, is noticeably thicker at the pyloric sphincter, and many of its muscle fibers diverge and travel towards the epithelium, where they join muscle bundles of the inner circular layer.

Innervations

The entire GI tract is innervated by the enteric nervous system (ENS), a subdivision of the peripheral nervous system, which regulates the involuntary contractions of these

muscle layers during digestion [34]. There are two major types of neuronal ganglia in the ENS: myenteric (Auerbach's) and sub-mucosal (Meissner's) plexuses. Myenteric plexuses are located in the muscularis externa between the inner circular and outer longitudinal layers, where they integrate sympathetic and parasympathetic motor innervations to the GI tract to control peristaltic contractions necessary for the movement of food content.

The muscular tone of the pyloric sphincter is under tight neural control. It is influenced by direct sympathetic and parasympathetic extrinsic innervations as well as intrinsic innervations and local neuropeptide release [34]. Extrinsic sympathetic and parasympathetic innervations are primarily via the splanchnic and vagus nerves, respectively. The majority of these innervations are afferent and relay information to the central nervous system from mucosal and muscle sense receptors. Mucosal receptors respond to changes in lumenal content (i.e., nutrient load), while muscle receptors detect the mechanical load on GI smooth muscle (i.e., stretch from distention). Short- or long-range reflex pathways govern the pyloric response to these afferent stimuli.

Contraction of the pyloric sphincter is stimulated by adrenergic and cholinergic neurotransmitter release from extrinsic efferent sympathetic and parasympathetic innervations, respectively, while inhibition occurs via release of a non-adrenergic, non-cholingergic neurotransmitter, possibly vasoactive intestinal peptide (VIP) or nitric oxide (NO) from extrinsic efferent parasympathetic innervations [34]. In addition to extrinsic neuronal control, intrinsic neural pathways originating in the proximal duodenum and

distal antrum regulate the muscular tone of the pyloric sphincter and coordinate gastrointestinal responses to local stimuli. These intrinsic innervations affect adrenergic- and cholinergic-dependent extrinsic innervations via the release of neuropeptides including VIP, cholecystokinin, substance P, and neuropeptide Y.

Infantile Hypertrophic Pyloric Stenosis

Pyloric dysfunction has a tremendous impact on digestion and human health. Infantile Hypertrophic Pyloric Stenosis (IHPS) is one of the most common gastrointestinal disorders in newborns; the incidence is approximately 3 per 1,000 live births [35]. It is associated with significant hypertrophy of the pyloric sphincter muscle and narrowing of the pyloric lumen. These anatomical changes lead to gastric outlet obstruction, which results in characteristic projectile vomiting at 2-8 weeks of age. Without proper medical care children are unable to feed and rapidly become dehydrated. In the early 20^{th} century IHPS was often fatal, but today IHPS can be effectively treated by pyloromyotomy surgery, during which the hypertrophied muscle is cut by a longitudinal incision in the anterior wall of the pylorus.

The precise etiology of IHPS is unknown. In biopsies of hypertrophied pyloric sphincter muscle from IHPS patients, the distribution of neuronal subtypes is significantly altered [35]. It is unclear, however, whether this neuronal phenotype is a cause or result of the muscular hypertrophy. Interestingly mice that lack *neuronal nitric oxide synthase* (*nNOS*) display hypertrophy of the pyloric sphincter, and a similar pyloric phenotype was obtained in newborn rats by administering L-NAME, an chemical inhibitor of nitric oxide

synthase, to pregnant dams [36, 37]. Given the potential role for NO in mediating relaxation of the pyloric sphincter muscle, this suggests that the muscular hypertrophy in IHPS patients may be due in part to decreased NO signaling. However no clear genetic cause has been identified for IHPS [35].

Embryonic development of the pylorus

The genetic network of pylorus development is well conserved across vertebrates, despite tremendous variation in GI tract morphology [38]. The mouse has two different types of stomach epithelium. The anterior stomach, which is also known as the forestomach, has stratified squamous epithelium and is similar to the esophagus. The posterior stomach, on the other hand, has simple columnar epithelium that is organized into deep glands. Within this region of glandular stomach epithelium, there are distinct domains with specialized functions, cell types, and gene expression. Cells of the anterior glandular epithelial region (i.e., the corpus) secrete digestive enzymes and acid, while those of the posterior glandular epithelial region (i.e., the antrum) secrete mucus and regulatory hormones (e.g., gastrin). The pylorus, the most distal part of the posterior stomach, is the site of the pyloric sphincter, and pyloric epithelium is similar to that of the antrum.

Unique gene expression domains in the posterior foregut mesoderm, site of the future pylorus

Above, I summarized the establishment of organ-specific gene expression in posterior foregut endoderm (see Chapter I.2). Similarly, distinctive gene expression patterns emerge in the posterior stomach and presumptive pyloric mesoderm: the transcription

factor BarH-like homeobox 1 (Barx1) is expressed throughout the stomach; the secreted signaling molecules Wnt5a and Bmp4 are restricted to the anterior glandular stomach; and the transcription factors sine oculis-related homeobox 2 homolog (Drosophila) (Six2) and NK3 homeobox 2 (Nkx3-2/Bapx1) are localized to the posterior glandular stomach [38]. On the other hand *Bmp4* and *Nkx2-3* are expressed in the intestine. This leaves a domain in the posterior glandular stomach that is devoid of Bmp4, and in this region, *Bmp4* expression may be inhibited by Bapx1 and/or Six2 [39, 40]. This domain corresponds precisely with the expression of *bone morphogenetic protein receptor, type 1A* (*Bmpr1a*) [38]. Gremlin (Grem1), a secreted inhibitor of Bmp signaling, is expressed specifically at the pylorus [41]. The expression of *gremlin*, coupled with the expression of *Bmpr1a* in the posterior glandular stomach and pylorus, suggests that Bmp signaling from the surrounding tissue may be important for patterning the posterior glandular stomach and/or the pylorus.

Molecular mechanisms of pylorus development

Activation of *Bmp4* in the mesenchyme of the intestine by secreted Hh from foregut endoderm is one of the earliest events in pylorus development [42]. Secreted Bmp4 ligand binds to BMP receptors (including Bmpr1a) in the posterior stomach and subsequent BMP pathway activity cell-autonomously activates *Nkx2-5* and *Sox9* [43, 44]. Since misexpression of *Nkx2-5* does not affect *Sox9* expression (and *vice versa*), it appears that these two pathways are independently activated by Bmp [41, 43].

In the chicken, the posterior glandular stomach is called the gizzard, and gizzard

epithelium can be distinguished from more posterior pyloric epithelium by its histological appearance. Misexpression of *Bmp4*, *Bmpr1b*, *Nkx2-5* or *Sox9* in gizzard mesoderm converts the overlying endoderm to pyloric epithelium-type histology [6, 41, 43, 44]. Conversely pyloric endoderm can be converted to gizzard-type epithelium by misexpression of dominant-negative forms of Nkx2-5 (enrepNkx2-5) or Sox9 (Sox9ΔCter) in pyloric mesoderm (although epithelial conversion by enrepNkx2-5 is incomplete) [41, 44].

Misexpression of *gremlin* in gizzard mesoderm yields a similar epithelial phenotype to *Bmp4*, *Bmpr1b*, *Nkx2-5* or *Sox9* misexpression, suggesting that Nkx2-5 and/or Sox9 may specify pyloric epithelium indirectly through regulation of *gremlin* expression [41]. In support of this hypothesis, misexpression of *Sox9* or *Sox9ΔCter* in stomach mesoderm expands or reduces *gremlin* expression, respectively. Since Nkx2-5 does not regulate *Sox9* and a regulatory relationship between Nkx2-5 and *gremlin* is unclear, the mechanism by which Nkx2-5 controls pyloric epithelial phenotype remains unresolved.

In the mouse, pyloric *Nkx2-5* expression first appears around E9.5 in discrete, bilateral mesodermal domains that flank the primitive gut tube [45]. A day later at E10.5 these domains expand to form a complete, circumferential ring around the gut at the site of the future pylorus. This domain also gives rise to the spleen. A condensation of splenic precursor cells (SPCs) that express *Nkx2-5* arises on the dorsal side of the pyloric band. From E10.5-E14.5 SPCs migrate anteriorly through the gastric mesothelium, along the greater curvature of the stomach. After E14.5 the migration of SPCs ends, and there are

distinct pyloric and splenic domains of *Nkx2-5* expression.

A number of genetic mouse models exist for the study of pyloric development. *Nkx2-5* null embryos have significant heart defects and die by E10, while *Sox9* heterozygous mice display skeletal abnormalities and die perinatally [46, 47]. Thus in the absence of appropriate conditional genetic models the lethality of these mutations has precluded a functional analysis of either *Nkx2-5* or *Sox9* during mouse pylorus development. *Gremlin* null mice show significant limb defects and die perinatally, but a pyloric phenotype has not been reported [48]. In contrast, mutant mouse models for *Six2*, *Bapx1*, and *Barx1* are all associated with loss of the pyloric constriction [39, 49, 50]. While neither loss of *Six2* nor *Bapx1* significantly alters *Nkx2-5* expression, the expression of Sox9 in pyloric mesenchyme is almost completely lost in *Six2* null embryos [39, 50]. In addition loss of *Six2* is associated with an expansion of *Bmp4* expression into posterior stomach mesoderm [39].

These data indicate that the genetic network of pylorus development is largely conserved between chicken and mouse. Still many questions remain about the molecular details of pylorus patterning. To address this issue, a part of my thesis work involved a microarray analysis of pylorus development (see Chapter II). From these studies we uncovered many more genes with exquisitely precise patterning at the pylorus and focused on the functional role of one of these, the zinc finger transcription factor GATA binding protein 3 (Gata3), in pylorus development. This study successfully revealed changes in several signaling pathways during pylorus development. One of the most dramatic changes was

the downregulation of Hh signaling in the intestine, but not the stomach. Previous studies in the *Xenopus* model had shown that a reduced level of Hh signaling in the intestine is required for proper cytodifferentiation of that organ [51]. While Hh is downregulated in intestine at this time, it is not extinguished, and Hh signaling continues to play important roles in smooth muscle development, lipid absorption, and inflammatory signaling in the adult intestine.

The fact that Hh signals are centrally involved in both embryonic and adult gut morphogenesis and function led me to further examine the molecular targets through which Hh signals impact gene expression. In the following sections, I will first review the roles of Hh in gut tube patterning. Second, I will discuss what is known about Hh-regulated transcriptional enhancers. Finally, I will analyze current *in silico* tools for detection of signaling-regulated enhancers.

I.4. Hh signaling in the late embryonic and adult GI tract

Throughout embryonic and adult life from the antrum to the colon, Hh signaling is exclusively paracrine: Hh ligands are secreted from the epithelium and signal to the surrounding mesenchyme [52]. Hh signaling is essential for the development and maintenance of mesenchymal and epithelial cell types in the gut; mice with perturbed Hh signaling have significant GI defects [8]. In much of the GI tract, Hh signals pattern the mesenchyme directly but epithelium indirectly.

Distinct mesenchymal cell populations in the gut transduce Hh signals [52]. For example, smooth muscle cells of the muscularis externa, muscularis muscosa, and villus core as well as mesothelial cells of the serosa respond to Hh signals, while neurons of the enteric nervous system do not. Myeloid cell populations in the colon, including dendritic cells and macrophages, can also directly respond to Hh signals [53]. Here I will summarize the molecular components of the Hh signaling pathway. Then, I will describe late developmental roles for Hh signaling in stomach and intestine and a homeostatic function for Hh in adult intestine.

The Hh pathway is highly conserved in eukaryotes

Invertebrates: Hh signaling in D. melanogaster

Components of the Hh pathway were first identified during a forward genetic screen for embryonic lethal mutations in *D. melanogaster* [54]. Genetic mutants for hedgehog (Hh), the sole *D. melanogaster* Hh ligand, or its transmembrane receptor patched (Ptc) have characteristic larval cuticle patterns arising from segmentation defects in the early embryo. In the absence of Hh ligand, Ptc inhibits the activation of the transmembrane protein smoothened (Smo). When Smo activation is inhibited, the downstream transcriptional effector cubitus interruptus (Ci) is proteolytically-cleaved to a repressor form (Ci^R) [55]. Upon binding of Hh to Ptc, Smo is activated and prevents formation of Ci^R. At high levels of Hh binding, Smo facilitates the production of activator Ci (Ci^A).

Vertebrates: Hh signaling in the mouse

With a few notable exceptions Hh signaling in vertebrates is directly analogous to *D.*

melanogaster. In the mouse there are three ligands [i.e., Shh, Ihh, and desert hedgehog (Dhh)], two receptors [i.e., patched homolog 1 (Ptch1) and Ptch2], and three intracellular transcriptional effectors [i.e., GLI-Kruppel family member GLI1 (Gli1), Gli2, and Gli3] [55]. In the absence of Hh ligand, Ptch1 inhibits smoothened homolog (Smo) (orthologous to *D. melanogaster* Smo), and Gli2 and Gli3 are converted to repressor forms (Gli^R). Hh ligand binding to Ptch1 releases its inhibition of Smo activation and promotes formation of activated Gli2 and Gli3 (Gli^A). Gli1 is an obligate transcriptional activator. Gli1 levels are not directly regulated by Smo, but instead are controlled transcriptionally via Gli^A.

Hh signaling affects epithelial and mesenchymal cells in the GI tract

Intestinal epithelium

While neither *Shh* nor *Ihh* is required *per se* for the formation of intestinal villi, a transgenic model of strong, intestine-specific Hh inhibition [expression of a soluble form of Hedgehog-interacting protein (Hhip), a pan-Hh inhibitor, in intestinal epithelium; 12.4KVillin-Hhip] shows significant epithelial defects, including loss of villi in animals in which *Hhip* is expressed at high levels [8, 56]. These mice exhibit neonatal wasting due to perturbations in enterocyte differentiation and poor intestinal absorption [56]. In mice expressing lower levels of the *Hhip* transgene, the crypt/villus axis pattern is perturbed: ectopic pre-crypt pockets are found on the upper villus, where they contribute to bizzarely branched villi. Both villus smooth muscle patterning and myofibroblast localization is perturbed as well.

Stomach epithelium

Loss of *Shh* leads to a dramatic overgrowth of glandular epithelium in the posterior stomach [8]. While *Gli3* null mutants show similar glandular overgrowth, *Gli2* null mutants have apparently normal stomachs [57]. This suggests that, in contrast to its predominant role as a repressor in other tissues (i.e., the neural tube), Gli3 may function primarily as an activator in the stomach.

Muscularis externa

Hh signaling from the endoderm is also required for the development and proper positioning of smooth muscle in the intestine. High levels of epithelial Hh signaling are thought to inhibit smooth muscle differentiation via activation of *Bmp4*, and exclusion of smooth muscle from the area immediately surrounding the gut tube may be required for the correct positioning of the muscularis externa [58]. Alternatively Hh may act directly on smooth muscle of the muscularis externa, as loss of either *Shh* or *Ihh* shows a reduced muscularis externa [8]. A similar smooth muscle phenotype is observed in the 12.4KVillin-Hhip model of decreased intestinal Hh signaling, as well as *Foxf2* null and *Foxf1*/*Foxf2* compound heterozygote mice [11, 56]. *Foxf1* and *Foxf2* may also be targets of Hh signaling in the intestine [11].

Villus smooth muscle

Unpublished data from our laboratory indicates that Hh signals control differentiation of the smooth muscle of the villus core, which is a distinct muscle population from the muscularis externa. Villus smooth muscle is dramatically expanded in postnatal

transgenic mice that overexpress Ihh in the intestine (Villin-Ihh). Conversely, villus smooth muscle differentiation is significantly reduced by inhibition of intestinal Hh signaling in postnatal bitransgenic mice (VFHhip; *Villin-Cre* x *Villin-loxP-LacZ-loxP-Hhip*). Exogenous application of Hh ligand or forced expression of constitutively active Gli2 in cultured smooth muscle precursor cells stimulates smooth muscle differentiation and upregulates the smooth muscle master regulator *myocardin* (*Myocd*).

Immune cells

Finally, recent evidence implicates Hh signaling in the homeostatic control of immune cells in the adult intestine [53, 59]. In three separate European sample populations, a non-synonymous SNP in the coding region of *GLI family zinc finger 1* (*GLI1*) (the human ortholog of *Gli1*) is associated with inflammatory bowel disease, in particular ulcerative colitis [53]. This mutation decreases GLI1 transactivation in transfected cells, and *Gli1* heterozygous mice exhibit increased susceptibility to an inflammatory stimulus. In addition reduced intestinal Hh signaling in VFHhip transgenic mice is associated with spontaneous inflammation [59]. Taken together these data suggest that reductions in Hh pathway activity may be pro-inflammatory.

I.5. Hh regulation of gene expression

The diversity of cellular responses to a single source of Hh signal in the GI tract suggests a complex regulation of Hh target gene activation in gut mesenchyme. Secreted Hh ligands can act as morphogens and elicit distinct transcriptional responses at discrete

distances from the source of the signal [60]. Examples of Hh morphogen activity are found throughout eukaryotic evolution, including the larval imaginal wing disc in *D. melanogaster* and the neural tube in mice [61, 62]. Cell type-specific transcriptional response of Hh target genes is tightly controlled by a number of mechanisms, including sequence-specific binding of the repressor (Ci^R/Gli^R) and activator (Ci^A/Gli^A) forms to cis-regulatory elements (i.e., enhancers) in genomic DNA [55]. Here I will introduce known Hh-regulated enhancers in *D. melanogaster* and outline *in silico* methods to identify novel enhancers from analysis of genomic sequence.

The structure of Hh-regulated enhancers in *D. melanogaster*

In *D. melanogaster* Hh signaling is essential for establishing segment polarity in the early embryo and A-P patterning in the larval imaginal discs [55]. During segmentation *hh* is expressed in repeating stripes of epidermal cells at the parasegmental borders [63]. Secreted Hh diffuses anteriorly and posteriorly to activate target gene expression. Thus many early embryo Hh targets are expressed in repeating stripes, including the Wnt ortholog *wingless* (*wg*). In the larval imaginal wing and leg discs, Hh is expressed in the posterior compartment and signals to the anterior compartment [55]. Hh activates expression of wing and leg disc target genes, including the Bmp ortholog *decapentaplegic* (*dpp*), in a stripe of cells along the A-P boundary. Hh-regulated enhancers have been characterized for *ptc*, *wg*, *dpp* and the Notch-regulated *Hes1* ortholog *hairy* [64-67]. Despite their shared responsiveness to Hh signals, these enhancers exhibit substantial structural variation.

ptc canonical enhancer

ptc is the canonical Hh target in *D. melanogaster* and one of few known orthologous targets (*Ptch1*) in the mouse [68, 69]. In Hh-responsive tissues activation of the Hh inhibitor Ptc serves as an important negative feedback mechanism that controls the range and duration of Hh signaling [70]. In larval imaginal wing discs of transgenic flies the *ptc* canonical enhancer activates reporter gene expression in a thin stripe of cells along the A-P compartmental boundary adjacent to Hh-expressing cells [64]. The *ptc* canonical enhancer is about one kilobase (kb) upstream of the *ptc* transcriptional start site and contains three closely-spaced, optimal consensus Ci binding sites (GACCACCCA). Deletion of these sites completely eliminates A-P compartmental boundary expression *in vivo*.

h leg disc enhancer

In leg imaginal discs the basic helix-loop-helix (bHLH) transcription factor hairy is expressed in a dorsal stripe along the A-P compartmental boundary, where it represses the formation of microchaete sensory bristles [71]. Hh is secreted by cells in the posterior compartment and signals to cells in the anterior compartment, and Hh pathway activity activates *hairy* expression in a narrow domain, just a few cells anterior to the A-P compartmental boundary. This pattern is regulated by the *hairy* leg disc enhancer, which is located between 30-35 kb downstream of the *hairy* transcriptional start site [67]. This enhancer contains two putative, non-optimal Ci binding sites: GACCTCCCA and GACCACCAT (nucleotide deviations from the optimal consensus are underlined). Directed mutagenesis of these sites reduced competitive binding to Ci protein *in vitro*,

and the presence of either mutant site (or both) eliminated reporter gene expression along the A-P compartmental boundary *in vivo*.

wg embryo enhancer

In the early embryo Wg regulates denticle formation, which leads to alternating bands of naked and denticular cuticle [54]. During segmentation, Hh is secreted from cells immediately posterior to the parasegmental borders, and Hh signaling in adjacent anterior cells activates *wg* expression via the *wg* embryo enhancer [65, 72]. This enhancer is located approximately five kb upstream of the *wg* transcriptional start site and contains four putative, non-optimal Ci binding sites: GAGCAGCCA, GTCCACGCT, GTTCACGCA, and GACCTCCCA (nucleotide deviations from the optimal consensus are underlined) [65]. Cumulative, directed mutagenesis of these sites reduces response to Hh signaling *in vitro* and specifically eliminates the segmental stripe pattern of reporter gene expression in the early embryo *in vivo*.

dpp wing disc enhancer

dpp is essential for proper A-P patterning of the wing imaginal disc and like *ptc* is expressed in a stripe along the A-P compartmental boundary [73, 74]. Unlike *ptc*, however, *dpp* is excluded from cells just anterior to the A-P compartmental boundary, likely via transcriptional repression by the homeodomain protein engrailed (en) [75]. Hh pathway activity in anterior compartment cells directs *dpp* expression via the *dpp* wing disc enhancer, which is located more than 20 kb downstream of the *dpp* locus and contains three putative, non-optimal Ci binding sites: GGCCACCTA, GACCGCCCG,

and TACCTCCCC (nucleotide deviations from the optimal consensus are underlined) [66]. In the wing disc, Hh signaling establishes opposing A-P gradients of Ci^R and Ci^A; Ci^R levels are highest furthest from the A-P compartmental boundary and *vice versa* for Ci^A. Deletion or directed mutagenesis of the putative Ci sites in the *dpp* wing disc enhancer expands (and shifts) reporter gene expression anteriorly *in vivo*. This suggests that in addition to Hh-independent local activators, input from both Ci^R and Ci^A are required for proper transcription control by the *dpp* wing disc enhancer. Ci^R defines the anterior *dpp* expression boundary by limiting the effect of local activators to the region transducing Hh signals, while Ci^A is necessary for activation close to the A-P compartmental boundary.

Summary: general properties of Hh-regulated enhancers?

It is clear from the varied location and structure of these enhancers that there is tremendous mechanistic diversity of Hh-regulated gene expression. At the same time, however, several important commonalities begin to emerge including the type and distribution of Ci binding sites. Despite being essential for proper reporter gene expression *in vivo*, most of the putative sites in these enhancers do not correspond to the optimal consensus. In fact of the enhancers described above, optimal consensus Ci binding sites are only found in the *ptc* canonical enhancer, which has three such sites clustered tightly together [64]. *ptc* is unique in that it is a universal Hh-responder (i.e., it responds to Hh signals in all cellular and tissue contexts). All other characterized Hh enhancers are context specific and may thus require non-optimal sites. Indeed, unpublished work from Dr. Scott Barolo's laboratory (Department of Cell and

Developmental Biology, University of Michigan) shows that substitution of optimal consensus Ci sites for non-optimal Ci sites in the *wg* embryo enhancer and *dpp* wing disc enhancer disrupts reporter gene expression. These results indicate that non-optimal Ci sites are *required* for proper activation of some Hh-regulated enhancers.

Another characteristic of these enhancers is the presence of multiple putative Ci binding sites. While the requirement for individual sites has not been rigorously examined in all of the enhancers, analysis of the *hairy* leg disc enhancer revealed that mutagenesis of either Ci site significantly altered reporter gene expression [67]. Similarly mutagenesis of the *dpp* wing disc enhancer demonstrated the individual requirement for two of its three putative Ci sites [66]. Taken together these data suggest that clustering of non-optimal Ci binding sites may be a general property of many cell type-specific Hh-regulated enhancers. Indeed, functional clustering of transcription factor binding sites has been noted in several other contexts [76-78].

Genome-wide methods of enhancer identification

Experimental approaches

The identification of enhancers is integral to the establishment of transcriptional relationships in gene regulatory networks, and until recently this was a laborious experimental task analogous to the mapping and cloning of genes [79]. For example in *D. melanogaster* large-scale, genome-wide “enhancer trapping” screens were carried out by random insertional transgenesis [80]. Reporter gene expression patterns of interest (e.g., larval imagine wing disc) were selected for further analysis, and once the genomic

position of the transgenic insertion was mapped, surrounding genomic regions could be analyzed for their ability to direct transgene expression in the expected pattern [81].

The emergence of chromatin immunoprecipitation (ChIP) techniques to characterize protein-DNA interactions in cells and tissues has changed the way enhancers are studied [82]. In particular, ChIP experiments that are combined with new technologies for genome-wide nucleotide profiling (i.e., ChIP-on-chip and ChIP-seq) have greatly increased the identification of cell- and tissue-specific DNA binding events for a given transcription factor [83]. However the degree of correlation between transcription factor binding events and enhancer activity for a given ChIP experiment is unpredictable and likely depends on both intrinsic biological properties and extrinsic experimental variables. Technical issues arising from the ChIP experiment itself (i.e., sensitivity and specificity of the immunoprecipitation step) can affect the reliability of the results, and the interpretation of ChIP data is subject to the spatiotemporal limitations of the analyzed cells or tissues. Recent Gli1 and Gli3 ChIP experiments in neuralized mouse embryoid bodies (EBs) and the developing mouse limb, respectively, have identified many novel Gli binding regions in the genome [84, 85]. While a surprising number of these regions are bound by Gli protein in both neuralized EBs and the limb, their Gli protein-binding status in intestinal mesenchyme, for example, is unclear.

In silico approaches: general considerations

The availability of high quality sequence data in the post-genome era, concomitant with the elucidation of DNA binding preferences for specific transcription factors, has

facilitated the development of genome-wide *in silico* methods to identify potential enhancers. A fundamental aspect of many of these approaches is finding the location of putative transcription factor binding sites in genomic sequence, which can be accomplished in a variety of ways. For example a set of consensus binding sites for a given transcription factor can be represented by an IUPAC nucleotide string, and matches to that string can be identified quickly in genomic sequence [86]. While simple and convenient, information about the binding preferences for that transcription factor is lost, including position-specific tolerance (or intolerance) to nucleotide substitution.

An alternative to IUPAC string representation is the construction of a mono-nucleotide distribution matrix, which contains the complete frequency distribution of nucleotides for a set of transcription factor binding sites [87]. In such matrices there are four rows, each pertaining to a specific nucleotide (A, C, G or T) and *n* columns (where *n* is the number of positions in the binding site). Each element in the matrix corresponds to the relative frequency (or weight) of a given nucleotide at a particular position. Hence mono-nucleotide distribution matrices are commonly referred to as position-weight matrices (PWMs). Matrix representation facilitates the analysis of putative binding sites and, in particular, non-optimal sites, including those that were not present in the initial set or are difficult to detect because of nucleotide substitutions at multiple positions. PWMs are also suitable for representing data from *in vitro* DNA binding assays [88].

There are at least two methods for evaluating binding sites with PWMs. The log-odds score compares the probability of a putative site given the distribution of nucleotides in a

PWM to the probability in "background" sequence. Large, positive values indicate a high degree of similarity to the PWM and potential binding [89]. A weakness of the log-odds score is its reliance on the estimation of background probability. Alternatively the MatInspector method gauges the similarity of a putative site to the optimal consensus site ("% matrix similarity"); a high % matrix similarity score indicates potential binding [90]. This method weights specific positions in a PWM by the "conservation" (i.e., relative distribution) of nucleotides. Positions with invariant nucleotides are maximally weighted, while those with evenly-distributed nucleotide frequencies are unweighted. Thus nucleotide substitutions at unweighted positions have little effect on the overall % matrix similarity, while those at maximally weighted positions significantly reduce % matrix similarity. In either case a minimum score threshold must be set in order to establish a set of potential binding sites. Changing the threshold alters the balance between sensitivity and specificity of binding site selection (i.e., raising the threshold increases specificity at the expense of sensitivity).

Once putative transcription factor binding sites have been identified in genomic sequence, their distribution can be analyzed for patterns correlated with known enhancers. For example because enhancers are often associated with regions of high interspecies sequence conservation, many *in silico* approaches utilize pairwise or multiple species sequence alignments.

In silico approaches: signaling-regulated enhancers

Two methods have been described for the identification of Hh-regulated enhancers in

vertebrates: Enhancer Element Locator (EEL) and Module Cluster Analysis (MCA) [84, 91]. EEL locates conserved binding sites in pairwise orthologous sequence alignments and uses PWMs and log-odds scores to estimate the energy of transcription factor binding (i.e., "primary interaction") [91]. It then assesses the potential energy from physical interaction between transcription factors (i.e., "secondary interaction") by measuring the predicted distance and angle between consecutive conserved sites in DNA. Finally EEL calculates a score to estimate the relative contributions of primary and secondary interactions to the total energy of an enhancer element, and this score can be used to rank potential enhancers. Using genome-wide, mouse-human pairwise sequence alignments and PWMs for Gli proteins, EEL identified a set of predicated Hh-regulated enhancers. In transgenic mice three of four tested enhancers showed reporter gene patterns consistent with Hh target gene activation.

In contrast to EEL, MCA analyzes orthologous sequence alignments of multiple, evolutionarily-distant vertebrate species (i.e., mouse, rat, human, dog, and zebrafish), and identifies highly conserved, non-coding sequence in the mouse genome [84]. It then uses PWMs and log-odds scores to locate transcription factor binding sites in the conserved sequence. The occurrence of sites in these sequences is modeled as a Poisson process, and MCA calculates a score to measure the enrichment of binding sites and rank potential enhancers. Using a PWM for Gli1 derived from ChIP-on-chip experiments in neuralized embryoid bodies (EBs), a genome-wide screen with MCA identified a set of predicted Hh-regulated enhancers, and a small subset (7 of 16) of these enhancers was associated with novel Gli1 ChIP peaks in neuralized EBs.

Another *in silico* approach, Site Clustering Over Random Expectation (SCORE), has successfully identified enhancers in *D. melanogaster*, and in contrast to EEL and MCA it does not consider sequence conservation [77]. Instead SCORE assesses the genome-wide distribution of binding sites and discriminates between statistically non-significant and significant clustering. SCORE tallied the length of each unique cluster of *n* sites in the *D. melanogaster* genome. It then compared the observed frequency of a particular length for a given cluster size *n* to the average expected frequency from repeated, independent Monte Carlo simulations to calculate a "purity" score [purity = (observed – average expected) / observed]. Monte Carlo simulations were performed by randomizing binding site positions within the entire genomic space and repeating the tallying process used for the *D. melanogaster* genome. A particular length for a given cluster size *n* was classified as statistically significant if the expected frequency was greater than the observed frequency in fewer than five out of 1,000 Monte Carlo simulations. The purity score of statistically significant clusters gives an estimation of the false positive rate (i.e., high purity scores have high specificity). Using a set of consensus sites for the Notch pathway transcriptional effector Suppressor of Hairless [Su(H)], SCORE identified a set of predicted Su(H)-regulated enhancers in *D. melanogaster*. These enhancers are correlated with known Su(H) transcriptional targets, and at least one enhancer directs reporter gene expression in a Su(H)-dependent manner.

Despite its success in detecting Notch-regulated enhancers in *D. melanogaster*, the SCORE method has a number of potential weaknesses. First, it relies on IUPAC string

representation to define the set of input binding site sequences, and consequently misses many non-optimal sites. Second, its clustering parameters are set manually prior to the analysis, instead of being derived from the sequence data itself. Third, the contribution of multiple overlapping statistically significant clusters to a single binding site is not assessed. Fourth, the Monte Carlo method of background estimation does not aptly account for the effect of %G-C content on the expected binding site distribution. Finally, SCORE analysis is not integrated with an assessment of evolutionary conservation, either sequence or functional. Taken together these factors reduce the sensitivity and specificity of SCORE, and indicate the need for computational methods that improve enhancer detection *in silico*. Such an approach is detailed in Chapter IV and should be an important tool in the identification of Hh-regulated enhancers.

I.6. Summary and thesis outline

Despite early embryonic patterning of posterior foregut endoderm by signaling molecules and the FoxA, Cdx, Sox2, and Pdx1 transcription factors, mature intestinal and stomach epithelial cell identities (i.e., expression of terminal differentiation genes) do not emerge until late embryogenesis. This suggests that a separate late patterning event may be required to establish epithelial cellular identity at the pylorus, and we hypothesized that such an event takes place between E14.5 and E16.5, concomitant with the restriction of *villin* expression to intestinal epithelial cells. In Chapter II, I describe transcriptomic changes surrounding the pylorus between E14.5 and E16.5 and report a unique domain of gene expression at the pylorus itself.

From the data presented in Chapter II, *Gata3* was identified as a novel pylorus-specific gene at E14.5 and E16.5. A role for *Gata3* in gut development has never been described, but we hypothesized that *Gata3* may be required for morphogenesis of the pylorus. In Chapter III, I describe the effect of loss of *Gata3* on the development of the pylorus and its smooth muscle sphincter.

Despite the paucity of confirmed Hh target genes in intestinal mesenchyme, Hh-dependent, cell-specific gene expression is essential for gut development and adult intestinal homeostasis. We hypothesized that Hh-regulated enhancers could be prospectively identified *in silico*. In Chapter IV, I outline a method to analyze the genome-wide distribution of Ci binding sites in *D. melanogaster* and describe the precision of subsequent enhancer predictions. Finally, in Chapter V, I will discuss the application of this thesis to future research.

I.7. Attribution

The first draft of this chapter was written by Aaron Udager. The text was revised with suggestions from Dr. Deborah Gumucio, Dr. Linda Samuelson, Dr. Gregory Dressler, and Mr. Ajay Prakash.

I.8. Publication

Text from this chapter has been revised by Aaron Udager, Mr. Ajay Prakash, and Dr. Deborah Gumucio for publication as a chapter in "Molecular Biology of Digestive Organs," a forthcoming volume in the Elsevier series *Progress in Molecular Biology and Translational Science*.

I.9. References

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

2. Tremblay, K.D., and Zaret, K.S. (2005). Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. Dev Biol *280*, 87-99.

3. Sherwood, R.I., Chen, T.Y., and Melton, D.A. (2009). Transcriptional dynamics of endodermal organ formation. Dev Dyn *238*, 29-42.

4. Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K., and Yamamura, K. (2002). Region-specific gastrointestinal Hox code during murine embryonal gut development. Dev Growth Differ *44*, 77-84.

5. Sekimoto, T., Yoshinobu, K., Yoshida, M., Kuratani, S., Fujimoto, S., Araki, M., Tajima, N., Araki, K., and Yamamura, K. (1998). Region-specific expression of murine Hox genes implies the Hox code-mediated patterning of the digestive tract. Genes Cells *3*, 51-64.

6. Smith, D.M., and Tabin, C.J. (1999). BMP signalling specifies the pyloric sphincter. Nature *402*, 748-749.

7. Barolo, S., and Posakony, J.W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes Dev *16*, 1167-1181.

8. Ramalho-Santos, M., Melton, D.A., and McMahon, A.P. (2000). Hedgehog signals regulate multiple aspects of gastrointestinal development. Development *127*, 2763-2772.

9. Litingtung, Y., Lei, L., Westphal, H., and Chiang, C. (1998). Sonic hedgehog is essential to foregut development. Nat Genet *20*, 58-61.

10. Hebrok, M., Kim, S.K., and Melton, D.A. (1998). Notochord repression of endodermal Sonic hedgehog permits pancreas development. Genes Dev *12*, 1705-1713.

11. Ormestad, M., Astorga, J., Landgren, H., Wang, T., Johansson, B.R., Miura, N., and Carlsson, P. (2006). Foxf1 and Foxf2 control murine gut development by limiting mesenchymal Wnt signaling and promoting extracellular matrix production. Development *133*, 833-843.

12. Friedman, J.R., and Kaestner, K.H. (2006). The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci *63*, 2317-2328.

13. Kaestner, K.H., Katz, J., Liu, Y., Drucker, D.J., and Schutz, G. (1999). Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. Genes Dev *13*, 495-504.

14. Shen, W., Scearce, L.M., Brestelli, J.E., Sund, N.J., and Kaestner, K.H. (2001). Foxa3 (hepatocyte nuclear factor 3gamma) is required for the regulation of hepatic GLUT2 expression and the maintenance of glucose homeostasis during a prolonged fast. J Biol Chem *276*, 42812-42817.

15. Filosa, S., Rivera-Perez, J.A., Gomez, A.P., Gansmuller, A., Sasaki, H., Behringer, R.R., and Ang, S.L. (1997). Goosecoid and HNF-3beta genetically interact to regulate neural tube patterning during mouse embryogenesis. Development *124*, 2843-2854.

16. Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.L., Wert, S., Stahlman, M.T., and Whitsett, J.A. (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. J Biol Chem *280*, 13809-13816.

17. Gao, N., LeLay, J., Vatamaniuk, M.Z., Rieck, S., Friedman, J.R., and Kaestner, K.H. (2008). Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. Genes Dev *22*, 3435-3448.

18. Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. Nature *435*, 944-947.

19. Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., and Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev *10*, 1670-1682.

20. Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A.V., and Zaret, K.S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev *23*, 804-809.

21. Kiefer, J.C. (2007). Back to basics: Sox genes. Dev Dyn *236*, 2356-2366.

22. Rizzoti, K., and Lovell-Badge, R. (2007). SOX3 activity during pharyngeal segmentation is required for craniofacial morphogenesis. Development *134*, 3437-3448.

23. Que, J., Okubo, T., Goldenring, J.R., Nam, K.T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development *134*, 2521-2531.

24. Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

25. Beck, F. (2004). The role of Cdx genes in the mammalian gut. Gut *53*, 1394-1396.

26. Subramanian, V., Meyer, B.I., and Gruss, P. (1995). Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. Cell *83*, 641-653.

27. Chawengsaksophak, K., James, R., Hammond, V.E., Kontgen, F., and Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. Nature *386*, 84-87.

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

29. 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.

30. Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., and Wright, C.V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development *122*, 983-995.

31. Mathan, M., Moxey, P.C., and Trier, J.S. (1976). Morphogenesis of fetal rat duodenal villi. Am J Anat *146*, 73-92.

32. 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 *127*, 3457-3466.

33. Braunstein, E.M., Qiao, X.T., Madison, B., Pinson, K., Dunbar, L., and Gumucio, D.L. (2002). Villin: A marker for development of the epithelial pyloric border. Dev Dyn *224*, 90-102.

34. Daniel, E.E. (1992). Sphincters : normal function--changes in diseases, (Boca Raton, Fla.: CRC Press).

35. Panteli, C. (2009). New insights into the pathogenesis of infantile pyloric stenosis. Pediatr Surg Int *25*, 1043-1052.

36. Huang, P.L., Dawson, T.M., Bredt, D.S., Snyder, S.H., and Fishman, M.C. (1993). Targeted disruption of the neuronal nitric oxide synthase gene. Cell *75*, 1273-

1286.

37. Barbosa, I.M., Ferrante, S.M., and Mandarim-De-Lacerda, C.A. (2001). [Role of nitric oxide synthase in the etiopathogenesis of hypertrophic pyloric stenosis in infants]. J Pediatr (Rio J) *77*, 307-312.

38. Smith, D.M., Grasty, R.C., Theodosiou, N.A., Tabin, C.J., and Nascone-Yoder, N.M. (2000). Evolutionary relationships between the amphibian, avian, and mammalian stomachs. Evol Dev *2*, 348-359.

39. Self, M., Geng, X., and Oliver, G. (2009). Six2 activity is required for the formation of the mammalian pyloric sphincter. Dev Biol *334*, 409-417.

40. Nielsen, C., Murtaugh, L.C., Chyung, J.C., Lassar, A., and Roberts, D.J. (2001). Gizzard formation and the role of Bapx1. Dev Biol *231*, 164-174.

41. Moniot, B., Biau, S., Faure, S., Nielsen, C.M., Berta, P., Roberts, D.J., and de Santa Barbara, P. (2004). SOX9 specifies the pyloric sphincter epithelium through mesenchymal-epithelial signals. Development *131*, 3795-3804.

42. Roberts, D.J., Smith, D.M., Goff, D.J., and Tabin, C.J. (1998). Epithelial-mesenchymal signaling during the regionalization of the chick gut. Development *125*, 2791-2801.

43. Theodosiou, N.A., and Tabin, C.J. (2005). Sox9 and Nkx2.5 determine the pyloric sphincter epithelium under the control of BMP signaling. Dev Biol *279*, 481-490.

44. Smith, D.M., Nielsen, C., Tabin, C.J., and Roberts, D.J. (2000). Roles of BMP signaling and Nkx2.5 in patterning at the chick midgut-foregut boundary. Development *127*, 3671-3681.

45. Burn, S.F., Boot, M.J., de Angelis, C., Doohan, R., Arques, C.G., Torres, M., and Hill, R.E. (2008). The dynamics of spleen morphogenesis. Dev Biol *318*, 303-311.

46. Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev *9*, 1654-1666.

47. Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrugghe, B. (2001). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci U S A *98*, 6698-6703.

48. Khokha, M.K., Hsu, D., Brunet, L.J., Dionne, M.S., and Harland, R.M. (2003). Gremlin is the BMP antagonist required for maintenance of Shh and Fgf signals during limb patterning. Nat Genet *34*, 303-307.

49. Kim, B.M., Miletich, I., Mao, J., McMahon, A.P., Sharpe, P.A., and Shivdasani, R.A. (2007). Independent functions and mechanisms for homeobox gene Barx1 in patterning mouse stomach and spleen. Development *134*, 3603-3613.

50. Verzi, M.P., Stanfel, M.N., Moses, K.A., Kim, B.M., Zhang, Y., Schwartz, R.J., Shivdasani, R.A., and Zimmer, W.E. (2009). Role of the homeodomain transcription factor Bapx1 in mouse distal stomach development. Gastroenterology *136*, 1701-1710.

51. Zhang, J., Rosenthal, A., de Sauvage, F.J., and Shivdasani, R.A. (2001). Downregulation of Hedgehog signaling is required for organogenesis of the small intestine in Xenopus. Dev Biol *229*, 188-202.

52. Kolterud, A., Grosse, A.S., Zacharias, W.J., Walton, K.D., Kretovich, K.E., Madison, B.B., Waghray, M., Ferris, J.E., Hu, C., Merchant, J.L., et al. (2009). Paracrine Hedgehog signaling in stomach and intestine: new roles for hedgehog in gastrointestinal patterning. Gastroenterology *137*, 618-628.

53. Lees, C.W., Zacharias, W.J., Tremelling, M., Noble, C.L., Nimmo, E.R., Tenesa, A., Cornelius, J., Torkvist, L., Kao, J., Farrington, S., et al. (2008). Analysis of germline GLI1 variation implicates hedgehog signalling in the regulation of intestinal inflammatory pathways. PLoS Med *5*, e239.

54. Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795-801.

55. Jiang, J., and Hui, C.C. (2008). Hedgehog signaling in development and cancer. Dev Cell *15*, 801-812.

56. Madison, B.B., Braunstein, K., Kuizon, E., Portman, K., Qiao, X.T., and Gumucio, D.L. (2005). Epithelial hedgehog signals pattern the intestinal crypt-villus axis. Development *132*, 279-289.

57. Kim, J.H., Huang, Z., and Mo, R. (2005). Gli3 null mice display glandular overgrowth of the developing stomach. Dev Dyn *234*, 984-991.

58. Sukegawa, A., Narita, T., Kameda, T., Saitoh, K., Nohno, T., Iba, H., Yasugi, S., and Fukuda, K. (2000). The concentric structure of the developing gut is regulated by Sonic hedgehog derived from endodermal epithelium. Development *127*, 1971-1980.

59. Zacharias, W.J., Li, X., Madison, B.B., Kretovich, K., Kao, J.Y., Merchant, J.L., and Gumucio, D.L. Hedgehog is an anti-inflammatory epithelial signal for the intestinal lamina propria. Gastroenterology.

60. Heemskerk, J., and DiNardo, S. (1994). Drosophila hedgehog acts as a morphogen in cellular patterning. Cell *76*, 449-460.

61. Strigini, M., and Cohen, S.M. (1997). A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. Development *124*, 4697-4705.

62. Hynes, M., Ye, W., Wang, K., Stone, D., Murone, M., Sauvage, F., and Rosenthal, A. (2000). The seven-transmembrane receptor smoothened cell-autonomously induces multiple ventral cell types. Nat Neurosci *3*, 41-46.

63. Sanson, B. (2001). Generating patterns from fields of cells. Examples from Drosophila segmentation. EMBO Rep *2*, 1083-1088.

64. Alexandre, C., Jacinto, A., and Ingham, P.W. (1996). Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. Genes Dev *10*, 2003-2013.

65. Von Ohlen, T., and Hooper, J.E. (1997). Hedgehog signaling regulates transcription through Gli/Ci binding sites in the wingless enhancer. Mech Dev *68*, 149-156.

66. Muller, B., and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. Development *127*, 2999-3007.

67. Kwon, C., Hays, R., Fetting, J., and Orenic, T.V. (2004). Opposing inputs by Hedgehog and Brinker define a stripe of hairy expression in the Drosophila leg imaginal disc. Development *131*, 2681-2692.

68. Goodrich, L.V., Johnson, R.L., Milenkovic, L., McMahon, J.A., and Scott, M.P. (1996). Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog. Genes Dev *10*, 301-312.

69. Forbes, A.J., Nakano, Y., Taylor, A.M., and Ingham, P.W. (1993). Genetic analysis of hedgehog signalling in the Drosophila embryo. Dev Suppl, 115-124.

70. Chen, Y., and Struhl, G. (1996). Dual roles for patched in sequestering and transducing Hedgehog. Cell *87*, 553-563.

71. Hays, R., Buchanan, K.T., Neff, C., and Orenic, T.V. (1999). Patterning of Drosophila leg sensory organs through combinatorial signaling by hedgehog, decapentaplegic and wingless. Development *126*, 2891-2899.

72. Von Ohlen, T., Lessing, D., Nusse, R., and Hooper, J.E. (1997). Hedgehog signaling regulates transcription through cubitus interruptus, a sequence-specific

DNA binding protein. Proc Natl Acad Sci U S A *94*, 2404-2409.

73. Nellen, D., Burke, R., Struhl, G., and Basler, K. (1996). Direct and long-range action of a DPP morphogen gradient. Cell *85*, 357-368.

74. Posakony, L.G., Raftery, L.A., and Gelbart, W.M. (1990). Wing formation in Drosophila melanogaster requires decapentaplegic gene function along the anterior-posterior compartment boundary. Mech Dev *33*, 69-82.

75. Sanicola, M., Sekelsky, J., Elson, S., and Gelbart, W.M. (1995). Drawing a stripe in Drosophila imaginal disks: negative regulation of decapentaplegic and patched expression by engrailed. Genetics *139*, 745-756.

76. Berman, B.P., Nibu, Y., Pfeiffer, B.D., Tomancak, P., Celniker, S.E., Levine, M., Rubin, G.M., and Eisen, M.B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc Natl Acad Sci U S A *99*, 757-762.

77. Rebeiz, M., Reeves, N.L., and Posakony, J.W. (2002). SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. Proc Natl Acad Sci U S A *99*, 9888-9893.

78. Markstein, M., Markstein, P., Markstein, V., and Levine, M.S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. Proc Natl Acad Sci U S A *99*, 763-768.

79. Arnone, M.I., and Davidson, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development *124*, 1851-1864.

80. Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., and Gehring, W.J. (1989). P-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev *3*, 1288-1300.

81. Gindhart, J.G., Jr., King, A.N., and Kaufman, T.C. (1995). Characterization of the cis-regulatory region of the Drosophila homeotic gene Sex combs reduced. Genetics *139*, 781-795.

82. Johnson, K.D., and Bresnick, E.H. (2002). Dissecting long-range transcriptional mechanisms by chromatin immunoprecipitation. Methods *26*, 27-36.

83. Massie, C.E., and Mills, I.G. (2008). ChIPping away at gene regulation. EMBO Rep *9*, 337-343.

84. Vokes, S.A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh,

W.J., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

85. Vokes, S.A., Ji, H., Wong, W.H., and McMahon, A.P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes Dev *22*, 2651-2663.

86. Cornish-Bowden, A. (1985). Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res *13*, 3021-3030.

87. Staden, R. (1984). Computer methods to locate signals in nucleic acid sequences. Nucleic Acids Res *12*, 505-519.

88. Hallikas, O., and Taipale, J. (2006). High-throughput assay for determining specificity and affinity of protein-DNA binding interactions. Nat Protoc *1*, 215-222.

89. Stormo, G.D., and Fields, D.S. (1998). Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci *23*, 109-113.

90. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res *23*, 4878-4884.

91. Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell *124*, 47-59.

CHAPTER II

DYNAMIC PATTERNING AT THE PYLORUS: FORMATION OF AN EPITHELIAL INTESTINE-STOMACH BOUNDARY IN LATE FETAL LIFE

In the adult mouse, distinct morphological and transcriptional differences separate stomach from intestinal epithelium. Remarkably, the epithelial boundary between these two organs is literally one cell thick. This discrete junction is established suddenly and precisely at E16.5, by sharpening a previously diffuse intermediate zone. In the present study, we define the dynamic transcriptome of stomach, pylorus and intestinal tissues between E14.5 and E16.5. We show that establishment of this boundary is concomitant with the induction of over a thousand genes in intestinal epithelium, and these gene products provide intestinal character. Hence, we call this process *intestinalization*. We identify specific transcription factors [i.e., hepatocyte nuclear factor 4, gamma (Hnf4γ), cAMP responsive element binding protein 3-like 3 (Creb3l3), and transcription factor EC (Tcfec)] and examine signaling pathways [i.e., Hh and Wnt] that may play a role in this process. Finally, we define a unique expression domain at the pylorus itself and detect novel pylorus-specific patterns for the transcription factor Gata3 and the secreted protein nephrocan (Nepn).

II.1. Introduction

The vertebrate GI tract consists of a series of connected organs (i.e., esophagus, stomach, small intestine, large intestine), each with a highly specialized epithelial surface that enables it to perform a distinct function during digestion. In adults, the epithelial boundaries between some of these adjacent organs are remarkably sharp. At the pylorus, for example, gastric and intestinal cells lie directly next to one another, without a transitional cell type (i.e., the epithelial pyloric border) [1].

Such discrete organ boundaries have fetal origins. In the embryo, the gut tube is molded from endoderm, along with its associated splanchnic mesoderm [2]. Anteroposterior patterning of the GI tract begins even before tube formation is complete; by E10 the developing gut tube has a clear Hox code that marks out the major organ domains and future sphincter locations [3]. Expression patterns of other gut transcription factors are also established early, including: the HMG-box protein Sox2 in early endoderm of the stomach domain [4], the caudal-related parahox factor Cdx2 in presumptive intestinal endoderm [5], and the homeodomain protein Nkx2-5 in a thin band of mesenchymal cells at the site of the future pylorus [6]. Despite this pattern, the epithelial surface exhibits few obvious morphological differences from stomach to intestine, even as late as E14.5.

While examining the expression pattern of villin, an intestine-specific actin bundling protein, we previously found that, at E14.5, it is expressed in a diminishing gradient (posterior to anterior) at the pyloric border. Two days later however, at E16.5, a sharp anterior expression boundary resolves: villin is detected at high levels in intestinal cells,

while neighboring gastric cells exhibit little or no expression [1]. We speculated that the formation of this boundary may reflect an important epithelial compartmentalization event in the GI tract. If so, it is remarkable for its late timing, more than five days after the initial establishment of the broad territorial domains that specify the location of stomach and intestine.

In the present study, we sought to determine whether compartmentalization of *villin* expression is accompanied by the formation of similar dramatic expression boundaries for other genes. Microdissection and microarray analysis was used to examine gene expression patterns at and around the pylorus at E14.5 and E16.5. Our data reveal that at E14.5, the transcriptomes of stomach, pylorus and intestine are only subtly different. At E16.5, however, hundreds of genes are coordinately upregulated in intestine. Remarkably, this transcriptional burst is seen in the intestinal epithelium but not the mesenchyme, and the batteries of activated genes are involved in the prototypical intestinal functions of absorption and metabolism. Interestingly, a similar large scale burst of gene induction does not occur in the stomach; the transcriptome of this organ changes little between E14.5 and E16.5. We identify several upregulated transcription factors (i.e., Hnf4γ, Tcfec, and Creb3l3), which, similar to *villin*, exhibit dramatic pyloric expression boundaries at E16.5. We also investigate signaling pathways (i.e., Hh and Wnt) that may be modulated during this compartmentalization event. Finally, we uncover novel genes with expression patterns that are restricted to the pyloric region itself and could participate in this patterning event. These include genes that encode the zinc finger

transcription factor Gata3 and the secreted transforming growth factor, beta (TGF-β) modulator nephrocan.

II.2. Materials and methods

Tissue collection

Embryos from C57BL/6J mice were collected from timed pregnant females, with the day of vaginal plug detection considered day 0.5. Intestine and stomach were removed and three contiguous segments were collected from antrum (stomach), pylorus, and duodenum (Figure II.1). For the microarray experiment, a total of 174 E14.5 embryos and 95 E16.5 embryos were dissected, and for each time/tissue group (e.g., E14.5 duodenum; six groups total), collected tissue was randomly pooled into one of three samples for replicate analysis. Total RNA was isolated with TRIzol and purified using the RNeasy kit, per the manufacturers' instructions. Separate collections were performed for RT-PCR and qPCR validation of microarray results.

Microarray processing

RNA samples were hybridized to MOE 430.2 microarrays (eighteen total chips: six time/tissue groups, three replicate chips for each group) by the University of Michigan Cancer Center Microarray Core Facility. Microarrays were scanned and processed using GCOS software (Affymetrix) and the resulting .CEL files were analyzed using RMA (Robust Multiarray Average; *affy* package in BioConductor; www.bioconductor.org), which does background adjustment, normalization and conversion of probeset intensity

data to log2 expression values [7, 8]. Probeset expression values were imported into MeV (Multi-experiment Viewer; www.tm4.org) for evaluation of statistical significance using the Student's t-test [9]. Replicate probeset expression values were averaged and log fold change (LFC) was determined by calculating the difference between any two groups [e.g., LFC = $\log2(\text{E14.5 duodenum})_{avg} - \log2(\text{E14.5 stomach})_{avg}$], which was then converted to numerical fold change (FC) [if LFC > 0, $FC = 2^{(LFC)}$; else, $FC = -(2^{(-LFC)})$]. Probesets with a p-value less than 0.05 and |FC| greater than 2 were selected for further analysis. To provide independent validation for the array results, we evaluated gene expression by RT-PCR. These results, shown in Figure II.2, reveal concordance with the array findings. Functional Annotation Clustering in DAVID, using the default options and "Medium" Classification Stringency, was performed with the following sets of Gene Ontology terms: GOTERM_BP_ALL, GOTERM_MF_ALL, PANTHER_BP_ALL and PANTHER_MF_ALL.

In situ hybridization

Staged E14.5 and E16.5 C57BL/6J embryos were dissected in DEPC-treated 1X PBS. Isolated gastrointestinal tracts were fixed overnight at 4°C in 4% PFA, embedded in paraffin, and cut into 5 μm sections. In situ hybridization was performed as described previously [10]. Specific probes used for in situ hybridization included: *Gata3* (NM_008091.3; 1028-1591 bp), *nephrocan* (NM_025684.2; 758-1450 bp), *Sfrp5* (NM_018780.2; 203-985 bp), *Creb3l3* (NM_145365.3; 683-1361 bp), *Tcfec* (NM_031198.2; 258-789), *Grem1* (NM_011824.4; 487-1281), and *Axin2* (NM_015732.4; 2227-3358).

Antibodies and immunostaining

Staged E14.5 and E16.5 C57BL/6J embryos were dissected in 1X PBS. Excised gastrointestinal tracts were fixed for overnight at 4°C in 4% PFA, embedded in paraffin, and cut into 5 μm sections. Vectastain ABC was used for Hnf4γ immunohistochemistry, per the manufacturer's instructions. Sox2 immunofluorescence was performed as described previously [11]. Briefly, sections were deparaffinized in xylene, rehydrated through increasing alcohol concentrations, and boiled for 10 minutes either in Antigen Unmasking Solution (Vector Laboratories, CA) (Sox2) or 10 mM sodium citrate, pH 6.0 (Hnf4γ). Slides were allowed to slowly cool down and then washed with 1X PBS. The sections were blocked with 10% donkey serum/0.01% Triton X-100 in 1X PBS for 1 hour at room temperature. Antibodies used were: rabbit polyclonal anti-Sox2 antibody (1:500, Chemicon, MA); goat polyclonal anti-Hnf4γ (1:500, Santa Cruz Biotechnology, CA); biotinylated horse anti-goat IgG (1:200, Vector); and, Alexa Fluor 488 donkey anti-rabbit IgG (1:1000, Invitrogen, OR). Primary antibodies were diluted in blocking solution and incubated on slides overnight at 4°C. Slides were washed in 1X PBS prior to incubation with appropriate secondary antibodies and DAPI (1:100) (Sox2) for 30 to 60 minutes at room temperature. After 1X PBS wash, coverslips were mounted with ProLong Gold Antifade Reagent (Sox2), or sections were lightly counterstained with hematoxylin, dehydrated with serial alcohol/xylene washes and coverslips were mounted with Permount. Imaging was done on a Nikon E800 microscope digital imaging system.

X-gal staining

Staged E14.5 or E16.5 embryos from genetic crosses of villin$^{lacZ/+}$ or Gli1$^{lacZ/+}$ males with C57BL/6J females were dissected on ice in 1X PBS [1, 12]. The gastrointestinal tract was excised, fixed with 4% PFA for 10 minutes at 4°C, and washed three times in 1X PBS. Tissues were saturated overnight with 30% sucrose (in 1X PBS), embedded and frozen in Optimal Cutting Temperature (OCT) compound, and cut into 5 μm sections. X-gal staining solution was prepared fresh, as follows: 1X PBS (pH 7.5), 2 mM magnesium chloride, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, and 1 mg X-gal (from 20 mg/ml stock in DMF). Sectioned tissue was incubated in staining solution at 37°C and monitored periodically to control staining intensity. Sections were then washed three times in 1X PBS, counterstained lightly with eosin and dehydrated with serial alcohol/xylene washes, and coverslips were mounted with Permount. Slides were imaged on an E800 microscope imaging system.

Real-time quantitative RT-PCR (qPCR)

Staged E14.5 and E16.5 C57BL/6J embryos were dissected in 1X PBS. Samples from the duodenum and antrum of individual embryos were collected and RNA was harvested by brief vortexing in MELT (Multi-Enzymatic Liquefaction of Tissue) digestion buffer (Applied Biosystems/Ambion, TX) and purified using the RNeasy kit. A total of eight E14.5 and nine E16.5 samples were obtained (for each tissue). cDNA was prepared with the iScript cDNA Synthesis Kit. qPCR reactions were set up using 2X iQ Supermix buffer and iQ iCycler 96-well plates and carried out in an iCycler real-time PCR detection system. Data were processed using the iCycler program and imported into Excel for normalization and statistical analysis. Gli1 Ct values were normalized to Hprt

(ΔCt) and then compared to the average ΔCt for E14.5 intestine (ΔΔCt). Average ΔΔCt values were calculated to determine relative fold difference to E14.5 tissue [2^(- ΔΔCt)]. 95% confidence intervals were determined using Excel (CONFIDENCE function). PCR conditions were optimized previously in the lab (i.e.- temperature, DMSO, etc; see Table II.1).

Reverse transcriptase PCR (RT-PCR)

Staged E14.5 and E16.5 C57BL/6J embryos were dissected in 1X PBS. Samples from four independent collections were separated by tissue type and time point (e.g. - E14.5 border, E16.5 duodenum, etc.) and then pooled randomly into two groups for replicate analysis. Additionally, fragments of contiguous tissue spanning the antrum, border and proximal duodenum were collected from E14.5 and E16.5 embryos as an input control. Tissue from each group was homogenized in 1 mL of TRIzol by drawing it through a 30-gauge needle. Total RNA was prepared according to the manufacturer's instructions, purified by consecutive phenol-chloroform (2X) and chloroform (2X) extractions, and quantitated by UV spectrophotometry using a NanoDrop. For each RNA sample, two independent cDNA preparations were performed using the iScript cDNA Synthesis Kit and pooled for subsequent analysis. Negative "No RT" controls for genomic contamination were prepared from whole E14.5 and E16.5 RNA in a similar manner. PCR was performed on cDNA samples using qRT-PCR-quality primers created by Beacon Designer. Individual primer conditions were optimized (i.e.- temperature, number of cycles, magnesium chloride, DMSO, etc; see Table II.1) prior to PCR. Products from PCR reactions were resolved under UV light with ethidium bromide-loaded, 2% TBE-

agarose gels. The band intensity of experimental genes was compared to the housekeeping gene Hprt.

II.3. Results

Diffuse pyloric expression boundaries at E14.5

Expression of Cdx2 (an intestinal marker) and Sox2 (a stomach marker) in early foregut endoderm was previously examined using whole mount confocal immunofluorescence [4]. By E9.5, the staining domains of these two proteins at the pylorus appear to be essentially distinct, prompting the authors of the study to propose that Cdx2 repression of *Sox2* establishes this boundary. Indeed, loss of *Cdx2* in the intestinal domain leads to increased Sox2 expression and conversion of the epithelium to an esophageal morphology [13]. If Cdx2 suppression of Sox2 is responsible for formation of the pyloric border and if the expression of these two proteins is compartmentalized across that border by E9.5, as suggested by Sherwood et al., then the compartmentalization of *villin* expression, which occurs seven days later at E16.5 (Braunstein et al, 2002), likely requires additional cues.

To investigate this more carefully, we examined pyloric *Cdx2* and Sox2 expression at E14.5 in sectioned material, and compared these patterns to that of *villin*. As shown in Figures II.3A-C, the boundaries of the expression domains of *Cdx2* and Sox2 are, like that of *villin*, diffuse at the pyloric region. Thus, it appears that while a regional pattern of endoderm identity (*Sox2* in stomach and *Cdx2* in intestine) is established by E14.5, the

precise boundaries of this pattern at the pylorus are not yet mature at the cellular level. The discrepancy between these results and those of Sherwood et al. may be explained by the use of sectioned material and the much higher magnification used here.

Gastric and duodenal transcriptomes are similar at E14.5 but distinct at E16.5

To learn more about the process of epithelial pyloric border formation, we analyzed global gene expression patterns at and around the developing pylorus. Using C57BL/6J mice, small (1-2mm) tissue fragments were microdissected from the pyloric region itself (easily recognized grossly by the muscular constriction) and from adjacent stomach and duodenum both before (E14.5) and after (E16.5) border formation (Figure II.1). Triplicate samples of extracted RNA were processed for microarray.

To assess the similarities and differences among the various tissue samples, and to determine the reproducibility of the replicate microarray chips, Principal Component Analysis (PCA) was applied, as described in the legend to Figure II.4. The clustering of replicate chip samples seen in Figure II.4A demonstrates that the collection process was reproducible. Importantly, the three E14.5 tissue samples (stomach, pylorus and duodenum) are tightly grouped, indicating that the transcriptomes of these tissues are quite similar to one another. In contrast, the three E16.5 groups are clearly different from one another and different from the E14.5 groups. Of the three E16.5 groups, the duodenum shows the most change from E14.5 to E16.5, as measured by the degree of separation between the points along the axis of Principal Component 1 (PC1). PC1, by

definition, contains the majority of the variation in the data [14]. In contrast, the stomach samples exhibit much less change along the PC1 coordinate.

A dramatic change in the duodenal transcriptome at E16.5

To identify gene expression changes underlying the emergence of distinct gastric and duodenal transcriptomes at E16.5, we assembled a list of all significant pairwise probeset enrichments, including comparisons of: a) each tissue at E14.5 to the same tissue at E16.5 (time axis; three comparisons); and, b) stomach, pylorus and duodenum to one another at either E14.5 or E16.5 (tissue axis; six comparisons). A total of 10,499 unique differentially expressed probesets were identified for all comparisons. (Due to space limitations, we do not present these data here; please see [15]). Of these, 9,137 showed significant changes in the time comparisons, 5,909 in the tissue comparisons, and 4,547 in both. We selected the group of probesets that are dynamic in both time and tissue dimensions for further analysis.

Although some genes were unique to the pyloric region (addressed below), the results of the PCA analysis (Figure II.4A) shows that the pyloric area generally (with some exceptions) exhibits probeset intensity values that are essentially the average of stomach and intestine values. Thus, for initial analysis, we compared temporally dynamic probesets in stomach and intestine to catch patterning differences that were emerging between these two tissues at E16.5. Of the 4,547 time and tissue dynamic probesets, 86% (3,917) showed enrichment in either stomach or duodenum at E16.5. Plotting these results in a Venn diagram (Figure II.4B) reveals that from E14.5 to E16.5, probesets

exhibiting transcriptional change in the duodenum far outnumber the probesets changing in stomach. In the stomach, 1,127 probesets show temporal change, but only 40% of these (452) are specific to the stomach. (675 probesets were changed in both stomach and duodenum.) In contrast, the robust temporal transcriptional change in the duodenum encompasses differences in 3,420 probesets, 80% of which (2,745) change only in the duodenum. These temporally dynamic duodenal probesets could be further annotated as enriched at E16.5 (i.e., depleted at E14.5) or depleted at E16.5 (i.e., enriched at E14.5). Tallying these groups, labeled D16 enriched or D16 depleted, respectively, reveals that D16 enriched probesets accounts for 61% (1,664 probesets) of the duodenum-specific temporal change. (Due to space limitations, we do not present these data here; please see [15]).

Upregulated genes in E16.5 duodenum are primarily epithelial, while downregulated genes are mesenchymal

We were next interested to determine whether the dramatic transcriptional change seen in duodenum at E16.5 was primarily due to changes in genes that are expressed in epithelium or mesenchyme. In an earlier study, we separated intestinal tissue using non-enzymatic methods and profiled gene expression in freshly isolated epithelium and mesenchyme to create a catalog of epithelial and mesenchymal genes [10]. Though the earlier study was done using E18.5 intestine, we reasoned that the epithelial/mesenchymal compartmentalization of genes would be largely similar at E16.5 and E18.5. Thus, using the earlier data, we tagged all D16 enriched and D16 depleted probesets as epithelial or mesenchymal. Some probesets could not be classified this way due to low expression or expression in both compartments. For classified D16 enriched

probesets, the vast majority (98.3%) were epithelial. These striking results predict that the formation of the distinct epithelial boundary between stomach and intestine is concomitant with a massive transcriptional inductive event in duodenal epithelium. Interestingly, and in contrast, 99.6% of compartment classified probesets that were depleted at E16.5 were mesenchymal (Figure II.4C).

Dramatic tissue rearrangement accompanies the emergence of intestinal villi between E14.5 and E16.5. Thus, we were concerned that the transcriptional changes observed might merely reflect an alteration in the ratio of epithelium to mesenchyme. We therefore examined the distribution of fold changes seen among D16 epithelial upregulated probesets or among D16 depleted mesenchymal probesets (Figure II.4D). This analysis revealed that 42% of the probesets in the D16 epithelial group are upregulated at levels greater than five-fold; in fact, 20% were upregulated over ten-fold. (Maximum upregulation of over 100-fold was seen in the case of four probesets.) However, only 8% of the D16 mesenchymal probesets are downregulated by five-fold or more and none are downregulated over ten-fold. In fact, nearly 75% of all D16 mesenchymal probesets show less than a four-fold change. Thus, we conclude that tissue rearrangement may account for some of the low level transcriptional change seen among the downregulated (and primarily mesenchymal) probesets. However, the dramatic inductive change observed in the epithelial compartment cannot be accounted for by this mechanism.

Duodenal gene expression changes at E16.5 correlate with functional differentiation of the small intestine

We used functional annotation clustering methods (DAVID; david.abcc.ncifcrf.gov) to classify D16 probesets [16]. Since there were only four probesets downregulated in epithelium [i.e., *Foxa1*, *glucagon* (*Gcg*), *melanoregulin* (*Mreg*), and *serine (or cysteine) peptidase inhibitor, clade A, member 1C* (*Serpina1c*)] we could not use DAVID analysis for this group. However, for the eighteen probesets upregulated in mesenchyme, the most enriched annotation terms were apoptosis, immune response and guanosine triphosphate (GTP) binding. Conversely, among the 73 mesenchymal probesets that are downregulated more than five-fold, we found that cell differentiation, neuronal development and cell migration are the most frequent terms represented (data not shown). Importantly, for the group of 431 epithelial upregulated probesets, genes involved in metabolic processes and cell transport are statistically overrepresented (Table II.2). This finding suggests that, between E14.5 and E16.5, there is a compartmentalized, functional switch in duodenal gene expression. The mesenchyme downregulates genes involved in development and morphogenesis, while the large burst of gene expression in the epithelium appears to prepare the intestine for its major role in absorption and metabolism. Thus, we can think of this significant epithelial change as a process of *intestinalization*.

Upregulated duodenal transcription factors display a sharp epithelial boundary at the pylorus

It is of interest to identify which transcription factors show significant change during of the epithelium. Table II.3 lists the most enriched transcription factors in the D16 epithelial group. Among these genes, *Tcfec*, *Creb3l3*, *nuclear receptor subfamily 2, group E, member 3* (*Nr2e3*), *v-maf musculoaponeurotic fibrosarcoma oncogene family,*

protein B (avian) (*Mafb*) and *Hnf4γ* exhibit expression levels that are over ten-fold enriched in duodenum relative to stomach. Hnf4γ is the lesser studied paralog of Hnf4α, a protein critical for epithelial function in pancreas, liver and colon [17-19]. *Hnf4α* is also highly expressed in intestinal epithelium, but is only two- to three-fold enriched in intestine relative to stomach at both E14.5 and E16.5 (data not shown). The basic leucine zipper (bZip) factor Mafb is mutated in kreisler mice, which exhibit morphogenic defects in hindbrain and inner ear [20, 21]. Interestingly, a recent genome-wide screen links *Mafb* to polygenic dyslipidemia [22]. *Nr2e3* (also known as *PNR* or *photoreceptor-specific nuclear receptor*) was initially implicated in eye development [23], but a recent study demonstrated that it is highly induced in intestinal epithelium during midgestation [24]. Neither *Tcfec* nor *Creb3l3* has previously been associated with intestinal gene regulation.

We looked carefully at the expression of three of these highly enriched D16 transcription factors to determine whether, similar to *villin*, they display a sharp, cell-specific anterior boundary at E16.5. Immunohistochemical staining for Hnf4γ shows both nuclear and cytoplasmic staining in the epithelium. At E16.5, cytoplasmic staining extends into the stomach, but strong nuclear staining is sharply demarcated at the pylorus (Figure II.5A). *In situ* hybridization reveals a sharp anterior expression boundary for *Creb3l3* (Figure II.5C) and *Tcfec* (Figure II.5E) at E16.5. Additionally, for both *Creb3l3* (Figure II.5D) and *Tcfec* (Figure II.5F), expression is primarily localized to the differentiated cells of the villus tips. Hnf4γ nuclear staining also appears to display a diminishing villus-crypt expression gradient (Figure II.5B). Thus, activation of these three transcription factors

(and likely others in Table II.3) in E16.5 duodenal epithelium appears to be concomitant with the maturation of the crypt/villus axis.

Modulation of developmental signaling pathways accompanies formation of the epithelial pyloric border

Soluble signaling factors play major roles in gut patterning, and previous data suggest that, in particular, Hh [25] and Wnt [26-28] signaling are critical in the context of intestinal differentiation. Thus, we examined the array results for expression of key elements of these pathways. We also directly tested, using transcriptional reporter readouts, whether there is differential activation of either of these two pathways across the pylorus at E14.5 or E16.5. The results are summarized below and presented in Table II.4 (Hh) and Table II.5 (Wnt).

Hh pathway

At E14.5, the Hedgehog ligands *Shh* and *Ihh* were expressed similarly in stomach and duodenum. However, by E16.5, *Shh*, which is expressed in the epithelium and signals in a paracrine manner to the mesenchyme [29, 30], was specifically downregulated in the duodenum (6.7 fold relative to stomach). In accordance with the drop in *Shh*, all three Gli factors (i.e., *Gli1*, *Gli2*, and *Gli3*) were reduced in the duodenum, as were the co-receptors *Ptch1* and *Smo* and the mouse Fused homolog *serine/threonine kinase 36 (fused homolog, Drosophila)* (*Stk36*). A number of recently identified pathway modulators [i.e., *growth arrest specific 1* (*Gas1*), *biregional cell adhesion molecule-related/down-regulated by oncogenes (Cdon) binding protein* (*Boc*), *cell adhesion molecule-related/down-regulated by oncogenes* (*Cdon*), and *RAB23, member RAS oncogene family*

(*Rab23*)] were also reduced in duodenum [31-33]. Only *megalin* (*Lrp2*), an endocytic receptor for Hedgehog ligands [34], is upregulated in epithelium of E16.5 duodenum. However, megalin is not dedicated to the transport of Hedgehog ligands and is known to transport cholesterol and vitamins among the over 50 molecules that it can bind [35].

We examined Hedgehog signaling across the pylorus at E14.5 and E16.5 using $Gli1^{lacZ/+}$ reporter mice as a direct readout of pathway activity [12]. Figure II.6A shows that, at E14.5, $Gli1^{lacZ/+}$ staining is similar in stomach and intestine. However, at E16.5, Gli1 activity is apparently reduced on the duodenal side of the pylorus, but maintained in the stomach (Figure II.6B), in agreement with the array results and with another recent study [30]. To validate this observation in a more quantitative manner, we assayed *Gli1* mRNA expression using real-time quantitative RT-PCR (qPCR). When normalized to the housekeeping gene *Hprt* and compared to levels in E14.5 intestine, *Gli1* expression was significantly decreased in E16.5 intestine (95% confidence interval: 0.8 to 0.3 fold change; Student's t-test: $p<0.05$; Figure II.6D). In contrast, *Gli1* expression in stomach did not vary significantly between E14.5 and E16.5 (Figure II.6C).

Wnt pathway

The expression of several Wnt ligands is modulated during pyloric border formation. At E16.5, *Wnt4*, *Wnt5a* and *Wnt11* show significant enrichment in stomach relative to duodenum. The Wnt receptors *frizzled homolog 1 (Drosophila)* (*Fzd1*), *Fzd2*, and *Fzd6* also show significant enrichment in stomach at this time, as do a number of Wnt pathway modulators (i.e., *secreted frizzled-related protein 1* (*Sfrp1*), *Sfrp2*, *Sfrp4*, *dickkopf*

homolog 2 (Xenopus laevis) (*Dkk2*), and *Dkk3*). However, the most robust tissue-specific expression pattern of all Wnt pathway components is that of the secreted modulator *Sfrp5*, which is highly duodenum-specific at both E14.5 (32 fold relative to stomach) and E16.5 (10.7 fold relative to stomach). Taken together, these findings predict that Wnt signaling activity, like that of Hh, is decreased in intestine at E16.5. To investigate this further, we explored some of these expression patterns by *in situ* hybridization.

We first examined *Sfrp5* expression. By E14.5, *Sfrp5* is robustly expressed in duodenal epithelium, with a soft anterior boundary of expression that extends a short distance into the stomach (Figure II.7A). By E16.5, when villus formation has begun, this expression domain resolves dramatically; *Sfrp5* expression is excluded from the villus tips and present only in cells of the proliferative intervillus region (Figure II.7B). The localized expression of this Wnt modulator (thought to be a Wnt inhibitor) in intervillus cells is consistent with the recent study of Kim et al., which concludes that, between E16.5 and birth, canonical Wnt signaling is excluded from intervillus regions and restricted to villus epithelium [26].

To clearly define the domain of active canonical Wnt signaling and establish whether a gradient of Wnt signaling indeed exists across the pylorus at either E14.5 or E16.5, we examined the expression of *Axin2*, a Wnt target gene and a commonly accepted read-out of canonical signaling activity [36]. Previous studies using an *Axin2* β-galactosidase (β-gal) reporter mouse suggested that *Axin2* is expressed mainly on villus tip cells and excluded from intervillus cells [26]. Because the detection of β-gal reporter activity in

intestinal epithelium can be complicated by the presence of an endogenous β-gal activity in this compartment (indeed, the endogenous β-galactosidase gene *Glb1* is more than six-fold upregulated, specifically in the duodenum at E16.5), we used *in situ* hybridization with an antisense *Axin2* riboprobe to directly examine this question. We found that at E14.5, *Axin2* expression is very low in both stomach and duodenum, indicating that the level of canonical Wnt pathway activity is low at this time (Figure II.7C). At E16.5, *Axin2* expression and therefore, canonical Wnt pathway activity increases considerably, but no visible gradient of canonical signaling activity can be detected across the pyloric region (Figure II.7D). Interestingly, in the duodenum, *Axin2* staining is exclusively seen in the intervillus region of E16.5 intestine (Figure II.7E); no signaling activity is detectable in villus tip epithelium.

A specific domain of gene expression at the pylorus

Using the group of tissue enriched probesets described earlier, we selected probesets that showed a pyloric expression pattern that was significantly different from both the stomach and duodenum at either E14.5 or E16.5. Of these probesets with pylorus-specific expression patterns, fifteen were detected at E14.5 and 79 were detected at E16.5. Known transcription factors and signaling pathway components among these pylorus-specific probesets are shown in Table II.6 and the remaining pyloric probesets are listed in Table II.7.

Pylorus-specific genes detectable at both E14.5 and E16.5 include the transcription factor *Nkx2-5* and the secreted BMP antagonist *gremlin*, both of which were previously shown

to exhibit pyloric expression in chick [37, 38]. However, the most pylorus-enriched gene detected at E14.5 was *Gata3*, which encodes a transcription factor not previously associated with pylorus patterning or function. Furthermore, at E16.5, the most enriched gene at the pylorus was *nephrocan*, a secreted regulator of TGF-β pathway activity [39]. We confirmed these novel findings by *in situ* hybridization.

Nephrocan and gremlin

Gremlin and nephrocan are both secreted modulators of TGF-β superfamily signaling and thus it was of interest to compare their expression patterns. At E14.5 and E16.5, *gremlin* is mesenchymal and expressed in a broad band at the pylorus (Figures 6A-B). Interestingly, *nephrocan* is expressed in E14.5 pyloric epithelium (Figure II.8C); its expression is slightly asymmetric with respect to the pylorus, with a broader expression domain on the stomach side. At E16.5, *nephrocan* becomes restricted to cells to the intervillus base of the intestine and cells at the base of developing antral glands. Interestingly, despite its epithelial localization, there is no clear boundary of expression across the pylorus for *nephrocan* at E16.5 (Figure II.8D).

Gata3

Gata3 expression is detectable in a narrow band at the pylorus at E14.5 (Figure II.8E). Staining is restricted to the mesenchyme and appears confined to cells outside of the thick inner circular smooth muscle of the distal stomach (Figure II.7F-G). Previous studies of *Nkx2-5* showed a similar well-demarcated mesenchymal pattern at the pylorus [40]. It

will be interesting to further compare the extent of possible overlap in both expression and function between these two transcription factors.

II.4. Discussion

In this study, we have investigated dynamic gene expression patterns at and around the developing pylorus. Although clear regional patterning of the stomach and intestine occurs prior to E14.5 (e.g., for *Sox2*, *Cdx2*, among other genes), this pattern does not play itself out in terms of the global transcriptomes of E14.5 stomach, duodenum, and pylorus tissues, all of which are surprisingly similar. In contrast, at E16.5, a dramatic burst of transcriptional induction occurs in duodenal epithelium and this event generates a distinct compartmentalization of gene expression on the duodenal side of the pyloric border. This genetic induction event coordinately activates hundreds of genes involved in absorption and metabolism. As a result, for the first time, epithelial cells of the intestine express genes that unambiguously distinguish their function during digestion from that of the stomach. We call this compartmentalized patterning step *intestinalization* and note that it occurs strikingly late in fetal development.

Though the vast majority of genes activated in the intestine are epithelial, a few mesenchymal genes are also upregulated at E16.5, several of which are involved in immune response function. This is interesting in light of recent parallel evidence from our laboratory showing that, in adult intestine, decreased Hh signal transduction increases inflammatory signaling [41]. Whether there is a direct connection between the

downregulation of the Hh pathway that we observe in duodenal mesenchyme at E16.5 (see Figures 4B,D) and activation of these mesenchymal inflammatory genes requires further investigation.

An unexpected finding from our array results is that only four genes (i.e., *Foxa1*, *Gcg*, *Mreg*, and *Serpina1c*) are specifically downregulated in intestinal epithelium at E16.5 (data not shown). Interestingly, *Foxa1* has been shown to be required for Shh expression in the developing lung, another foregut endoderm-derived organ [42]. Because of its concomitant downregulation with *Shh* in intestinal epithelium, it is tempting to postulate that the attenuation of *Foxa1* expression is responsible for reduced *Shh* expression during intestinalization. If so, it will be important to understand the transcriptional regulation of *Foxa1* expression and determine whether downregulation of this gene is required for intestinalization.

Given the ample published evidence of a role for Wnt signaling during foregut specification [26-28], we were surprised to find that the *Axin2* expression pattern suggests little or no difference in canonical Wnt signaling across the pylorus, either before or after pyloric border formation. In fact, at E14.5, very little canonical Wnt activity is detectable at all in either distal stomach or intestine. The lack of canonical Wnt signal in E14.5 duodenum is concordant with the phenotype of *transcription factor 7-like 2, T-cell specific, HMG-box* (*Tcf7l2*/*Tcf4*) null mice, which exhibit no apparent defect in E14.5 intestine [43]. At E16.5, we see that intestinal *Axin2* expression is confined to intervillus regions (see Figure II.7E). This is in accordance with the finding that these

proliferative cells are dependent upon canonical Wnt signals, as shown by the loss of this proliferative population in the face of either *Tcf4* deficiency [43] or *Dkk1* overexpression [44, 45]. Together, these findings and the results of our analysis do not support the idea [26] that a canonical Wnt signaling compartment exists on villus tips at E16.5.

In this regard, the expression pattern of *Sfrp5* at E16.5 is interesting. Recent work indicates that Sfrp5 can modulate either canonical or non-canonical Wnt signals in the Xenopus foregut [46]. In that system, Sfrp5 was able to bind the ortholog of the non-canonical protein Wnt5a. Intriguingly, both overexpression of *Sfrp5* (in Xenopus) and deficiency of *Wnt5a* (in mouse) result in shortened hindgut [46, 47]. Data presented above show that, at E16.5, *Sfrp5* expression is restricted to the intervillus region (see Figure II.7B), the same area that is positive for *Axin2* staining (and, hence, canonical pathway activity). Thus, it will be of interest to test whether Sfrp5 functions in the intervillus zone to modulate canonical signals, non-canonical signals, or both.

Several epithelial transcription factors are dramatically upregulated during intestinalization and may participate directly in large-scale induction of absorptive and metabolic activity in the intestine. The *Hnf4*γ paralog *Hnf4*α has been previously implicated in a similarly late developmental maturation event in the liver. In that system, Hnf4α upregulates a large number of structural genes and is thought to be important for the re-epithelialization of hepatic cells following their migration out of the gut tube proper and into the septum transversum [48]. It is tempting to speculate that intestinalization is a similar event. Even though intestinal epithelial cells never leave the

confines of the epithelial sheet, as developing hepatoblasts do, the epithelium itself is drastically reorganized during the process of villus formation. Perhaps *Hnf4α* and *Hnf4γ* are critical in the final stabilization of this remodeled state. Certainly, binding sites for these factors are highly enriched in the promoters of intestine-specific and epithelial-specific genes [10].

Two previously unstudied factors were among the most upregulated D16 epithelial transcription factors. Creb3l3 (also known as Creb-H), a member of the bZip family of transcription factors, is involved in the endoplasmic reticulum (ER) stress response [49], and interestingly, its expression in the developing liver is dependent upon Hnf4α [50]. Since intestinalization involves transcriptional activation of hundreds of genes, several of which are expressed at tremendously high levels, the ER of intestinal epithelial cells may abruptly require a much higher degree of organization and efficiency to deal with the translational onslaught that follows. Indeed, we show that *Creb3l3* is expressed in epithelial cells of the villi, exactly the population in which differentiated gene expression is induced. The idea that ER stress might accompany cell differentiation and might activate mediators of the response pathway in order to coordinate protein biosynthesis remains functionally untested here, but has been well documented for a number of secretory cell types [51].

Another transcription factor that is highly induced in E16.5 intestinal epithelium is *Tcfec*. This bHLH-Zip factor is a member of the MiT family (with *Mitf*, *Tcfeb*, and *Tcfe3*), several of which are expressed in a highly tissue- and cell-specific manner. Often these

proteins are responsible for the expression of signature proteins that are critical to organ or tissue development and function. For example, MiT family members regulate tartrate-resistant alkaline phosphatase in osteoclasts [52], melanin in pigment cells [53] and proteases in mast cells [54]. The MiT proteins form both homo- and hetero-dimers, a fact that may explain why mouse knockout models of several family members show no phenotypes, despite the apparent transcriptional importance of these genes [55]. According to our microarray data, the related family member *Tcfeb* is also expressed in the intestine and is upregulated during intestinalization, though not as dramatically as *Tcfec*.

Intestinalization occurs concomitantly with formation of a sharp boundary of epithelial gene expression at the pylorus. For genes like *villin*, *Cdx2*, and *Sox2*, a broad domain of expression with a diffuse boundary is detectable early and reflects the regional divisions of organ territory in the developing gut tube. But at E16.5, the boundary of expression sharpens exquisitely to allow differentiated intestinal cells to lie directly next to future stomach cells. An interesting question for further analysis is whether boundary formation and intestinalization actually constitute separate events. The process of intestinalization might reflect maturation of the vertical axis of the villus; differentiating cells exiting the proliferative compartment of the last villus next to the stomach may travel only a set distance from that crypt, coming to rest immediately next to a less differentiated neighbor derived from the stomach progenitor compartment. Alternatively, the pyloric border region itself could propagate a signal that promotes intestinalization, similar to the function of a classical organizer [56].

Our data confirm and extend earlier studies [6] that reveal a characteristic domain of gene expression at the pylorus. We show that this domain is present both prior to and after intestinalization. We report here the novel finding that, similar to *Nkx2-5*, *Gata3* is expressed in a narrow band at the pylorus (see Figure II.7E). Given that Nkx and Gata family members are known to interact in other developmental systems, these factors may collaboratively regulate pyloric patterning and organogenesis [57-59]. In addition, we describe the pyloric expression patterns of *gremlin* (mesenchyme; see Figures 6A-B) and *nephrocan* (epithelium; see Figures 6C-D), two secreted modulators of TGF-β superfamily signaling. To our knowledge, nephrocan is the first secreted signaling protein to be identified in pyloric epithelium. In this regard, pyloric border patterning might be similar to the boundary patterning events observed at the midbrain-hindbrain border in the developing brainstem and the atrioventricular boundary of the heart. These events involve formation of straight, sharp expression boundaries [60-62], and in both cases, the border region itself has signaling activity that influences neighboring tissues [61, 63].

The intestinalization event that we document occurs without a similar maturation process in the stomach, where only limited transcriptional change occurs. The fact that the intestine, a more posterior tissue, matures prior to the stomach is somewhat surprising, given the tendency of embryonic development to progress in an anterior to posterior direction. Indeed, it is possible that this finding has evolutionary roots, as the stomach is believed to be an added character that first appeared in primitive fish [38]. It is possible then, that the stomach epithelium at E16.5 is governed by a program designed to

represses the emerging intestinal state in order to preserve the primitive stomach epithelium for the later reception of instructive signals to differentiate as stomach. Though this notion is entirely speculative, it has interesting implications for intestinal metaplasia, a pathological lesion in which patches of epithelium with intestinal character emerge in the stomach. The possibility that active repression of intestinal differentiation in stomach exists during pyloric border formation (and persists throughout adult life) will become amenable to further investigation now that the transcriptomes of stomach and intestine are available during this important developmental event.

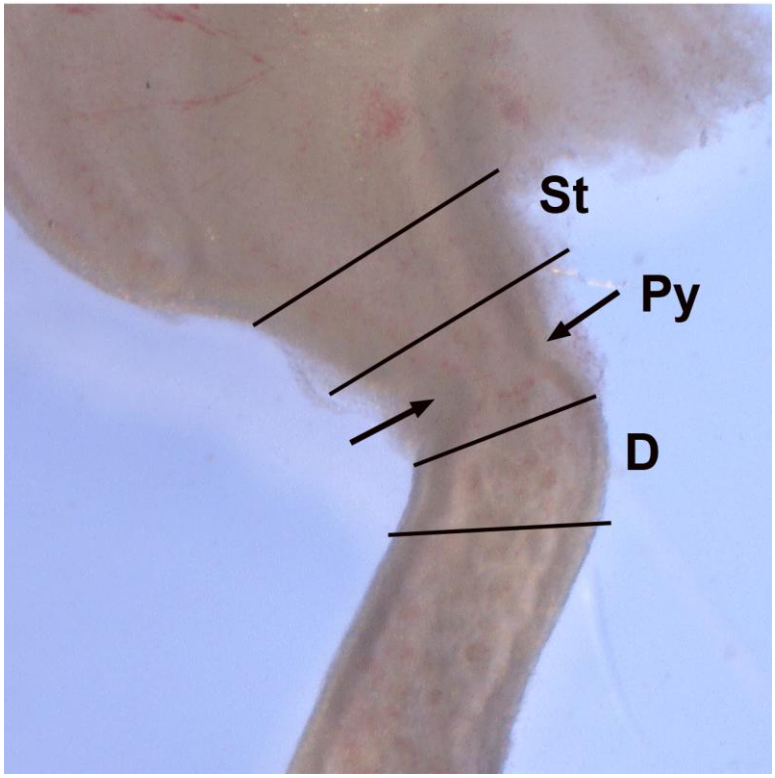

Figure II.1. Diagram of microdissection for microarray experiment. A photomicrograph showing E14.5 foregut. The pylorus was identified grossly by its muscular constriction (arrows). Small, contiguous pieces of tissue were dissected from the pylorus and surrounding stomach and duodenum. St = stomach, Py = pylorus and D = duodenum. Photograph by Dr. Tracy (Xiaotan) Qiao.

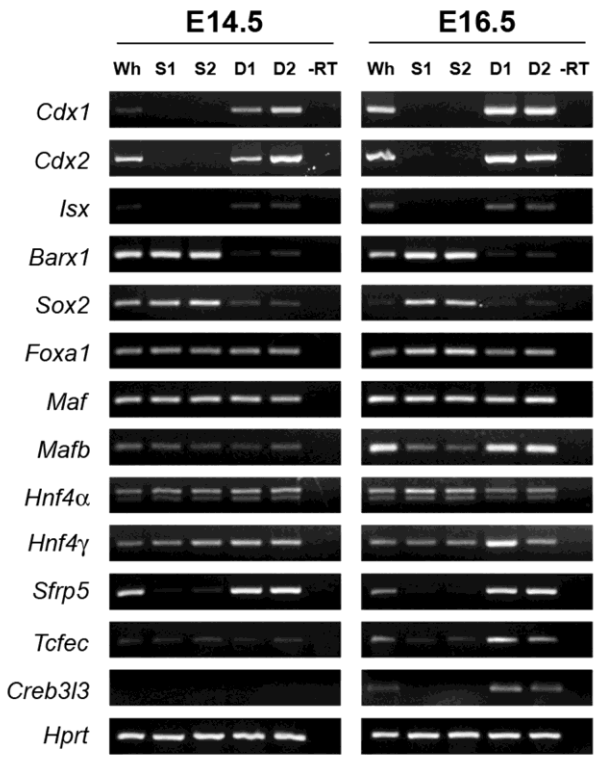

Figure II.2. Independent validation of microarray data by RT-PCR. Semi-quantitative RT-PCR was carried out as described in Supplemental Methods. *Hprt* is used as a control; it does not vary with time and tissue. Tissue specific regulation of *Cdx1*, *Cdx2*, *Isx*, *Barx1*, *Sox2*, and *Sfrp5* is set by E14.5 and does not vary with time. *Mafb* and *Hnf4γ* are preferentially, but not specifically, expressed in the intestine. *Tcfec* and *Creb3l3* are greatly induced in E16.5 duodenum. RT-PCR by Dr. Neil Richards.

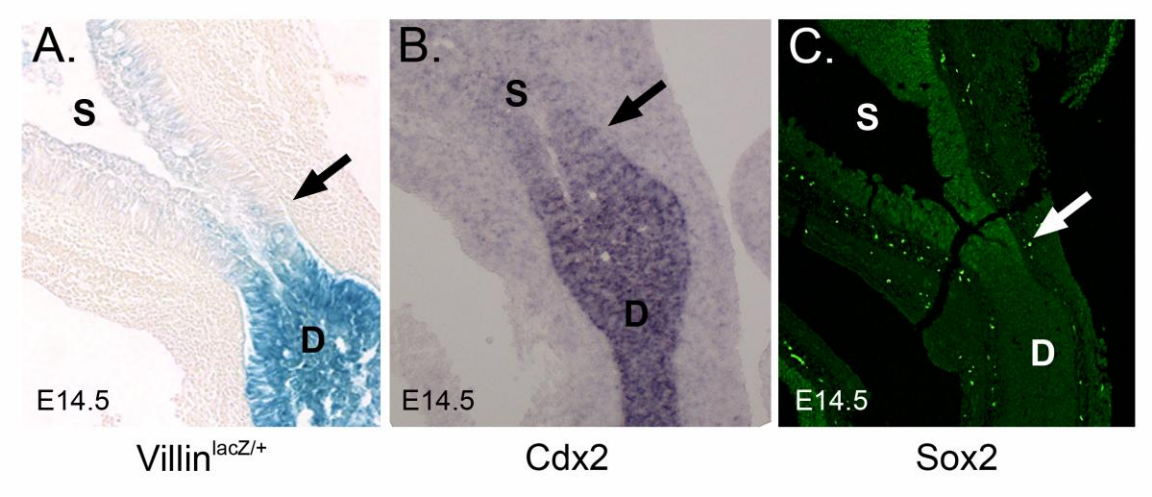

Figure II.3. The epithelial pyloric boundary is diffuse at E14.5. A) X-gal staining for β-gal expression in *villin*$^{lacZ/+}$ mice [1]. B) *In situ* hybridization for *Cdx2*. C) Immunofluorescence (green) staining for Sox2. Staining was performed on sectioned E14.5 material. The presumptive pyloric border is indicated by the arrow. D = duodenum and S = stomach. X-gal staining and immunostaining by Dr. Tracy (Xiaotan) Qiao.

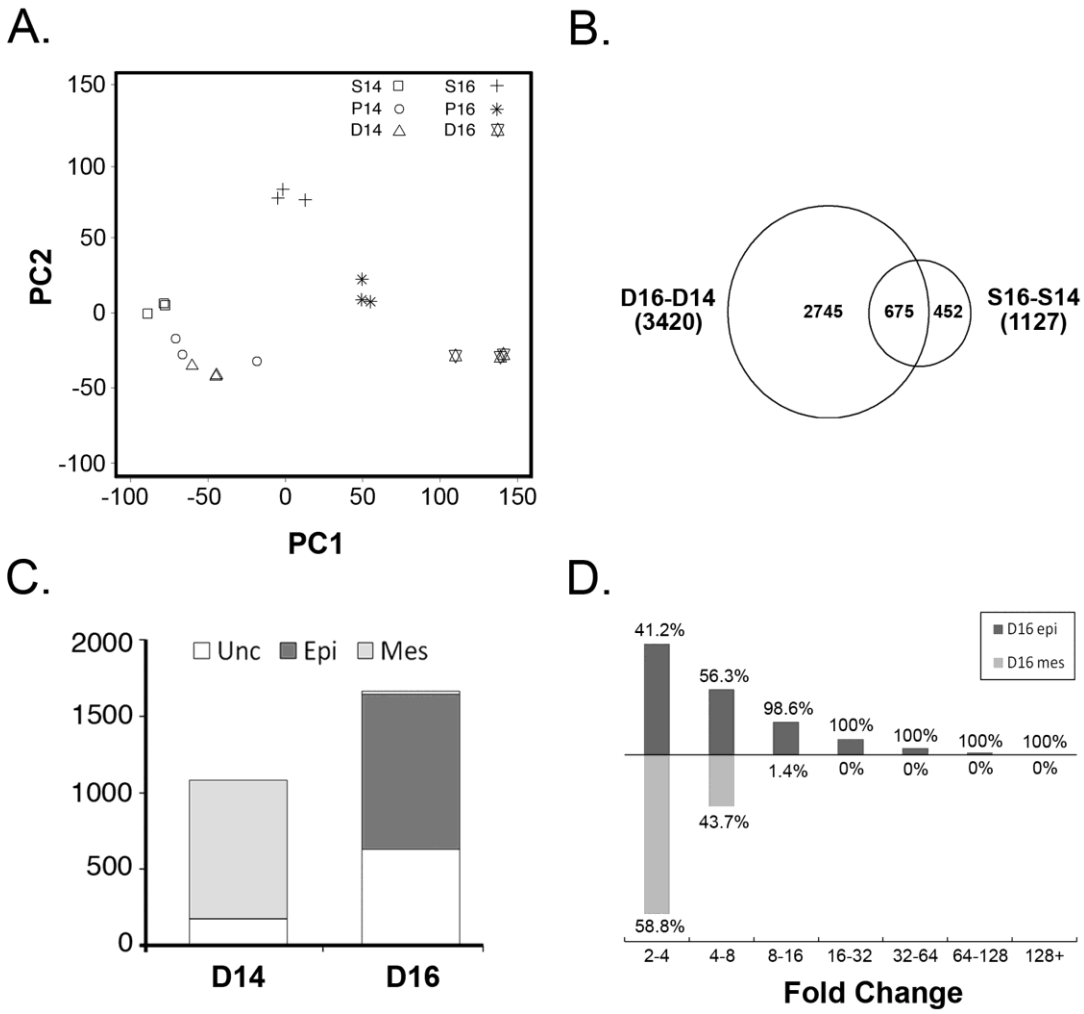

Figure II.4. Dramatic upregulation of gene expression in E16.5 duodenal epithelium. A) Principal components analysis of individual microarray chips. The first two principal components (PC1 and PC2), which together represent the majority of the sample variance, are plotted. Note grouping of E14.5 tissues. B) Venn diagram of temporal changes (e.g., D16-14, S16-S14) for tissue-enriched probesets at E16.5. Overlap between these groups indicates probesets that change in both tissues from E14.5 to E16.5. C) Epithelial (Epi) or mesenchymal (Mes) compartmentalization of enriched (D16) or depleted (D14) probesets in E16.5 duodenum. Probesets with low expression or expression in both compartments are considered unclassified (Unc). D) Histogram of duodenal (D16-D14) fold changes for E16.5 upregulated epithelial probesets (D16 epi) and downregulated mesenchymal probesets (D16 mes). The height of the bar is proportional to the absolute number of probesets. Percentages indicate the relative proportion of D16 epi or D16 mes probesets represented by each bar for each fold change category. Principal components analysis by Dr. Xing Li.

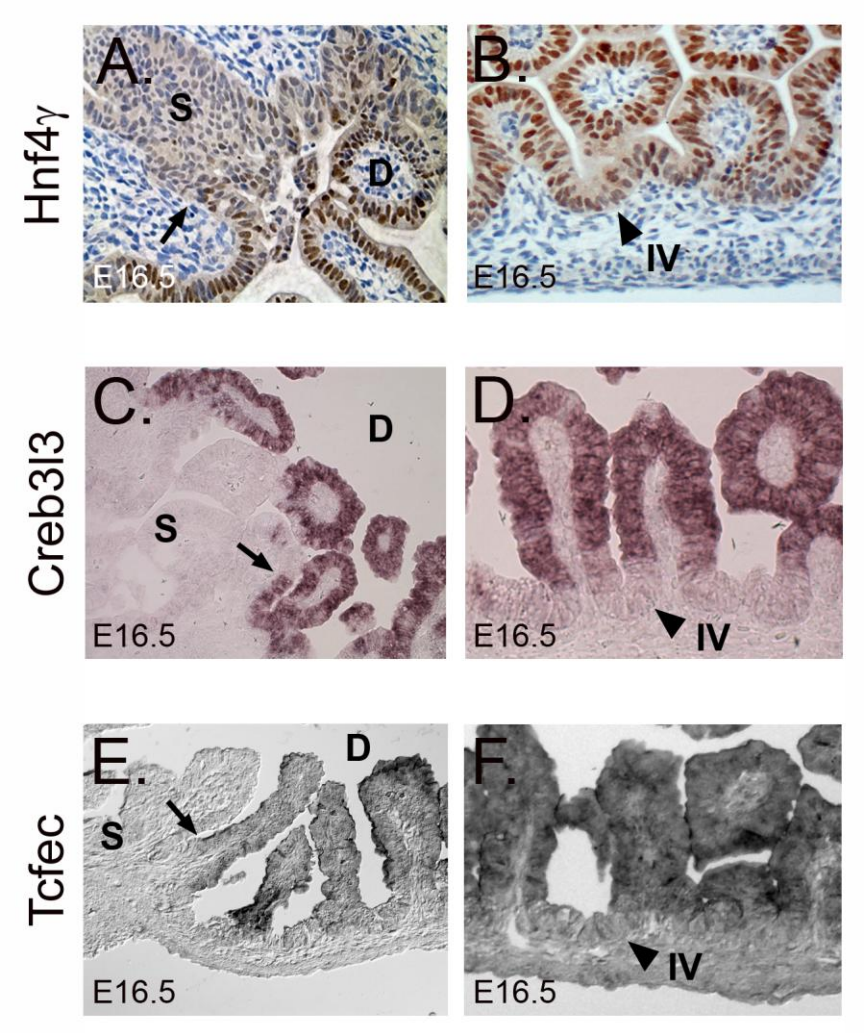

Figure II.5. The duodenal epithelial transcription factors Hnf4γ, Creb3l3, and Tcfec have sharp anterior expression boundaries at E16.5. A) Immunohistochemical staining for Hnf4γ reveals strong nuclear staining in the intestine, but only weak nuclear staining in the stomach. B) A gradient of nuclear Hnf4γ staining from intervillus (weak) to villus tip (strong) epithelium is present. C) *In situ* hybridization for *Creb3l3* shows a dramatic boundary at the stomach-intestine border at E16.5. *Creb3l3* is not expressed at E14.5 (data not shown). D) *Creb3l3* is restricted to villus epithelium at E16.5. E) *In situ* hybridization for *Tcfec* reveals a discrete boundary of expression at E16.5. F) *Tcfec* is expressed predominately in differentiated duodenal villus epithelium and not intervillus epithelium. Arrows denote stomach-intestine expression boundary. Arrowheads indicate intervillus (IV) epithelium. D = duodenum and S = stomach.

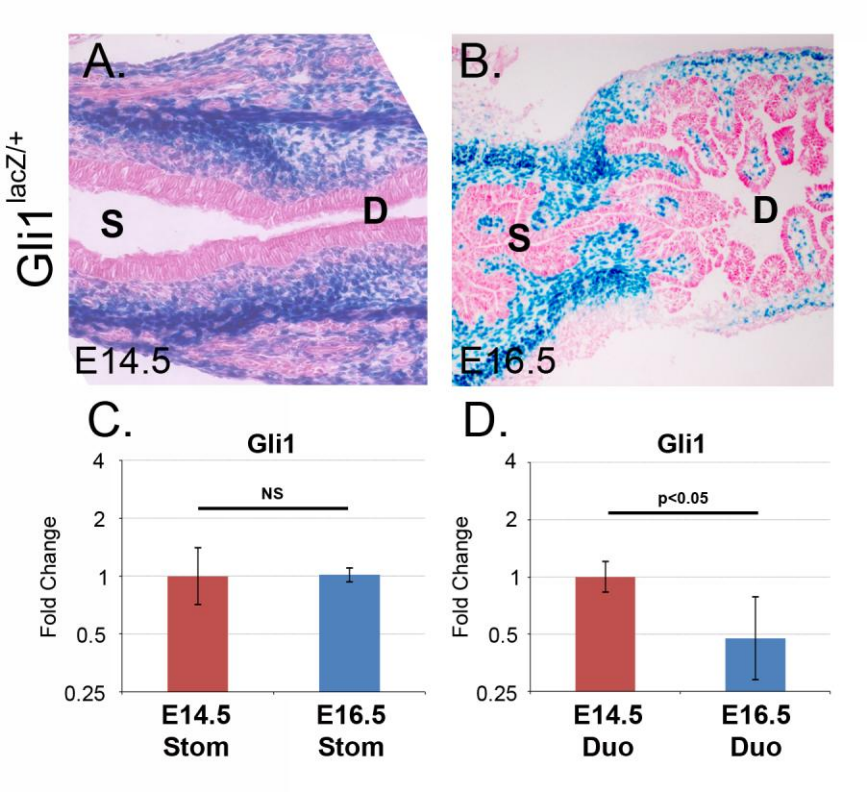

Figure II.6. Downregulation of Hedgehog signaling in duodenum at E16.5. A) X-gal staining for β-gal expression in *Gli1*$^{lacZ/+}$ mice at E14.5 reveals similar levels of Hh signaling activity across the pyloric border. B) At E16.5, X-gal staining shows an apparent difference in stomach and intestinal Hh signaling. C-D) Histogram of fold change values for normalized *Gli1* qPCR data. Compared to the same tissue at E14.5, *Gli1* expression was significantly decreased in E16.5 intestine but not in stomach. NS = not statistically significant. D, duodenum; S, stomach. X-gal staining by Dr. Åsa Kolterud.

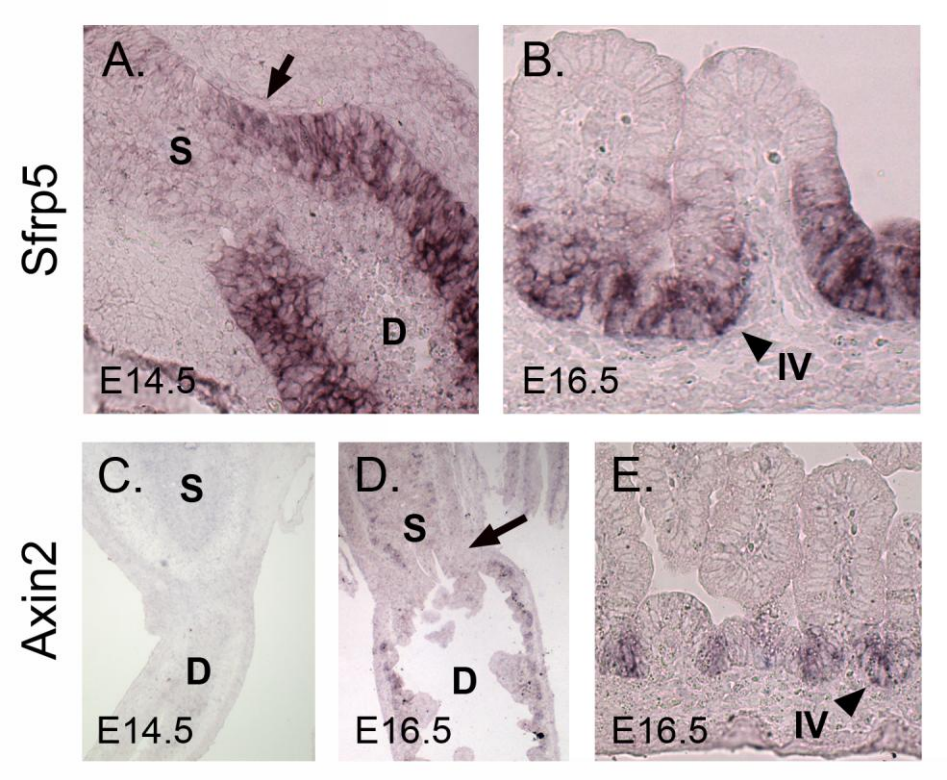

Figure II.7. Canonical Wnt signaling is active across the pylorus at E16.5 and restricted to intervillus epithelium in the duodenum. A) *In situ* hybridization for *Sfrp5* at E14.5. Note diffuse boundary at pyloric border (arrow). B) *Sfrp5* expression in the E16.5 intestine reveals that expression is strong in intervillus but not villus tip epithelium. C) *In situ* hybridization for *Axin2* at E14.5; little or no signal is seen. D) *Axin2* expression at 16.5 shows faint staining in both the stomach and duodenal epithelium, with little difference between these tissues. The presumptive pyloric border is indicated by the arrow. E) Higher magnification of *Axin2* intestinal staining at E16.5 reveals that staining is confined to the intervillus epithelium. Arrowheads indicate intervillus (IV) epithelium. D = duodenum and S = stomach.

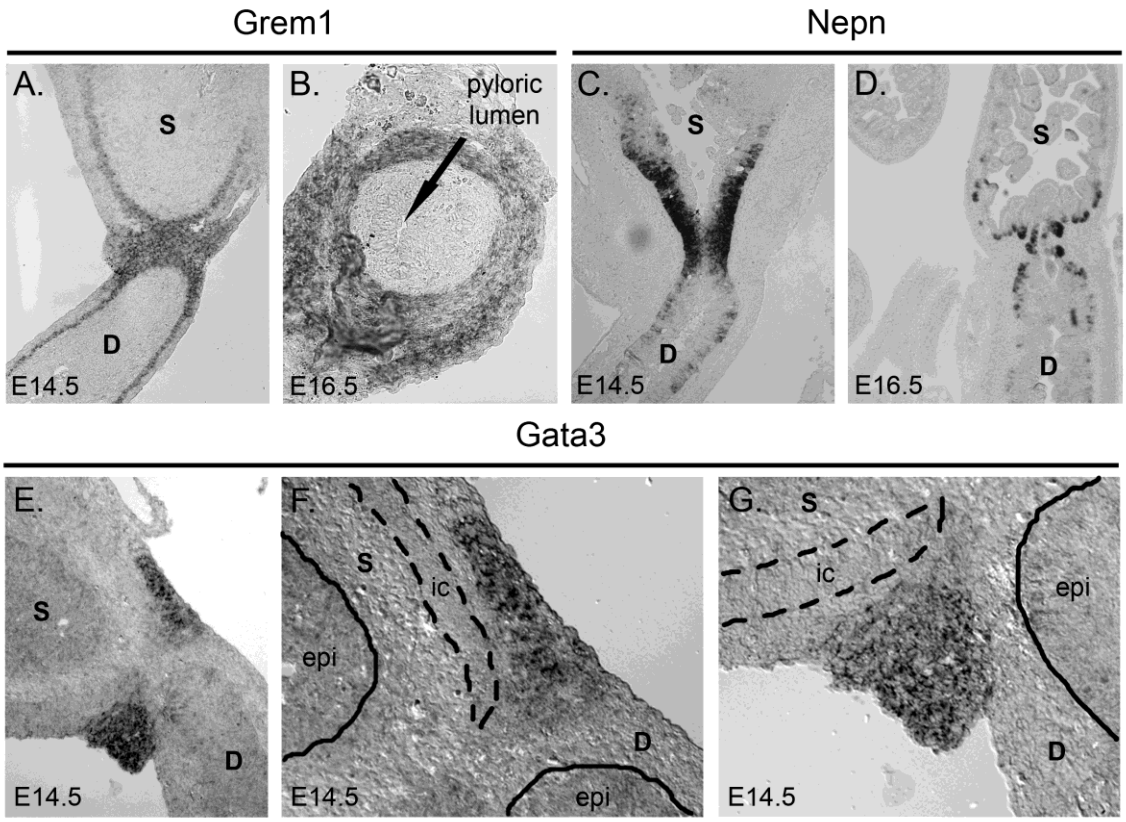

Figure II.8. Pylorus-specific expression of Gata3, gremlin, and nephrocan. A) *In situ* hybridization for *gremlin* (*Grem1*) at E14.5. Mesenchymal expression is strong at the pylorus and continues in the inner circular muscle of the intestine and more weakly in the inner circular muscle of the stomach. B) *Gremlin* expression is also mesenchymal at E16.5, as seen in a cross-section through the pylorus. Arrow shows the pyloric lumen. C) *In situ* hybridization for *nephrocan* (*Nepn*) at E14.5 reveals epithelial expression at the pylorus. Expression is more robust towards the stomach, and is primarily localized to epithelial cells closer to the basement membrane. D) *Nephrocan* expression at E16.5 is restricted to the base of the developing epithelial glands in stomach and developing villi in intestine. E) *In situ* hybridization for *Gata3* at E14.5 reveals expression in a tight band of cells at the pylorus. F-G) Higher magnification of *Gata3* pyloric staining reveals expression in mesenchyme but not epithelium. Epithelium (epi) is demarcated by the solid line, while the inner circular (ic) smooth muscle is outlined by the dashed line. D = duodenum and S = stomach. *In situ* hybridization by Dr. Chunbo Hu.

Table II.1. Summary of RT-PCR primer sequence and optimized conditions.

Gene Symbol	Forward primer	Reverse primer	Amplicon length (bp)	Temp (deg C)	Cycles	$MgCl_2$ (mM)	DMSO[1,3]
Barx1	GGAGTCGCACCGTATTCACTGAGC	CCGCCACCTTGCAGCACTATTTTC	187	58	30	2.5	NO
Cdx1	CGGACGCCCTACGAATGGATG	CTGCTGCTGCTGCTGTTTCTTC	276	61	30	1.5	YES
Cdx2	GAAACCTGTGCGAGTGGATGC	TTGTTGCTGCTGCTGCTGTTG	284	58	30	1.5	YES
Creb3l3	CAGTGGCATCTCTGAGGATCTACC	CAGTGAGGTTGAAGCGGGAGG	266	61	30	1.5	YES
Foxa1	GGCATGAGAGCAACGACTGG	TAGGTGTTCATGGAGTTCATAGAGC	115	57	30	2.5	NO
Gli1	CGCCAAGCACCAGAATCGG	CCGAGACACAAGGTCCTTCATCC	419	61	40	6[2]	YES
Hnf4a	GATGCTTCTCGGAGGGTCTGC	TTGGTGGTGATGGCTGTGGAG	200	60	30	1.5	NO
Hnf4g	CGCAGCATTCGGAAGAGTCATG	CCGCTTGTGCCAGAGTGTTTATG	220	60	30	1.5	NO
Hprt	AGTCCCAGCGTCGTGATTAGC	ATAGCCCCCCTTGAGCACACAG	204	61	30	2	YES
Isx	ACTTCACCCATTACCCTGACATCC	TCTTCTCCTGCTTCCTCCACTTG	123	55	30	1.5	YES
Maf	CAACGGCTTCCGAGAAAACG	CCCACGGAGCATTTAACAAGG	111	56	30	2.5	YES
Mafb	GCAACAGCTACCCACTAGCC	AGCTGGTCATCAGAGAAGCG	108	60	30	2.5	YES
Sfrp5	CCCTGGACAACGACCTCTGC	CACAAAGTCACTGGAGCACATCTG	143	59	30	2.5	YES
Sox2	CTCGCAGACCTACATGAACG	AGTGGGAGGAAGAGGTAACC	146	59	28	3	NO
Tcfec	ATGAACCCATGAGCCCAGACAG	AGCATCCGTGAGACCAGCATTAG	173	56	30	2	YES

[1]To be used at a 2% final concentration.

[2]iQ Supermix contains 6 mM $MgCl_2$ (qPCR only).

[3]To be used at a 4% final concentration (qPCR only).

Table II.2. DAVID analysis of D16 enriched epithelial genes.

Annotation Cluster 1	Enrichment Score: 5.05	Count	P-value	Fold Enrichment	FDR
cellular lipid metabolic process		29	2.0E-7	3.1	0
Annotation Cluster 2	Enrichment Score: 4.64	Count	P-value	Fold Enrichment	FDR
transporter		40	2.4E-15	4.7	0
Annotation Cluster 3	Enrichment Score: 4.1	Count	P-value	Fold Enrichment	FDR
hydrolase activity, acting on glycosyl bonds		13	1.2E-6	6.2	0
Annotation Cluster 4	Enrichment Score: 3.95	Count	P-value	Fold Enrichment	FDR
carbohydrate metabolism		30	2.2E-10	4.2	0
Annotation Cluster 5	Enrichment Score: 3.44	Count	P-value	Fold Enrichment	FDR
acyltransferase activity		12	1.7E-4	4.1	0.3
Annotation Cluster 6	Enrichment Score: 3.18	Count	P-value	Fold Enrichment	FDR
organic acid metabolic process		22	9.3E-5	2.6	0.2

Table II.3. Summary of transcription factor gene expression changes in E16.5 duodenal epithelium. For each comparison, the label (e.g., D14, D16, S16, Epi, Mes) refers to the time and/or tissue of maximum expression and the number in parentheses is the fold change.

Probeset ID	Symbol	Description	D16-S16	D16-D14	Epi-Mes
1419537_at	Tcfec	transcription factor EC	D16 (41.65)	D16 (5.45)	Epi (18.27)
1424688_at	Creb3l3	cAMP responsive element binding protein 3-like 3	D16 (18.95)	D16 (25.92)	Epi (11.26)
1423631_at	Nr2e3	nuclear receptor subfamily 2, group E, member 3	D16 (18.68)	D16 (23.80)	Epi (4.60)
1451716_at	Mafb	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B (avian)	D16 (14.49)	D16 (14.49)	Epi (2.93)
1460127_at	Hnf4g	hepatocyte nuclear factor 4, gamma	D16 (13.19)	D16 (3.66)	Epi (49.82)
1449051_at	Ppara	peroxisome proliferator activated receptor alpha	D16 (9.09)	D16 (5.04)	Epi (6.58)
1425392_a_at	Nr1i3	nuclear receptor subfamily 1, group I, member 3	D16 (8.46)	D16 (8.22)	Epi (6.08)
1419185_a_at	Mlxipl	MLX interacting protein-like	D16 (7.87)	D16 (2.93)	Epi (2.40)
1437473_at	Maf	avian musculoaponeurotic fibrosarcoma (v-maf) AS42 oncogene homolog	D16 (6.89)	D16 (5.38)	Epi (2.41)
1417244_a_at	Irf7	interferon regulatory factor 7	D16 (6.77)	D16 (8.09)	Epi (2.18)
1425528_at	Prrx1	paired related homeobox 1	D16 (5.35)	D16 (9.57)	Epi (2.73)
1434416_a_at	Solh	Small optic lobes homolog (Drosophila)	D16 (4.34)	D16 (2.64)	Epi (2.84)
1417519_at	Plagl2	pleiomorphic adenoma gene-like 2	D16 (3.89)	D16 (2.38)	Epi (7.06)
1440831_at	Bach1	BTB and CNC homology 1	D16 (3.84)	D16 (2.32)	Epi (2.67)
1449854_at	Nr0b2	nuclear receptor subfamily 0, group B, member 2	D16 (3.31)	D16 (5.04)	Epi (2.55)
1420808_at	Ncoa4	nuclear receptor coactivator 4 /// predicted gene, EG627557	D16 (2.98)	D16 (2.53)	Epi (2.72)
1435991_at	Nr3c2	nuclear receptor subfamily 3, group C, member 2	D16 (2.55)	D16 (2.57)	Epi (2.36)
1443100_at	Thrb	Thyroid hormone receptor beta	D16 (2.27)	D16 (2.20)	Epi (3.34)
1440870_at	Prdm16	PR domain containing 16	D16 (2.20)	D16 (3.13)	Epi (5.24)
1426690_a_at	Srebf1	sterol regulatory element binding factor 1	D16 (2.10)	D16 (2.54)	Epi (2.63)
1418437_a_at	Mlx	MAX-like protein X	D16 (2.09)	D16 (2.33)	Epi (2.05)
1419052_at	Ovol1	OVO homolog-like 1 (Drosophila)	D16 (2.04)	D16 (2.14)	Epi (2.14)

Table II.4. Summary of gene expression changes in Hedgehog signaling pathway components. NC means no significant change (significant change is FC≥2 and p<0.05). For each comparison, the label (e.g., D14, D16, S14, S16) refers to the time and/or tissue of maximum expression and the number in parentheses is the fold change.

Probeset ID	Symbol	Description	D14-S14	D16-S16	S16-S14	D16-D14
1426869_at	Boc	biregional cell adhesion molecule-related/down-regulated by oncogenes (Cdon) binding protein	NC (1.05)	S16 (4.09)	NC (1.25)	D14 (4.86)
1434957_at	Cdon	cell adhesion molecule-related/down-regulated by oncogenes	NC (1.15)	S16 (2.18)	NC (1.40)	D14 (2.64)
1448494_at	Gas1	growth arrest specific 1	NC (1.11)	S16 (2.36)	NC (1.67)	D14 (4.36)
1449058_at	Gli1	GLI-Kruppel family member GLI1	NC (1.04)	S16 (3.06)	NC (1.14)	D14 (3.65)
1459211_at	Gli2	GLI-Kruppel family member GLI2	NC (1.04)	S16 (2.12)	NC (1.99)	D14 (4.40)
1455154_at	Gli3	GLI-Kruppel family member GLI3	NC (1.26)	S16 (3.07)	S14 (2.70)	D14 (6.56)
1450704_at	Ihh	Indian hedgehog	NC (1.80)	S16 (2.19)	NC (1.60)	NC (1.32)
1427133_s_at	Lrp2	low density lipoprotein receptor-related protein 2	NC (1.45)	D16 (7.29)	NC (1.24)	D16 (4.04)
1439663_at	Ptch1	Patched homolog 1	NC (1.31)	NC (1.63)	NC (1.89)	D14 (2.35)
1454876_at	Rab23	RAB23, member RAS oncogene family	NC (1.02)	S16 (2.28)	NC (1.65)	D14 (3.69)
1436869_at	Shh	sonic hedgehog	NC (1.75)	S16 (6.66)	NC (1.30)	D14 (8.94)
1427049_s_at	Smo	smoothened homolog (Drosophila)	NC (1.03)	S16 (2.21)	NC (1.61)	D14 (3.65)
1434733_at	Stk36	serine/threonine kinase 36 (fused homolog, Drosophila)	NC (1.12)	NC (1.35)	S14 (2.15)	D14 (2.58)

Table II.5. Summary of gene expression changes in Wnt signaling pathway components. NC means no significant change (significant change is FC≥2.0 and p<0.05). For each comparison, the label (e.g., P14, P16, D14, D16, S14, S16) refers to the time and/or tissue of maximum expression and the number in parentheses is the fold change.

Probeset ID	Symbol	Description	D16-D14	S16-S14	D14-S14	D16-S16
1437351_at	Cxxc4	CXXC finger 4	D14 (5.01)	NC (1.93)	NC (1.15)	S16 (2.25)
1438884_at	D830007B15Rik	RIKEN cDNA D830007B15 gene	D14 (3.70)	NC (1.36)	D14 (3.62)	NC (1.40)
1420512_at	Dkk2	dickkopf homolog 2 (Xenopus laevis)	D14 (6.56)	S14 (2.05)	NC (1.06)	S16 (3.01)
1417312_at	Dkk3	dickkopf homolog 3 (Xenopus laevis)	D14 (2.57)	NC (1.11)	NC (1.24)	S16 (2.29)
1437284_at	Fzd1	frizzled homolog 1 (Drosophila)	D14 (3.15)	NC (1.04)	NC (1.42)	S16 (2.14)
1418532_at	Fzd2	frizzled homolog 2 (Drosophila)	D14 (5.61)	S14 (2.00)	NC (1.62)	S16 (4.55)
1449730_s_at	Fzd3	frizzled homolog 3 (Drosophila)	D14 (2.24)	NC (1.43)	NC (1.12)	NC (1.76)
1417301_at	Fzd6	frizzled homolog 6 (Drosophila)	NC (1.57)	NC (1.14)	NC (1.45)	S16 (2.59)
1450044_at	Fzd7	frizzled homolog 7 (Drosophila)	D14 (4.02)	S14 (2.46)	NC (1.19)	NC (1.95)
1451022_at	Lrp6	low density lipoprotein receptor-related protein 6	NC (1.76)	S14 (2.02)	NC (1.36)	NC (1.19)
1417278_a_at	Nkd1	naked cuticle 1 homolog (Drosophila)	D14 (2.14)	NC (1.14)	NC (1.46)	NC (1.67)
1452249_at	Prickle1	prickle like 1 (Drosophila)	D14 (2.70)	NC (1.51)	NC (1.19)	NC (1.50)
1428808_at	Prickle2	prickle-like 2 (Drosophila)	D14 (3.94)	NC (1.50)	NC (1.62)	NC (1.62)
1446780_at	Ror2	Receptor tyrosine kinase-like orphan receptor 2	NC (1.23)	S14 (2.07)	NC (1.11)	NC (1.52)
1423986_a_at	Scotin	scotin gene	D16 (3.66)	S16 (3.35)	NC (1.03)	NC (1.06)
1448395_at	Sfrp1	secreted frizzled-related sequence protein 1	D14 (5.45)	S14 (3.24)	NC (1.29)	S16 (2.17)
1448201_at	Sfrp2	secreted frizzled-related protein 2	D14 (7.70)	NC (1.55)	NC (1.28)	S16 (6.37)
1451031_at	Sfrp4	secreted frizzled-related sequence protein 4	NC (1.62)	S16 (2.13)	NC (1.91)	S16 (2.52)
1436075_at	Sfrp5	secreted frizzled-related sequence protein 5	D14 (4.71)	NC (1.58)	D14 (31.69)	D16 (10.66)
1450117_at	Tcf3	transcription factor 3	D14 (6.03)	NC (1.63)	NC (1.45)	S16 (5.35)
1441756_at	Tcf7l2	Transcription factor 7-like 2, T-cell specific, HMG-box	NC (1.87)	S14 (2.45)	NC (1.87)	NC (1.43)
1423852_at	Tmem46	transmembrane protein 46	D14 (5.03)	NC (1.02)	D14 (2.79)	NC (1.84)
1438426_at	Tmem58	transmembrane protein 58	D14 (3.61)	NC (1.64)	NC (1.03)	S16 (2.25)
1436118_at	Vangl2	vang-like 2 (van gogh, Drosophila)	D14 (4.56)	NC (1.81)	NC (1.10)	S16 (2.78)
1450772_at	Wnt11	wingless-related MMTV integration site 11	NC (1.29)	NC (1.93)	NC (1.48)	S16 (2.21)
1450782_at	Wnt4	wingless-related MMTV integration site 4	NC (1.97)	S16 (6.34)	NC (1.35)	S16 (4.35)
1436791_at	Wnt5a	wingless-related MMTV integration site 5A	D14 (2.99)	NC (1.59)	S14 (2.34)	S16 (4.38)
1436978_at	Wnt9a	wingless-type MMTV integration site 9A	D14 (2.17)	NC (1.20)	NC (1.22)	NC (1.47)

Table II.6. Summary of pyloric transcription factors and signaling molecules. NC means no significant change (significant change is FC≥2 and p<0.05). For each comparison, the label (e.g., P14, P16, D14, D16, S14, S16) refers to the time and/or tissue of maximum expression and the number in parentheses is the fold change.

Probeset ID	Symbol	Description	P14-S14	P14-D14	P16-S16	P16-D16
A. Transcription factors						
1447500_at	Cutl2	cut-like 2 (Drosophila)	NC (1.67)	NC (-1.23)	P16 (2.09)	P16 (2.44)
1448886_at	Gata3	GATA binding protein 3	P14 (6.75)	P14 (10.35)	P16 (2.55)	P16 (3.39)
1449566_at	Nkx2-5	NK2 transcription factor related, locus 5 (Drosophila)	P14 (5.64)	P14 (6.58)	P16 (2.49)	P16 (2.76)
1431899_at	Nkx6-3	NK6 transcription factor related, locus 3 (Drosophila)	NC (1.51)	P14 (4.56)	P16 (2.00)	P16 (2.08)
1451569_at	Nr2c2	nuclear receptor subfamily 2, group C, member 2	NC (1.48)	NC (1.79)	P16 (2.02)	P16 (2.25)
B. Signaling molecules						
1451991_at	Epha7	Eph receptor A7	NC (1.16)	NC (1.01)	P16 (2.14)	P16 (2.10)
1424007_at	Gdf10	growth differentiation factor 10	NC (1.54)	NC (1.23)	P16 (2.74)	P16 (8.62)
1425357_a_at	Grem1	gremlin 1	P14 (7.13)	P14 (2.26)	P16 (4.90)	P16 (2.69)
1419065_at	Nepn	Nephrocan	NC (1.10)	P14 (6.31)	P16 (17.06)	P16 (16.73)
1426561_a_at	Npnt	Nephronectin	NC (1.33)	NC (1.13)	S16 (2.85)	D16 (2.05)
1422553_at	Pten	phosphatase and tensin homolog	NC (1.98)	NC (1.94)	P16 (3.36)	P16 (2.02)
1442067_at	Ror1	Receptor tyrosine kinase-like orphan receptor 1	NC (1.20)	NC (1.27)	P16 (2.29)	P16 (3.33)
1436892_at	Spred2	sprouty-related, EVH1 domain containing 2	NC (1.54)	NC (1.79)	P16 (2.87)	P16 (3.22)

Table II.7. Summary of pyloric genes. NC means no significant change (significant change is FC≥2 and p<0.05). For each comparison, the label (e.g., P14, P16, D14, D16, S14, S16) refers to the time and/or tissue of maximum expression and the number in parentheses is the fold change.

Probeset ID	Symbol	Description	P14-S14	P14-D14	P16-S16	P16-D16
1459266_at	---	---	P14 (2.05)	P14 (2.08)	NC (1.53)	NC (1.47)
1429286_at	1190003M12Rik	RIKEN cDNA 1190003M12 gene	NC (1.12)	NC (1.64)	P16 (2.48)	P16 (12.56)
1424439_at	1810065E05Rik	RIKEN cDNA 1810065E05 gene	NC (1.76)	NC (1.67)	P16 (4.51)	P16 (6.56)
1425233_at	2210407C18Rik	RIKEN cDNA 2210407C18 gene	P14 (5.01)	P14 (2.87)	NC (1.00)	P16 (17.46)
1427437_at	2610203C20Rik	RIKEN cDNA 2610203C20 gene	NC (1.54)	P14 (2.87)	P16 (2.40)	P16 (2.20)
1432556_a_at	3100002J23Rik	RIKEN cDNA 3100002J23 gene	P14 (6.30)	NC (1.46)	P16 (2.65)	P16 (2.19)
1438798_at	4931406P16Rik	RIKEN cDNA 4931406P16 gene	NC (1.62)	NC (1.75)	P16 (2.92)	P16 (2.18)
1435163_at	9030612M13Rik	RIKEN cDNA 9030612M13 gene	P14 (2.19)	P14 (2.45)	P16 (3.37)	NC (1.75)
1438531_at	A730054J21Rik	RIKEN cDNA A730054J21 gene	NC (1.69)	P14 (2.65)	P16 (2.47)	P16 (3.95)
1415927_at	Actc1	actin, alpha, cardiac	NC (1.91)	NC (1.24)	P16 (2.04)	P16 (2.58)
1451675_a_at	Alas2	aminolevulinic acid synthase 2, erythroid	P14 (3.26)	P14 (2.03)	P16 (2.89)	P16 (3.41)
1416649_at	Ambp	alpha 1 microglobulin/bikunin	P14 (2.20)	NC (1.33)	P16 (8.72)	P16 (2.39)
1459253_at	Arrdc3	Arrestin domain containing 3	P14 (2.57)	P14 (2.84)	NC (1.27)	NC (1.08)
1443801_at	AW822216	Expressed sequence AW822216	NC (1.55)	NC (1.84)	P16 (3.37)	P16 (2.10)
1438663_at	Bat2d	BAT2 domain containing 1	NC (1.32)	n/1 (1.87)	P16 (2.34)	P16 (2.28)
1440990_at	BC056349	cDNA sequence BC056349	NC (1.83)	P14 (2.34)	P16 (2.25)	P16 (3.26)
1457458_at	BC057627	cDNA sequence BC057627	NC (1.50)	NC (1.77)	P16 (2.14)	P16 (2.10)
1450624_at	Bhmt	betaine-homocysteine methyltransferase	NC (1.96)	NC (1.64)	P16 (2.56)	P16 (2.89)
1439040_at	Cenpe	centromere protein E	NC (1.40)	NC (1.55)	P16 (2.25)	P16 (3.42)
1427767_a_at	Cftr	cystic fibrosis transmembrane conductance regulator homolog	NC (1.40)	NC (1.04)	P16 (3.73)	P16 (2.33)
1436343_at	Chd4	chromodomain helicase DNA binding protein 4	NC (1.37)	NC (1.63)	P16 (2.46)	P16 (2.19)
1439427_at	Cldn9	claudin 9	NC (1.32)	NC (1.41)	P16 (2.53)	P16 (2.67)
1428571_at	Col9a1	procollagen, type IX, alpha 1	P14 (2.76)	P14 (2.48)	P16 (4.63)	P16 (10.41)
1448326_a_at	Crabp1	cellular retinoic acid binding protein I	NC (1.63)	P14 (5.02)	P16 (2.14)	P16 (5.53)
1424598_at	Ddx6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6	NC (1.49)	P14 (2.13)	P16 (2.68)	P16 (3.54)
1437403_at	E130306M17Rik	RIKEN cDNA E130306M17 gene	P14 (2.30)	P14 (2.23)	P16 (2.80)	P16 (4.29)
1456319_at	EG665081	predicted gene, EG665081	P14 (2.19)	P14 (2.36)	NC (1.58)	NC (1.72)
1449077_at	Eraf	erythroid associated factor	P14 (2.46)	NC (1.71)	P16 (2.88)	P16 (3.60)

Probeset ID	Symbol	Description	P14-S14	P14-D14	P16-S16	P16-D16
(Table II.7 continued)						
1456326_at	Gm784	gene model 784, (NCBI)	NC (1.46)	NC (1.04)	P16 (2.24)	P16 (4.08)
1423436_at	Gsta3	glutathione S-transferase, alpha 3	P14 (4.12)	NC (1.01)	P16 (2.97)	P16 (2.24)
1423016_a_at	Gypa	glycophorin A	P14 (2.86)	NC (1.96)	P16 (4.00)	P16 (4.59)
1448716_at	Hba-x	hemoglobin X, alpha-like embryonic chain in Hba complex	P14 (2.14)	NC (1.33)	P16 (2.17)	P16 (2.70)
1427866_x_at	Hbb-b1	hemoglobin, beta adult major chain	P14 (2.46)	NC (1.92)	P16 (2.44)	P16 (3.60)
1418199_at	Hemgn	hemogen	P14 (3.02)	P14 (2.07)	P16 (3.14)	P16 (3.79)
1426114_at	Hnrpab	heterogeneous nuclear ribonucleoprotein A/B	NC (1.39)	NC (1.29)	P16 (2.67)	P16 (2.00)
1458492_x_at	Hnt	neurotrimin	NC (1.26)	P14 (2.59)	P16 (2.32)	P16 (3.77)
1423276_at	Ildr1	immunoglobulin-like domain containing receptor 1	NC (1.41)	NC (1.41)	P16 (2.85)	P16 (2.07)
1442368_at	Kctd12b	potassium channel tetramerisation domain containing 12b	NC (1.72)	NC (1.80)	P16 (2.49)	P16 (3.28)
1460258_at	Lect1	leukocyte cell derived chemotaxin 1	P14 (4.27)	P14 (4.45)	NC (1.06)	NC (1.52)
1422071_at	Lgals6	lectin, galactose binding, soluble 6	P14 (7.12)	P14 (4.57)	P16 (2.79)	P16 (2.29)
1421278_s_at	LOC630963	similar to spectrin alpha 1	P14 (4.19)	NC (1.50)	P16 (5.80)	P16 (5.54)
1418188_a_at	Malat1	Metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)	NC (1.27)	P14 (2.21)	P16 (3.23)	P16 (4.01)
1451989_a_at	Mapre2	microtubule-associated protein, RP/EB family, member 2	NC (1.31)	NC (1.01)	P16 (2.13)	P16 (2.14)
1422643_at	Moxd1	monooxygenase, DBH-like 1	P14 (2.12)	NC (1.07)	P16 (2.30)	P16 (3.95)
1437250_at	Mreg	melanoregulin	S14 (3.33)	D14 (3.12)	S16 (-4.30)	NC (1.01)
1435521_at	Msi2	Musashi homolog 2 (Drosophila)	NC (1.51)	P14 (2.03)	P16 (2.25)	P16 (3.03)
1436309_at	Neto2	neuropilin (NRP) and tolloid (TLL)-like 2	P14 (2.36)	NC (1.00)	P16 (2.24)	P16 (2.98)
1448290_at	Pap	pancreatitis-associated protein	P14 (2.67)	D14 (2.09)	P16 (63.15)	P16 (2.80)
1428952_at	Pdia2	protein disulfide isomerase associated 2	NC (1.41)	NC (1.07)	S16 (5.69)	D16 (2.88)
1460332_at	Pln	phospholamban	NC (1.91)	NC (1.60)	P16 (2.00)	P16 (3.63)
1448186_at	Pnliprp2	pancreatic lipase-related protein 2	P14 (2.49)	NC (1.50)	P16 (23.05)	P16 (2.39)
1449876_at	Prkg1	protein kinase, cGMP-dependent, type I	NC (1.82)	NC (1.19)	P16 (2.33)	P16 (2.97)
1460633_at	Prpf19	PRP19/PSO4 pre-mRNA processing factor 19 homolog (S. cerevisiae)	NC (1.81)	NC (1.66)	P16 (2.02)	P16 (2.12)
1454791_a_at	Rbbp4	retinoblastoma binding protein 4	NC (1.94)	P14 (2.11)	P16 (2.56)	P16 (2.44)
1438069_a_at	Rbm5	RNA binding motif protein 5	NC (1.75)	P14 (2.16)	P16 (3.15)	P16 (2.88)

Probeset ID	Symbol	Description	P14-S14	P14-D14	P16-S16	P16-D16
(Table II.7 continued)						
1442263_at	Rgs13	regulator of G-protein signaling 13	NC (1.03)	NC (1.01)	P16 (3.50)	P16 (2.90)
1419014_at	Rhag	Rhesus blood group-associated A glycoprotein	P14 (2.84)	P14 (2.01)	P16 (2.12)	P16 (2.33)
1422552_at	Rprm	reprimo, TP53 dependent G2 arrest mediator candidate	NC (1.76)	P14 (5.06)	P16 (2.14)	P16 (3.67)
1460623_at	Skap2	src family associated phosphoprotein 2	P14 (2.56)	P14 (2.06)	NC (1.19)	NC (1.04)
1447517_at	Skiv2l2	superkiller viralicidic activity 2-like 2 (S. cerevisiae)	P14 (3.88)	P14 (3.84)	NC (1.31)	NC (1.51)
1420334_at	Slc12a8	solute carrier family 12 (potassium/chloride transporters), member 8	NC (1.22)	NC (1.22)	P16 (2.77)	P16 (2.04)
1434502_x_at	Slc4a1	solute carrier family 4 (anion exchanger), member 1	P14 (2.69)	NC (1.71)	P16 (4.11)	P16 (5.28)
1436853_a_at	Snca	synuclein, alpha	NC (1.90)	NC (1.72)	P16 (2.40)	P16 (3.87)
1421277_at	Spna1	spectrin alpha 1	P14 (2.96)	NC (1.22)	P16 (2.73)	P16 (2.09)
1441858_at	Uck2	Uridine-cytidine kinase 2	NC (1.23)	P14 (2.03)	P!6 (2.19)	P16 (2.42)
1439174_at	Unc5c	Unc-5 homolog C (C. elegans)	NC (1.77)	NC (1.09)	P16 (2.48)	P16 (2.28)
1426305_at	Upk1a	uroplakin 1A	P14 (2.73)	NC (1.27)	P16 (4.91)	P16 (4.09)
1446886_at	Usp3	Ubiquitin specific peptidase 3	P14 (2.23)	P14 (2.86)	P16 (2.56)	NC (1.20)
1419195_at	Wfdc15	WAP four-disulfide core domain 15	P14 (2.50)	NC (1.20)	P16 (2.90)	P16 (2.34)

II.5. Attribution

The first draft of this chapter was written by Dr. Xing Li and is included in his doctoral dissertation “Bioinformatic analysis of epithelial:mesenchymal crosstalk during mouse gut development and patterning.” Subsequent extensive revisions to the text were made by Aaron Udager, in consultation with Dr. Deborah Gumucio.

II.6. Publication

With the exception of minor revisions, the text of this chapter has been published in *Developmental Dynamics* [2009 Dec;238(12):3205-17; PMID: 19877272].

II.7. References

1. Braunstein, E.M., Qiao, X.T., Madison, B., Pinson, K., Dunbar, L., and Gumucio, D.L. (2002). Villin: A marker for development of the epithelial pyloric border. Dev Dyn *224*, 90-102.

2. Wells, J.M., and Melton, D.A. (1999). Vertebrate endoderm development. Annual review of cell and developmental biology *15*, 393-410.

3. Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K., and Yamamura, K. (2002). Region-specific gastrointestinal Hox code during murine embryonal gut development. Development, growth & differentiation *44*, 77-84.

4. Sherwood, R.I., Chen, T.Y., and Melton, D.A. (2009). Transcriptional dynamics of endodermal organ formation. Dev Dyn *238*, 29-42.

5. 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.

6. Smith, D.M., Nielsen, C., Tabin, C.J., and Roberts, D.J. (2000). Roles of BMP signaling and Nkx2.5 in patterning at the chick midgut-foregut boundary. Development (Cambridge, England) *127*, 3671-3681.

7. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome biology *5*, R80.

8. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic acids research *31*, e15.

9. Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. BioTechniques *34*, 374-378.

10. Li, X., Madison, B.B., Zacharias, W., Kolterud, A., States, D., and Gumucio, D.L. (2007). Deconvoluting the intestine: molecular evidence for a major role of the mesenchyme in the modulation of signaling cross talk. Physiological genomics *29*, 290-301.

11. Que, J., Okubo, T., Goldenring, J.R., Nam, K.T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development (Cambridge, England) *134*, 2521-2531.

12. Park, H.L., Bai, C., Platt, K.A., Matise, M.P., Beeghly, A., Hui, C.C., Nakashima, M., and Joyner, A.L. (2000). Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. Development (Cambridge, England) *127*, 1593-1605.

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

14. Ringner, M. (2008). What is principal component analysis? Nature biotechnology *26*, 303-304.

15. Li, X., Udager, A.M., Hu, C., Qiao, X.T., Richards, N., and Gumucio, D.L. (2009). Dynamic patterning at the pylorus: formation of an epithelial intestine-stomach boundary in late fetal life. Dev Dyn *238*, 3205-3217.

16. Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome biology *4*, P3.

17. 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.

18. Li, J., Ning, G., and Duncan, S.A. (2000). Mammalian hepatocyte differentiation requires the transcription factor HNF-4alpha. Genes & development *14*, 464-474.

19. Miura, A., Yamagata, K., Kakei, M., Hatakeyama, H., Takahashi, N., Fukui, K., Nammo, T., Yoneda, K., Inoue, Y., Sladek, F.M., et al. (2006). Hepatocyte nuclear factor-4alpha is essential for glucose-stimulated insulin secretion by pancreatic beta-cells. The Journal of biological chemistry *281*, 5246-5257.

20. McKay, I.J., Muchamore, I., Krumlauf, R., Maden, M., Lumsden, A., and Lewis, J. (1994). The kreisler mouse: a hindbrain segmentation mutant that lacks two rhombomeres. Development (Cambridge, England) *120*, 2199-2211.

21. Cordes, S.P., and Barsh, G.S. (1994). The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor. Cell *79*, 1025-1034.

22. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nature genetics *41*, 56-65.

23. Chen, J., Rattner, A., and Nathans, J. (2005). The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes. J Neurosci *25*, 118-129.

24. Choi, M.Y., Romer, A.I., Hu, M., Lepourcelet, M., Mechoor, A., Yesilaltay, A., Krieger, M., Gray, P.A., and Shivdasani, R.A. (2006). A dynamic expression survey identifies transcription factors relevant in mouse digestive tract development. Development (Cambridge, England) *133*, 4119-4129.

25. Zhang, J., Rosenthal, A., de Sauvage, F.J., and Shivdasani, R.A. (2001). Downregulation of Hedgehog signaling is required for organogenesis of the small intestine in Xenopus. Developmental biology *229*, 188-202.

26. 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.

27. Kim, B.M., Buchner, G., Miletich, I., Sharpe, P.T., and Shivdasani, R.A. (2005). The stomach mesenchymal transcription factor Barx1 specifies gastric epithelial identity through inhibition of transient Wnt signaling. Developmental cell *8*, 611-622.

28. Okubo, T., and Hogan, B.L. (2004). Hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm. Journal of biology *3*, 11.

29. Madison, B.B., Braunstein, K., Kuizon, E., Portman, K., Qiao, X.T., and Gumucio, D.L. (2005). Epithelial hedgehog signals pattern the intestinal crypt-villus axis. Development (Cambridge, England) *132*, 279-289.

30. Kolterud, Å., Grosse, A.S., Zacharias, W.J., Walton, K.D., Kretovich, K.D., Madison, B.B., Waghray, M., Ferris, J.E., Hu, C., Merchant, J.L., et al. (2009). Paracrine Hedgehog signaling in stomach and intestine: new roles for Hedgehog in gastrointestinal patterning. In Press, Gastroenterology.

31. Allen, B.L., Tenzen, T., and McMahon, A.P. (2007). The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. Genes & development *21*, 1244-1257.

32. Eggenschwiler, J.T., Espinoza, E., and Anderson, K.V. (2001). Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. Nature *412*, 194-198.

33. Tenzen, T., Allen, B.L., Cole, F., Kang, J.S., Krauss, R.S., and McMahon, A.P. (2006). The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. Developmental cell *10*, 647-656.

34. McCarthy, R.A., Barth, J.L., Chintalapudi, M.R., Knaak, C., and Argraves, W.S. (2002). Megalin functions as an endocytic sonic hedgehog receptor. The Journal of biological chemistry *277*, 25660-25667.

35. Kozyraki, R., and Gofflot, F. (2007). Multiligand endocytosis and congenital defects: roles of cubilin, megalin and amnionless. Current pharmaceutical design *13*, 3038-3046.

36. Yan, D., Wiesmann, M., Rohan, M., Chan, V., Jefferson, A.B., Guo, L., Sakamoto, D., Caothien, R.H., Fuller, J.H., Reinhard, C., et al. (2001). Elevated expression of axin2 and hnkd mRNA provides evidence that Wnt/beta -catenin signaling is activated in human colon tumors. Proceedings of the National Academy of Sciences of the United States of America *98*, 14973-14978.

37. Moniot, B., Biau, S., Faure, S., Nielsen, C.M., Berta, P., Roberts, D.J., and de Santa Barbara, P. (2004). SOX9 specifies the pyloric sphincter epithelium through mesenchymal-epithelial signals. Development (Cambridge, England) *131*, 3795-3804.

38. Smith, D.M., Grasty, R.C., Theodosiou, N.A., Tabin, C.J., and Nascone-Yoder, N.M. (2000). Evolutionary relationships between the amphibian, avian, and mammalian stomachs. Evolution & development *2*, 348-359.

39. Mochida, Y., Parisuthiman, D., Kaku, M., Hanai, J., Sukhatme, V.P., and Yamauchi, M. (2006). Nephrocan, a novel member of the small leucine-rich repeat protein family, is an inhibitor of transforming growth factor-beta signaling. The Journal of biological chemistry *281*, 36044-36051.

40. Smith, D.M., and Tabin, C.J. (1999). BMP signalling specifies the pyloric sphincter. Nature *402*, 748-749.

41. Lees, C.W., Zacharias, W.J., Tremelling, M., Noble, C.L., Nimmo, E.R., Tenesa, A., Cornelius, J., Torkvist, L., Kao, J., Farrington, S., et al. (2008). Analysis of germline GLI1 variation implicates hedgehog signalling in the regulation of intestinal inflammatory pathways. PLoS medicine *5*, e239.

42. Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.L., Wert, S., Stahlman, M.T., and Whitsett, J.A. (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. The Journal of biological chemistry *280*, 13809-13816.

43. 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. Nature genetics *19*, 379-383.

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

45. 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. Proceedings of the National Academy of Sciences of the United States of America *101*, 266-271.

46. Li, Y., Rankin, S.A., Sinner, D., Kenny, A.P., Krieg, P.A., and Zorn, A.M. (2008). Sfrp5 coordinates foregut specification and morphogenesis by antagonizing both canonical and noncanonical Wnt11 signaling. Genes Dev *22*, 3050-3063.

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

48. Parviz, F., Matullo, C., Garrison, W.D., Savatski, L., Adamson, J.W., Ning, G., Kaestner, K.H., Rossi, J.M., Zaret, K.S., and Duncan, S.A. (2003). Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. Nature genetics *34*, 292-296.

49. Zhang, K., Shen, X., Wu, J., Sakaki, K., Saunders, T., Rutkowski, D.T., Back, S.H., and Kaufman, R.J. (2006). Endoplasmic reticulum stress activates cleavage of CREBH to induce a systemic inflammatory response. Cell *124*, 587-599.

50. Luebke-Wheeler, J., Zhang, K., Battle, M., Si-Tayeb, K., Garrison, W., Chhinder, S., Li, J., Kaufman, R.J., and Duncan, S.A. (2008). Hepatocyte nuclear factor 4alpha is implicated in endoplasmic reticulum stress-induced acute phase response by regulating expression of cyclic adenosine monophosphate responsive element binding protein H. Hepatology (Baltimore, Md *48*, 1242-1250.

51. Wu, J., and Kaufman, R.J. (2006). From acute ER stress to physiological roles of the Unfolded Protein Response. Cell death and differentiation *13*, 374-384.

52. Partington, G.A., Fuller, K., Chambers, T.J., and Pondel, M. (2004). Mitf-PU.1 interactions with the tartrate-resistant acid phosphatase gene promoter during osteoclast differentiation. Bone *34*, 237-245.

53. Tachibana, M. (2000). MITF: a stream flowing for pigment cells. Pigment cell research / sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society *13*, 230-240.

54. Nechushtan, H., and Razin, E. (2002). The function of MITF and associated proteins in mast cells. Molecular immunology *38*, 1177-1180.

55. Steingrimsson, E., Tessarollo, L., Pathak, B., Hou, L., Arnheiter, H., Copeland, N.G., and Jenkins, N.A. (2002). Mitf and Tfe3, two members of the Mitf-Tfe family of bHLH-Zip transcription factors, have important but functionally redundant roles in osteoclast development. Proceedings of the National Academy of Sciences of the United States of America *99*, 4477-4482.

56. Meinhardt, H. (2008). Models of biological pattern formation: from elementary steps to the organization of embryonic axes. Curr Top Dev Biol *81*, 1-63.

57. Charron, F., and Nemer, M. (1999). GATA transcription factors and cardiac development. Seminars in cell & developmental biology *10*, 85-91.

58. Peterkin, T., Gibson, A., and Patient, R. (2003). GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation. The EMBO journal *22*, 4260-4273.

59. Zhang, Y., Rath, N., Hannenhalli, S., Wang, Z., Cappola, T., Kimura, S., Atochina-Vasserman, E., Lu, M.M., Beers, M.F., and Morrisey, E.E. (2007). GATA and Nkx factors synergistically regulate tissue-specific gene expression and development in vivo. Development (Cambridge, England) *134*, 189-198.

60. Canning, C.A., Lee, L., Irving, C., Mason, I., and Jones, C.M. (2007). Sustained interactive Wnt and FGF signaling is required to maintain isthmic identity. Developmental biology *305*, 276-286.

61. Chi, N.C., Shaw, R.M., De Val, S., Kang, G., Jan, L.Y., Black, B.L., and Stainier, D.Y. (2008). Foxn4 directly regulates tbx2b expression and atrioventricular canal formation. Genes & development *22*, 734-739.

62. Joyner, A.L., Liu, A., and Millet, S. (2000). Otx2, Gbx2 and Fgf8 interact to position and maintain a mid-hindbrain organizer. Current opinion in cell biology *12*, 736-741.

63. Bai, C.B., Auerbach, W., Lee, J.S., Stephen, D., and Joyner, A.L. (2002). Gli2, but not Gli1, is required for initial Shh signaling and ectopic activation of the Shh pathway. Development (Cambridge, England) *129*, 4753-4761.

CHAPTER III

GATA3 IS ESSENTIAL FOR SMOOTH MUSCLE STRUCTURES AT THE PYLORUS

The smooth muscle pyloric sphincter is a critical regulator of gastroduodenal flux during digestion. Recent analysis of the pyloric transcriptome has identified a number of novel pyloric genes including *Gata3*, which encodes a zinc finger transcription factor. In the present study we report that *Gata3* is required for normal pyloric morphogenesis. We show that Gata3 and Nkx2-5 are expressed in specific smooth muscle cell populations at the pylorus, including pylorus smooth muscle proper as well as two previously undescribed, superficial muscular cords that form on the lesser curvature of the stomach. Loss of *Gata3* results in the loss of these specific muscle populations and loss of the pyloric constriction. Finally we use a reporter-marked *Gata3* allele to show that in the absence of *Gata3*, cells that co-express the reporter-marked *Gata3* locus and Nkx2-5 persist. This indicates that *Gata3* is not epistatic to *Nkx2-5* at the pylorus.

III.1. Introduction

The pylorus is the anatomical junction between the stomach and the small intestine, and its muscular sphincter helps to regulate the passage of food between these organs during

digestion. Similar to other organs of the gastrointestinal tract, the pylorus has a distinctive histology. The pyloric lumen is lined by endoderm-derived mucosal epithelium, and it is here that a tight, one-cell-thick boundary between intestinal and gastric epithelial cells (i.e., the epithelial pyloric border) is established during late embryogenesis [1]. The mucosal epithelium is surrounded by several layers of mesoderm-derived tissue. The loose sub-mucosal layer directly supports the overlying epithelium and contains nervous tissue as well as a number of mesenchymal cell types (i.e., myofibroblasts, endothelial cells, lymphatic cells, etc). Concentric inner circular and outer longitudinal layers of smooth muscle cells ensheathe the sub-mucosa, providing the main structural support for the pylorus.

The orientation of smooth muscle cells in the outer longitudinal layer is parallel to the long axis of the lumen, while the cells of the inner circular layer are oriented circumferentially. Throughout most of the gastrointestinal tract, these muscle layers are mostly (if not entirely) distinct. At the pylorus, however, the inner circular and outer longitudinal layers coalesce to form the pyloric sphincter, a complex muscle structure that controls the lumenal size of the gastric outlet by integrating numerous nervous innervations and local hormonal feedback [2-5]. In the common human congenital pathology infantile hypertrophic pyloric stenosis (IHPS), both the structure and function of the pyloric sphincter are abnormal [5]. Yet despite the significant incidence of IHPS (approximately three per 1,000 live births), the pathology of the disease is not completely understood. A better understanding of the events involved in the development of the pyloric region could shed light on the molecular etiology involved.

The pylorus forms at a precise point between the stomach and the duodenum. Anteroposterior patterning events are directed by extracellular signaling molecules in combination with the "gastrointestinal Hox code". Together these integrated molecular programs establish broad organ domains, each of which display unique endodermal and mesodermal gene expression patterns [6, 7]. For example by embryonic day (E) 10.5 three members of the Nkx family of homeodomain proteins (Nkx2-3, Nkx2-5, and Nkx3-2/Bapx1) subdivide the mesoderm of the distal foregut. *Nkx2-3* and *Bapx1* mark the proximal small intestine and distal stomach, respectively, while *Nkx2-5* is expressed in a narrow domain at the pylorus [8, 9].

Experiments in the chick embryo have established a basic genetic network for pylorus development [10-14]. Hh signaling from the intestinal endoderm activates *Bmp4* expression in the underlying mesenchyme. Secreted Bmp4 then independently activates *Nkx2-5* and *Sox9* in pyloric mesenchyme. Subsequently Sox9, but not Nkx2-5, upregulates the Bmp signaling modulator *gremlin*, which patterns the overlying pyloric endoderm. Germline loss of *Bapx1* or *Six2* is each associated with significant pyloric sphincter defects in the mouse [15, 16].

Recent transcriptional profiling of E14.5 and E16.5 pyloric tissues confirmed a number of these pylorus-specific genes in mouse, including *Nkx2-5* and *gremlin*, and identified several novel pyloric genes, such as *Gata3* [1]. *Gata3* is of particular interest because it encodes a zinc finger transcription factor and is expressed in a narrow domain at the

pylorus, similar to *Nkx2-5*. In the present study we compare the spatiotemporal and cell type-specific expression of Gata3 and Nkx2-5 at the pylorus and show that both mark subdomains of smooth muscle. Both factors are expressed in superficial, cord-like structures on the ventral stomach surface. We also demonstrate that the loss of *Gata3* perturbs pyloric morphogenesis and results in the loss of specific smooth muscle structures, including portions of the outer longitudinal layer and the ventral pyloric cords.

III.2. Materials and methods

Mice

The generation of *Gata3*$^{lacZ/+}$ and *Nkx2-5*$^{lacZ/+}$ mice have been described previously [17-19]. *Gata3*$^{lacZ/+}$ intercrosses result in wild type (*Gata3*$^{+/+}$), *Gata3* heterozygous (*Gata3*$^{lacZ/+}$), and *Gata3* null (*Gata3*$^{lacZ/lacZ}$) genotypes. *Gata3* null embryos die at E11 of sympathoadrenal insufficiency, but can be rescued by administration of a pharmacological cocktail containing α- and β-adrenergic agonists to the pregnant dam, as previously described [20, 21]. Rescue solution was prepared fresh: 15 mg each of isoproterenol (Sigma I-5627) and phenylephrine (Sigma P-6126) was added to 50 mL of ddH_20 and supplemented with 100 mg of ascorbic acid and 2 g of sucrose. This solution (which supplanted normal drinking water) was administered once daily to timed pregnant dams, beginning at E7.5. Rescued embryos survived until birth, but died shortly thereafter, likely due to palatal defects and failure to suckle. Although Gata3 null embryos can be identified phenotypically, all genotypes were confirmed by RT-PCR.

Antibodies

Primary antibodies used were: Cy3-conjugated mouse monoclonal to αSMA (1:500, Sigma C6198), rabbit polyclonal to desmin (1:500 Abcam ab8952), goat polyclonal to Nkx2-5 (1:100, Santa Cruz sc-8697), mouse monoclonal to Gata3 (1:100, Santa Cruz sc-268), and chicken polyclonal to β-gal (1:1000, Abcam ab9361). Secondary antibodies used were: biotinylated horse anti-universal (mouse/rabbit/goat) IgG (Vector BA-1400), biotinylated rabbit anti-chicken IgY (1:200, Abcam ab6752-1500), and Alexa Fluor 488 donkey anti-rabbit IgG (1:500 Invitrogen A-21206). Tertiary antibodies used were: Cy3/Cy5-conjugated mouse monoclonal anti-biotin (1:100, Jackson Immunoresearch 200-162-211/200-172-211).

Whole mount X-gal staining

X-gal staining was performed as described previously [22]. Staged embryos from intercrosses of *Gata3*$^{lacZ/+}$ mice, *Gata3*$^{lacZ/+}$ and wild type (C57BL/6J), or *Nkx2-5*$^{lacZ/+}$ and wild type mice were dissected in 1X PBS. The gastrointestinal tract was fixed with 4% PFA for 10 minutes at 4°C, and washed in 1X PBS. X-gal staining solution was prepared fresh: 1X PBS (pH 7.5), 2 mM magnesium chloride, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, and 1 mg/mL X-gal (from 20 mg/ml stock in DMF) were combined. Fixed tissue was incubated in staining solution at 37°C and monitored periodically to control staining intensity. Stained tissue was washed in 1X PBS and then post-fixed overnight at 4°C in 4% PFA. Tissue was cleared with increasing concentrations of glycerol in 1X PBS.

Whole mount immunofluorescence

Dissected foreguts from adult mice were fixed overnight in 4% PFA. Samples were cut in half with microsurgical scissors and embedded in 2% low melt agarose. While still in agarose, tissues were washed and stained in a 1X PBS blocking solution containing 0.1% Triton-X and 10% BSA. The first primary antibody staining (rabbit polyclonal to desmin) was done overnight at 4°C. Tissues were washed with blocking solution and stained with the secondary antibody (Alexa Fluor 488 donkey anti-rabbit IgG) at room temperature (RT) for 2 hours. Tissues were washed with blocking solution and stained with the second primary antibody (Cy3-conjugated mouse monoclonal to αSMA) at RT for 2 hours. Tissues were washed in 1XPBS prior to microscopy.

Histology

Staged embryos were dissected in 1X PBS. Gastrointestinal tracts were fixed overnight at 4°C in 4% PFA, embedded in paraffin, cut into 5-10 μm sections and adhered to Superfrost Plus glass slides (Fisher 22-034-979).

Hematoxylin and eosin staining

For hematoxylin and eosin (H&E) staining standard laboratory methods were used. Sections were deparaffinized in xylene, rehydrated through decreasing alcohol concentrations, stained with hematoxylin, incubated in bluing solution, counterstained in eosin, dehydrated through increasing alcohol concentrations, and equilibrated with xylene. Glass coverslips were mounted with Permount (Fisher SP15-100).

Immunofluorescence

Sections were deparaffinized in xylene, rehydrated through decreasing alcohol concentrations, and boiled for 10 minutes in 10 mM sodium citrate, pH 6.0. Slides were allowed to slowly cool down and then washed with 1X PBS. Sections were pre-blocked sequentially using the Avidin/Biotin Blocking Kit (Vector SP-2001) and the MOM Kit (Vector FMK-2201). The sections were then blocked in 10% animal serum/0.03% Triton X-100 in 1X PBS for 1 hour at RT. Primary antibodies were diluted in blocking solution prior to incubation on slides. Slides were washed in 1X PBS prior to incubation with appropriate secondary antibodies and DAPI (1:100). After 1X PBS washes, glass coverslips were mounted with ProLong Gold Antifade Reagent. See Table 1 for details of specific immunostaining experiments.

Microscopy

Whole mount tissue was photographed in 1X PBS or 80% glycerol (in 1X PBS) on a Leica dissecting microscope with a Leica CCD camera. Slides were photographed on a Nikon E800 microscope with a Nikon CCD camera.

III.3. Results

Gata3 expression in the developing foregut

Previously we reported a novel pyloric expression domain for *Gata3* at E14.5 using *in situ* hybridization [1]. To better characterize the spatiotemporal pattern of pyloric *Gata3* during embryogenesis, we analyzed $Gata3^{lacZ/+}$ embryos by whole mount X-gal staining.

At E10.5 there is robust *Gata3* expression in several tissues but no discrete staining in the foregut, while at E11.5 weak *Gata3* expression is occasionally seen in the foregut region (Figures III.1A-B). By E12.5 an obvious band of *Gata3*-expressing cells appears in the distal stomach near the gastroduodenal junction (Figure III.1C). This band forms a nearly circumferential ring around the gut with a small gap on the lesser curvature (ventral side) of the stomach. A day later at E13.5 the *Gata3* expression domain is wider, and two small, cellular prominences are discernible on the ventral side of the band at either end of the gap (Figure III.1D). These clusters of *Gata3*-expressing cells lie along the lesser curvature of the stomach and extend slightly anteriorly towards the esophagus. Pyloric *Gata3* expression continues to expand at E14.5, particularly on the greater curvature (dorsal side) of the stomach, where a small, distinct nodule of *Gata3*-expressing cells protrudes from the band (Figure III.1E). On the ventral aspect, the cellular prominences first observed a day earlier have coalesced into thin, bilateral cellular cords that extend anteriorly from the pylorus to the esophagus along the lesser curvature of the stomach. At E14.5, if viewed anteriorly from the duodenum, the pyloric *Gata3* expression domain resembles a horseshoe (Figure III.1E). Between E15.5 and E18.5 there are relatively few significant changes in this pattern. The dorsal nodule disappears by E16.5 (data not shown), and the ventral cords lengthen, such that by E18.5 they extend to the level of the gastroesophageal junction (Figure III.1F).

To confirm and extend these observations, we analyzed *Gata3* protein expression in sectioned wild type tissue by immunostaining. At E14.5 and E18.5 nuclear Gata3 signal is detected in a narrow domain of mesenchymal cells at the gastroduodenal junction

(Figure III.2). Gata3-expressing cells are positioned within the outer mesenchymal layer of the gut but not the loose layer of sub-mucosal mesenchyme. Consistent with our whole mount X-gal staining results, at E14.5 Gata3-positive cells were found in the dorsal nodule as well as the ventral cords, while at E18.5 the dorsal nodule was absent and Gata3 was only expressed in the pyloric mesoderm and ventral cords. No epithelial staining was observed at either time point.

Nkx2-5 is expressed in a similar domain to Gata3 in the foregut

Nkx2-5 is a known pyloric gene in a number of vertebrate species, including the mouse [9]. The expression pattern of *Nkx2-5* in early foregut mesenchyme and spleen morphogenesis was recently described, but the embryonic expression of *Nkx2-5* at the pylorus has not been fully explored [23]. To examine the spatiotemporal pattern of pyloric *Nkx2-5* expression, we analyzed *Nkx2-5*$^{lacZ/+}$ embryos by whole mount X-gal staining [17]. *Nkx2-5* is expressed in a discrete domain in the foregut at E10.5 and E11.5, 48 hours earlier than *Gata3* expression is detected (Figures III.3A-B). Between E12.5 and E13.5 *Nkx2-5*-positive, splenic precursor cells (SPCs) migrate anteriorly from the pyloric mesenchyme and through the mesogastrium along the greater curvature of the stomach to populate the developing spleen [23]. *Gata3* is not expressed in this migrating splenic population. By E14.5 the migration of SPCs ceases and *Nkx2-5* is expressed in a band at the pylorus (Figure III.3C). Like *Gata3, Nkx2-5* is expressed in the dorsal nodule and ventral cords at E14.5, and between E15.5 and E18.5 there are few changes in the *Nkx2-5* pyloric expression pattern (Figure III.3D). At all time points, however, the *Nkx2-5*

domain is apparently broader than the corresponding *Gata3* domain, and in contrast to *Gata3,* the *Nkx2-5* band forms a complete circumferential ring around the gut.

Nkx2-5 protein expression was next examined in sectioned wild type tissue by immunostaining. At E14.5 and E18.5, nuclear Nkx2-5 is detected in a narrow domain of mesenchymal cells at the gastroduodenal junction (Figure III.4). This domain is broader than the corresponding Gata3 domain. Nkx2-5-expressing cells are positioned within the inner and outer smooth muscle layers of the gut, but unlike Gata3, Nkx2-5-positive cells are also present in the loose layer of sub-mucosal mesenchyme. Consistent with our whole mount X-gal staining results, Nkx2-5-positive cells were found at E14.5 in the dorsal nodule, as well as the ventral cords, while at E18.5 the dorsal nodule was absent and Nkx2-5 was only expressed in the ventral cords. [At E14.5 some residual migrating SPCs were also detected in the mesogastrium attached to the pylorus (Figure III.4A).] No epithelial staining was observed at either time point.

Gata3 and Nkx2-5 are expressed in smooth muscle cell populations at the pylorus

We analyzed Nkx2-5 co-expression with two molecular markers of smooth muscle [i.e., actin, alpha 2, smooth muscle, aorta (Acta2/αSMA) and desmin (Des)] by immunostaining. Co-staining with these markers identifies four possible cell populations: non-smooth muscle ($\alpha SMA^{neg}/desmin^{neg}$), myofibroblast ($\alpha SMA^{+}/desmin^{neg}$), smooth muscle precursor ($\alpha SMA^{neg}/desmin^{+}$), and differentiated smooth muscle cells ($\alpha SMA^{+}/desmin^{+}$). At E18.5 the majority of Nkx2-5-positive cells are $\alpha SMA^{+}/desmin^{+}$, while a subset are $\alpha SMA^{+}/desmin^{neg}$ (Figures III.5A-B).

The Gata3 expression domain also appeared to correspond to the location of smooth muscle cell populations at the pylorus. To confirm this we analyzed Gata3 co-expression with desmin, a molecular marker of differentiated smooth muscle cells as well as smooth muscle precursors. At E16.5 the vast majority of Gata3-positive cells throughout the pylorus are also desmin-positive (Figures III.5C-D). Thus both Nkx2-5 and Gata3 are expressed in a subset of smooth muscle cells at the pylorus.

The ventral pyloric cords are a novel smooth muscle structure

To our knowledge the bilateral, *Gata3*- and *Nkx2-5*-positive ventral cords described above are novel structures. We examined these cords in whole mount and sectioned wild type tissue using markers for smooth muscle populations (αSMA and desmin). At E18.5, the majority of cells in the cords are $\alpha SMA^{+}/desmin^{+}$, indicating that they are composed of differentiated smooth muscle (Figures III.6A-B). At the anterior tip of the cord near the esophagus there is a diffuse meshwork of $\alpha SMA^{neg}/desmin^{+}$ smooth muscle precursor cells. The fact that the cords apparently lengthen in a posteroanterior direction, from the pylorus towards the esophagus, suggests that this $\alpha SMA^{neg}/desmin^{+}$ cell population may be a source of precursors for the developing smooth muscle cells within the cord. In whole mount specimens stained at postnatal day (P) 30, the ventral cords are readily observable, indicating that they persist after birth (Figure III.6C). Similar to E18.5 cells within the cords are $\alpha SMA^{+}/desmin^{+}$, but at the anterior tip a subset of cells near the esophagus are $\alpha SMA^{neg}/desmin^{+}$. Filamentous structures originating from $\alpha SMA^{+}/desmin^{+}$ cells extend from the cords into this population of $\alpha SMA^{neg}/desmin^{+}$

cells. At the posterior (pyloric) end the differentiated smooth muscle cells of the cords integrate into the outer smooth muscle layer of the gut (data not shown). These ventral cords are a novel smooth muscle structure of the pylorus, although whether they contribute to the physiological regulation of pyloric function is not yet clear.

Gata3 is required for pylorus morphogenesis

To determine the requirement for *Gata3* during pylorus development we intercrossed *Gata3*$^{lacZ/+}$ mice. We compared the structure of the pylorus in wild type, *Gata3* heterozygous, and *Gata3* null embryos in whole mount and H&E-stained, sectioned material. In whole mount E18.5 Gata3 null embryos, the gastroduodenal angle is altered, and the pyloric constriction is lost (Figures III.7A-B). These characteristics are confirmed in H&E-stained, sectioned samples. The gastroduodenal angle of *Gata3* null embryos is more acute than wild type littermates, and the diameter of the inner ring of the pyloric sphincter, measured as the distance between inner smooth muscle layers on opposite sides of the gut, is wider in *Gata3* null embryos (Figures III.7C-D). In addition, the outer longitudinal smooth muscle at the pylorus appears perturbed (Figures III.7E-F). These results indicate that *Gata3* is required for proper morphogenesis of the pylorus and suggest that *Gata3* may directly or indirectly effect pyloric mesenchymal cell populations, including smooth muscle.

Loss of Gata3 is associated with the absence of specific smooth muscle structures at the pylorus

To correlate the pyloric phenotype of *Gata3* null embryos with the spatiotemporal pattern of *Gata3* expression, we analyzed intercross matings of *Gata3*$^{lacZ/+}$ mice by whole mount

X-gal staining. Compared to *Gata3* heterozygous littermates at E16.5, *Gata3* null mice displayed obvious differences in the pattern of X-gal staining (Figures III.8A-B). In *Gata3* null mice, the band of X-gal staining around the pylorus is more diffuse and extends into associated pancreatic tissue on the dorsal side of the stomach. The ventral cords are perturbed but still present. At day later at E17.5, however, there is complete absence of the ventral cords (Figures III.8C-D). Instead the ventral side of the stomach contains a small prominence of X-gal positive cells. This indicates that Gata3 is required for the maintenance of the ventral pyloric cords.

To investigate the function of *Gata3* in smooth muscle cell populations at the pylorus, we analyzed wild type and *Gata3* null embryos by immunostaining of sectioned tissue for αSMA and desmin. Compared to wild type littermates at E18.5, the morphology of the inner circular smooth muscle and loose sub-mucosal layers in *Gata3* null embryos are similar, despite the obvious difference in constriction of the pyloric smooth muscle sphincter (Figure III.9). In wild type and *Gata3* null tissue, $aSMA^{+}/desmin^{+}$ differentiated smooth muscle cells populate the inner smooth muscle layer and $aSMA^{+}/desmin^{neg}$ myofibroblasts are present in the sub-mucosal layer. Even so, there are subtle differences in the shape of the inner smooth muscle layer, with loss of the knob-like shape of this muscle at the pylorus. Additional changes are apparent in the outer smooth muscle layer on the dorsal side of the stomach. Whereas wild type tissue has a distinct outer smooth muscle layer at the pylorus, *Gata3* null tissue does not (Figures III.9B and D). In contrast, the thin outer layer of longitudinal smooth muscle persists in both the duodenum and the

stomach of Gata3 null mice. Thus, *Gata3* is required for the persistence of smooth muscle cells in the outer longitudinal layer of the pyloric sphincter.

Gata3 and Nkx2-5 are co-expressed but Gata3 is not epistatic to Nkx2-5 at the pylorus

Our immunostaining experiments indicate that Gata3 and Nkx2-5 are both expressed in a subset of smooth muscle cells at the pylorus (see Figure III.5). To test whether *Gata3* is required for *Nkx2-5* expression, we intercrossed *Gata3*$^{lacZ/+}$ mice and analyzed *Gata3* heterozygous (*Gata3*$^{lacZ/+}$) and *Gata3* (*Gata3*$^{lacZ/lacZ}$) null progeny by immunostaining sectioned tissue for Nkx2-5 and β-gal (the protein encoded by *lacZ*). In *Gata3* heterozygotes, cells that express β-gal also express Nkx2-5 (Figures III.10A and C). This is strong indirect evidence for co-expression of Gata3 and Nkx2-5 at the pylorus. Cells expressing β-gal were reduced in number but were still present in *Gata3* null embryos, indicating that at least a subset of the cells that express *Gata3* transcripts do not require *Gata3* protein for survival (Figures III.10B and D). Additionally, the vast majority of these β-gal-expressing, *Gata3* null cells also express Nkx2-5, demonstrating that at least in this surviving cell population, *Gata3* is not required for Nkx2-5 expression at the pylorus.

III.4. Discussion

In this study, we examined the role of *Gata3* during the development of the mouse pylorus. We report that germline loss of *Gata3* leads to pyloric dysmorphogenesis: *Gata3* null embryos have an altered gastroduodenal angle and reduced pyloric constriction. This

phenotype suggests a defect in the development of the pyloric sphincter, and we show that loss of *Gata3* is correlated with the absence of specific smooth muscle structures at the pylorus. At E18.5 Gata3-positive cells are found in the inner circular and outer longitudinal smooth muscle layers, and many of these cells co-express the smooth muscle marker desmin. In addition Gata3-positive cells form bilateral cords that originate at the pylorus and lie along the lesser curvature of the stomach; these ventral pyloric cords co-express markers of differentiated smooth muscle. In *Gata3* null embryos the ventral pyloric cords and the outer longitudinal smooth muscle layer of the pylorus are absent. Thus we propose that the loss of these specific smooth muscle structures results in the pyloric phenotype of *Gata3* null embryos.

Interestingly, despite the fact that Gata3 is apparently expressed in both inner circular and outer longitudinal smooth muscle layers of the pylorus, only the outer longitudinal layer is lost in *Gata3* null embryos. Careful analysis shows that while the pyloric muscle proper appears to be an extension of the inner circular muscle, the bulk of the muscle at the pylorus is actually contributed by the outer longitudinal muscle. In both the proximal small intestine and distal stomach, the outer longitudinal layer is quite thin, especially when compared to the inner circular layer. At the pylorus, however, within the narrow domain of *Gata3* and *Nkx2-5* expression, the outer smooth muscle layer is almost as thick as the inner circular layer. In addition, many of the cells in this hyperplastic region are no longer oriented parallel to the long axis of the gut, but rather are oblique or even perpendicular (data not shown). These observations are supported by anatomical findings from guinea pigs and cadaveric human pylori, which reveal hyperplasia of the outer

longitudinal layer and fascicular connections to the inner circular layer [2, 4]. Whether *Gata3* and/or *Nkx2-5* direct this hyperplasia of the outer longitudinal muscle at the pylorus remains to be tested. Certainly the absence of this thickened muscle in Gata3 animals demonstrates that Gata3 is critical for the survival of its component smooth muscle cell population. The deterioration of the ventral pyloric cords in Gata3 null embryos is further evidence that Gata3 is required for the survival of these smooth muscle populations. A survival role for Gata3 is not without precedent as Gata3 is also required for the survival of adrenal chromaffin cells [24].

The pattern of pyloric expression of *Nkx2-5* is very similar to that of *Gata3*. However, *Nkx2-5* expression begins earlier and its domain is broader, which suggests that *Nkx2-5* may be epistatic to *Gata3* at the pylorus. Germline *Nkx2-5* null embryos are embryonic lethal by E10, which precludes further analysis of its role in pylorus development or a complete assessment of epistasis between *Nkx2-5* and *Gata3* [25]. We show here that Nkx2-5 is present in cells that express transcripts from the *Gata3* locus; thus it is possible that Nkx2-5 regulates *Gata3*. It is clear that *Gata3* is not epistatic to *Nkx2-5*. Nkx and Gata family factors interact in a number of other tissues including cardiac muscle, where they directly regulate each other's transcription and/or cooperatively activate gene expression [26-36]. For example, a translocation in human T-cell acute lymphoblastic leukemia (T-ALL) results in co-expression of Nkx2-5 and Gata3. In these leukemic cells, Nkx2-5 and Gata3 physically interact and may cooperatively regulate gene expression [37]. Thus at least in some cellular contexts Nkx2-5 and Gata3 may interact to control gene transcription.

Gata3 null animals lose the pyloric constriction, likely due to smooth muscle changes. Similar pyloric phenotypes are reported for germline loss of either *Bapx1* or *Six2*. These genes encode posterior stomach transcription factors that regulate *Bmp4* expression. Misexpression of *Bapx1* in the anterior glandular stomach (proventriculus) of chicken embryos inhibits *Bmp4*, while germline loss of *Six2* de-represses Bmp4 in the posterior glandular stomach [8, 15]. The expression pattern of *Gata3* (i.e., in the Bmp-responsive domain of pyloric mesenchyme) suggests that, similar to *Nkx2-5*, it could potentially be a target of Bmp signaling. Alternatively, as discussed above Nkx2-5 may regulate *Gata3*. In such a scenario the pyloric phenotype of *Bapx1* and *Six2* null embryos could potentially be due to changes in *Gata3* expression, downstream of Bmp4. Pyloric *Nkx2-5* expression is altered but not extinguished in either *Bapx1* or *Six2* null embryos [15, 16]. In follow-up studies it will be important to explore the potential relationship between *Gata3* and Bmp signaling, *Six2*, and *Bapx1*, so that Gata3 can be properly placed within the genetic network of pylorus development.

To our knowledge, this study is the first to report the existence of the ventral pyloric cords. While it is unclear whether these cords are present in humans, they appear analogous in structure and development, although not position, to the Ligament of Treitz. Also known as the suspensory muscle of the duodenum, the Ligament of Treitz contains a thin smooth muscle that connects the duodenum at the duodenojejunal flexure to the abdominal aorta between the celiac and superior mesenteric arteries [38]. At the duodenojejunal flexure the muscle of the Ligament of Treitz integrates into both the outer

longitudinal and inner circular smooth muscle layers; at the other end it connects to the aorta via dense fibers of connective tissue. Contraction of the muscle of the Ligament of Treitz alters the angle of the duodenojejunal flexure, facilitating passage of food from the duodenum to jejunum [39]. The nascent Ligament of Treitz first appears in early fetal development as a band of loose connective tissue (i.e., superior retention band). Later, smooth muscle cells populate the band and form the muscle of the Ligament of Treitz [38, 40]. The similarities between the ventral pyloric cords and the Ligament of Treitz are striking and suggest that the function of the cords may be to regulate the tone of the pyloric sphincter. Follow-up studies will be necessary to more specifically examine this possibility *in vivo*.

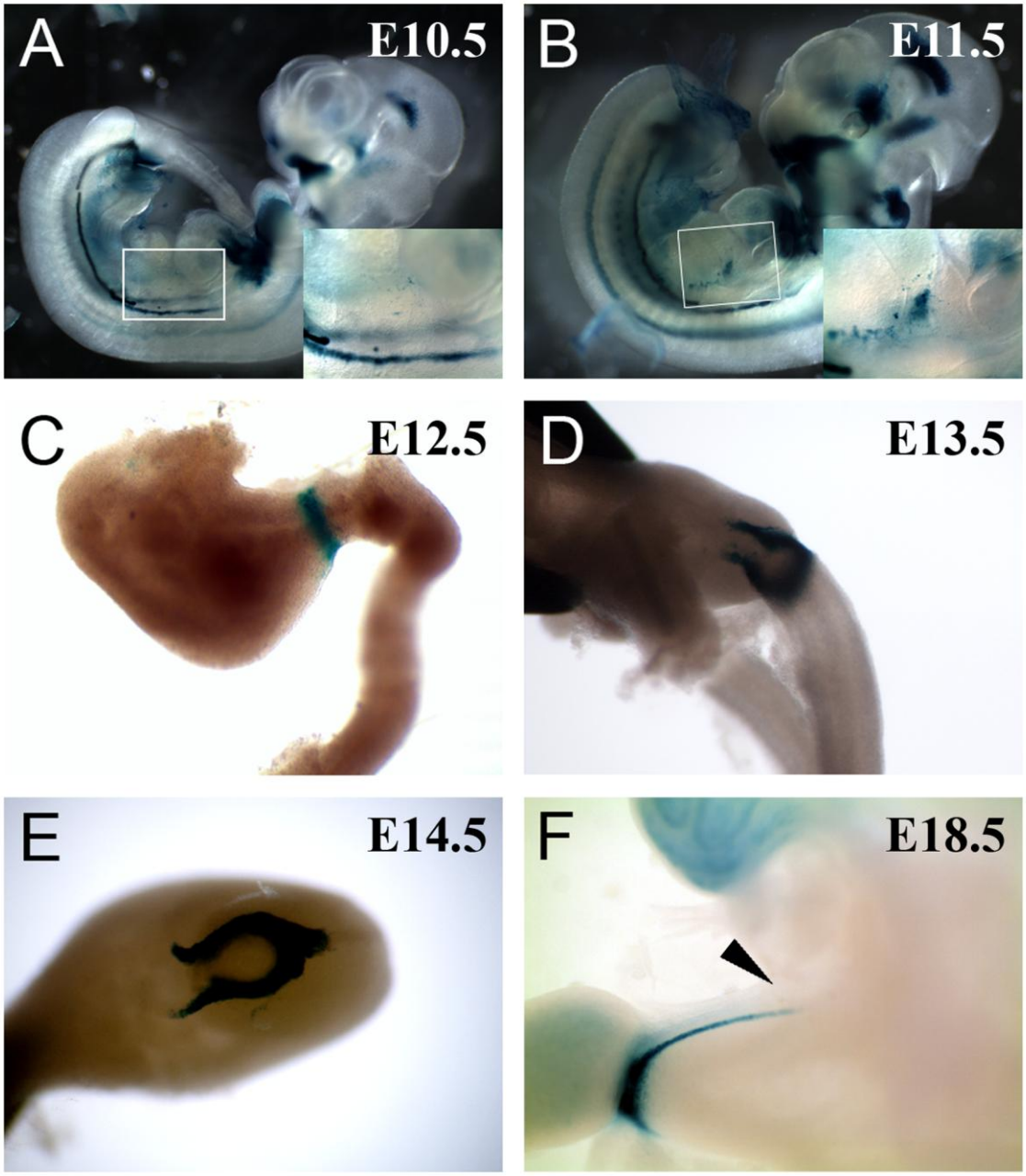

Figure III.1. Embryonic expression of Gata3 in the posterior foregut. Whole mount X-gal staining of *Gata3*$^{lacZ/+}$ embryos or foreguts between E10.5 and E18.5: A) E10.5, B) E11.5, C) E12.5, D) E13.5, E) E14.5, and F) E18.5. Blue X-gal staining represents *Gata3* expression. The white boxes in A) and B) roughly correspond to the area of high magnification in the inset. Notice that the circumferential band of X-gal staining is incomplete in D) and E). In C) and E) the stomach is on the left and right, respectively. D) is a ventral view of the lesser curvature of the stomach. E) is an anterior view of the pyloric lumen from the position of the duodenum. Note than in F) the cords of X-gal staining extend anteriorly to the esophagus (black arrowhead). X-gal staining in F) by Mr. Ajay Prakash.

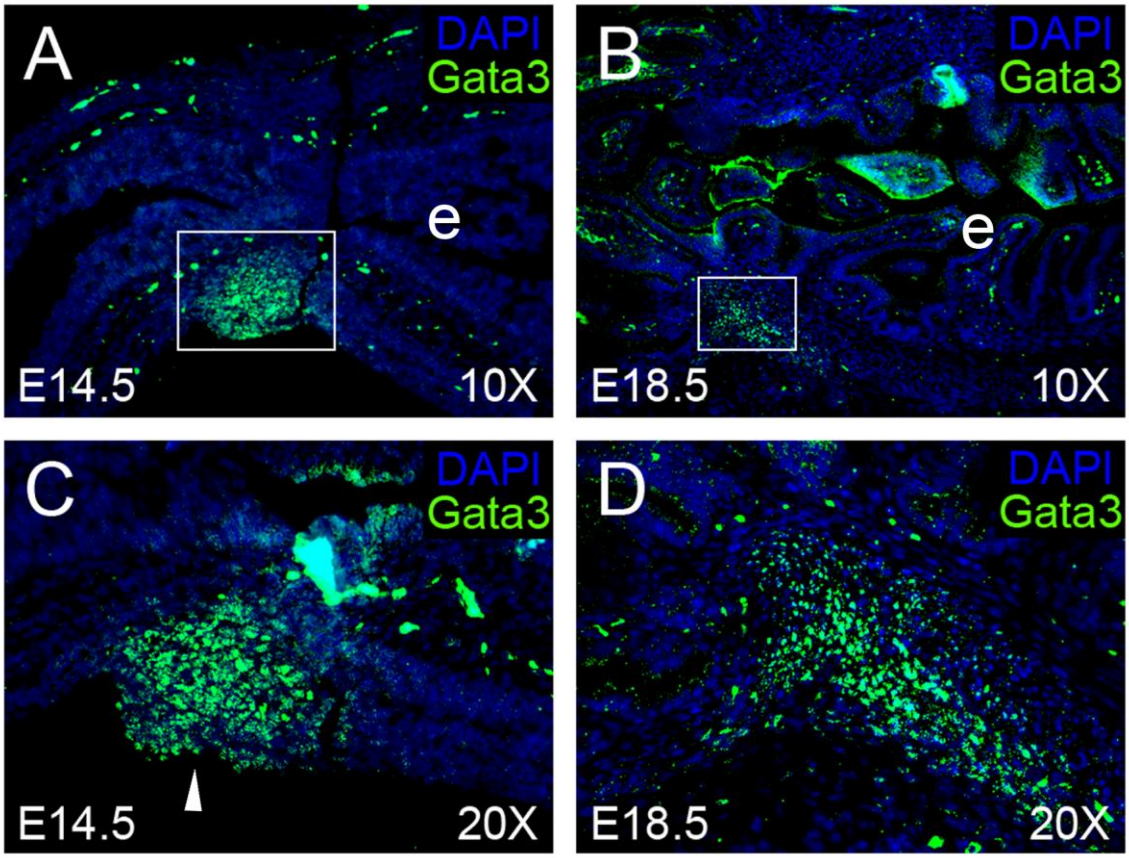

Figure III.2. Gata3 is expressed in the outer mesenchymal layer at the pylorus. Immunofluorescence for Gata3 in sectioned wild type foregut at E14.5 and E18.5: A) E14.5 (10X), B) E18.5 (10X), C) E14.5 (20X), and D) E18.5 (20X). Gata3 staining is in green, while nuclei are counterstained blue with DAPI. Nuclear Gata3 staining is cerulean. Non-nuclear green staining in the epithelium and mesenchyme is background. The white boxes in A) and B) roughly correspond to the area of higher magnification in C) and D), respectively. In all panels, stomach is on the right and intestine is on the left. Notice that there is no Gata3 staining in epithelium (e) or the loose sub-mucosal mesenchyme. The bulge of Gata3-positive cells in C) is the dorsal nodule (white arrowhead). Immunostaining by Mr. Ajay Prakash.

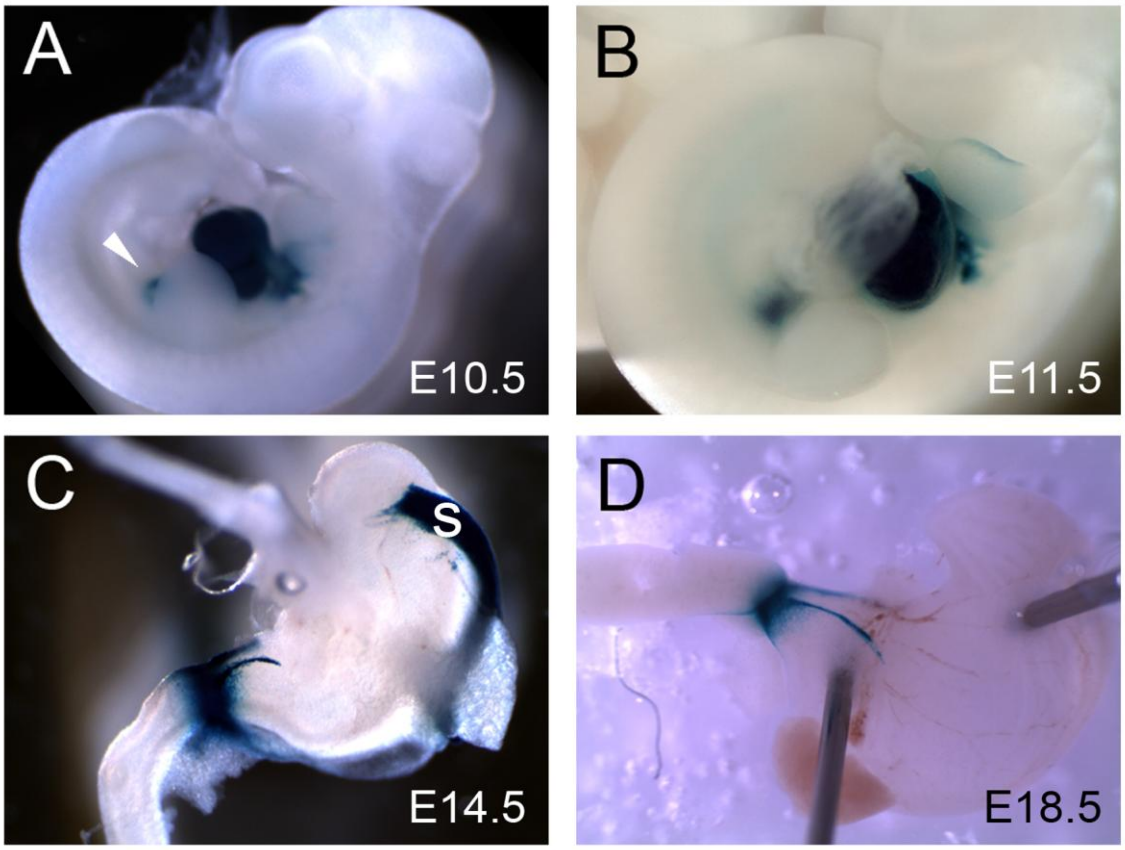

Figure III.3. Embryonic expression of Nkx2-5 in the posterior foregut. Whole mount X-gal staining of *Nkx2-5*$^{lacZ/+}$ embryos or foreguts between E10.5 and E18.5: A) E10.5, B) E11.5, C) E14.5, and D) E18.5. Blue X-gal staining represents *Nkx2-5* expression. Notice the discrete domain of X-gal staining in A) (white arrowhead) and B). Also note the complete circumferential band of X-gal staining in C) and D) and the X-gal stained spleen (s) in C). Note that the X-gal stained ventral cords extend towards the esophagus in C) and D). In those panels the stomach is on the right and the intestine is on the left. X-gal staining in D) by Mr. Ajay Prakash.

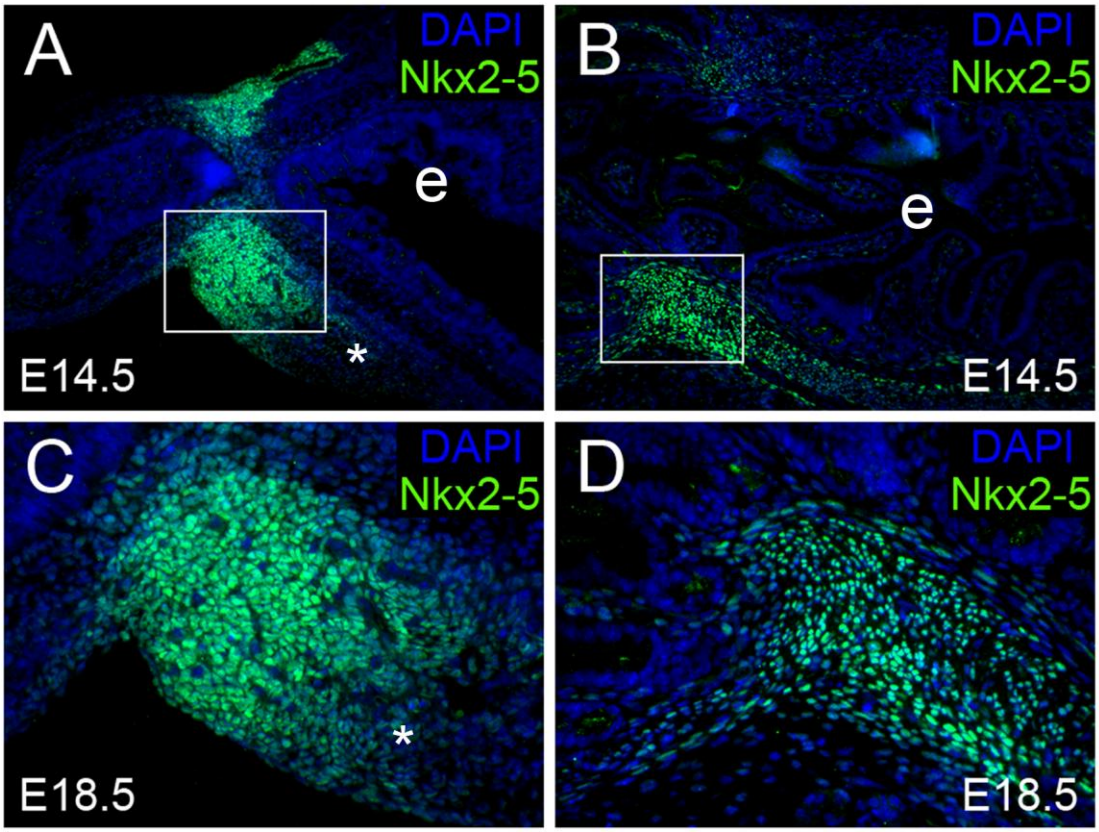

Figure III.4. Nkx2-5 is expressed broadly in the mesenchyme at the pylorus.
Immunofluorescence for Nkx2-5 in sectioned wild type foregut at E14.5 and E18.5: A) E14.5 (10X), B) E18.5 (10X), C) E14.5 (20X), and D) E18.5 (20X). Nkx2-5 staining is in green, while nuclei are counterstained blue with DAPI. Nuclear Nkx2-5 staining is cerulean. Non-nuclear green staining in the epithelium and mesenchyme is background. The white boxes in A) and B) roughly correspond to the area of higher magnification in C) and D), respectively. Notice that there is no Nkx2-5 staining in epithelium (e). The mesogastrium is denoted by the white asterisks in A) and C). In all panels, stomach is on the right and intestine is on the left. Immunostaining by Mr. Ajay Prakash.

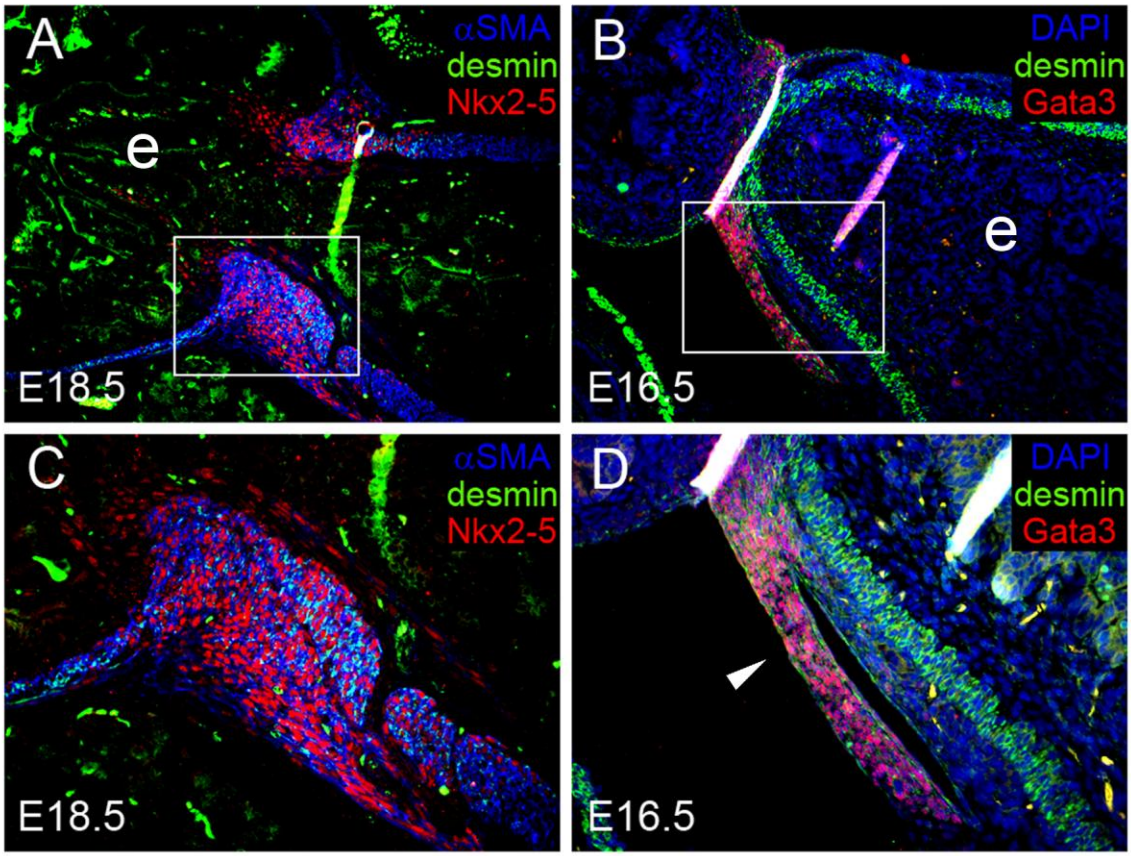

Figure III.5. Nkx2-5 and Gata3 are expressed in smooth muscle cell populations at the pylorus. Immunofluorescence for Nkx2-5, desmin, and αSMA or Gata3 and desmin in sectioned wild type foregut at E18.5 and E16.5, respectively: A) Nkx2-5, desmin, and αSMA at E18.5 (10X), B) Gata3 and desmin at E16.5 (10X), C) Nkx2-5, desmin, and αSMA at E18.5 (20X), and D) Gata3 and desmin at E16.5 (20X). In A) and C) red is Nkx2-5, blue is αSMA, and green is desmin. Blue staining is myofibroblasts, green staining is smooth muscle precursor cells, and cerulean staining is differentiated smooth muscle. Green staining in the epithelium (e) is background. Notice in C) that red nuclei co-localize with cerulean staining in the smooth muscle layers and red staining in the loose sub-mucosal mesenchyme. In B) and D), red is Gata3, green is desmin, and the nuclei are counterstained blue with DAPI. Magenta staining is nuclear Gata3 and green staining is smooth muscle precursors and differentiated smooth muscle. Notice that magenta nuclei co-localize with green staining in the smooth muscle layers at the pylorus. The white boxes in A) and B) roughly correspond to the area of higher magnification in C) and D), respectively. In all panels, stomach is on the right and intestine is on the left. The ventral cord is apparent is B) and D) (white arrowhead). Immunostaining by Mr. Ajay Prakash.

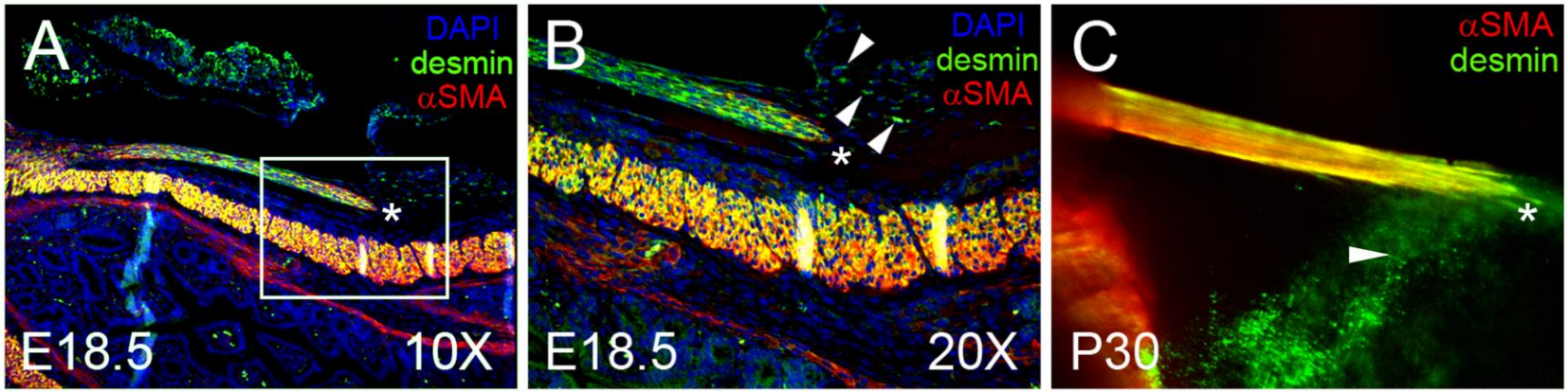

Figure III.6. The ventral pyloric cords are smooth muscle structures. Immunofluorescence for desmin and αSMA in sectioned and whole mount wild type tissue at E18.5 and P30, respectively: A) E18.5 (10X), B) E18.5 (20X), and C) P30. In A) and B) green is desmin, red is αSMA, and nuclei are counterstained in blue with DAPI. In C) green is desmin and red is αSMA. Red staining is myofibroblasts, green staining is smooth muscle precursor cells, and yellow staining is differentiated smooth muscle. Notice that nuclei in the ventral pyloric cord co-localize with yellow staining. The anterior tip of the cord is denoted by an asterisk in all panels. Notice that nuclei in loose tissue just outside of the tip co-localizes with green staining in B) and C) (white arrowheads). The white box in A) roughly corresponds to the area of higher magnification in B). Immunostaining by Mr. Ajay Prakash.

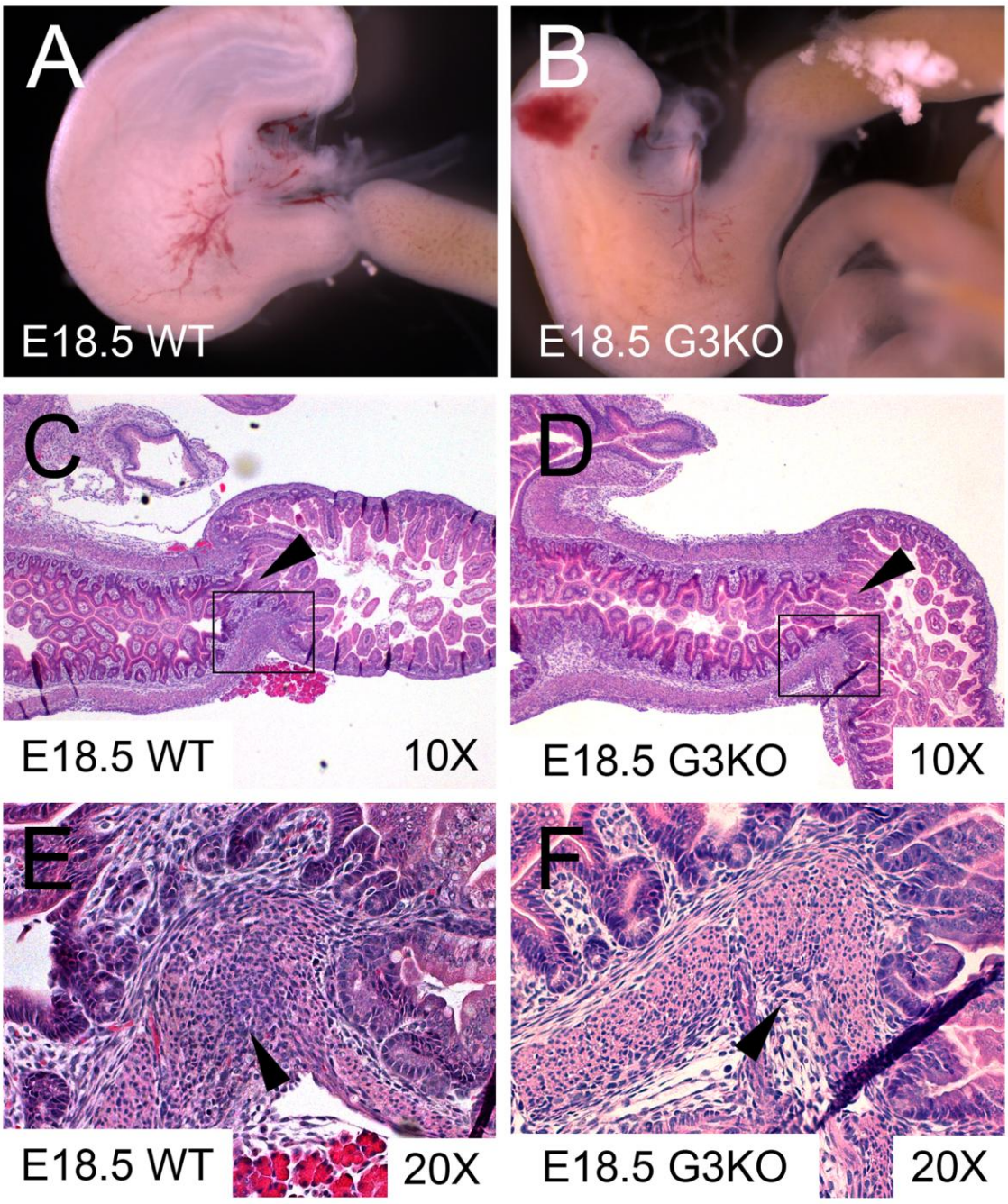

Figure III.7. Gata3 is required for pylorus morphogenesis. Whole mount photomicroscopy and H&E staining of sectioned wild type and *Gata3* null (G3KO) foreguts at E18.5: A) WT whole mount at E18.5, B) G3KO whole mount at E18.5, C) WT H&E at E18.5 (10X), D) G3KO H&E at E18.5 (20X), E) WT H&E at E18.5 (20X), and F) G3KO H&E at E18.5 (20X). Notice the attenuation of the pyloic constriction in D) and apparent loss of smooth muscle in F) (black arrowheads). The white boxes in C) and D) roughly correspond to the area of higher magnification in E) and F), respectively.

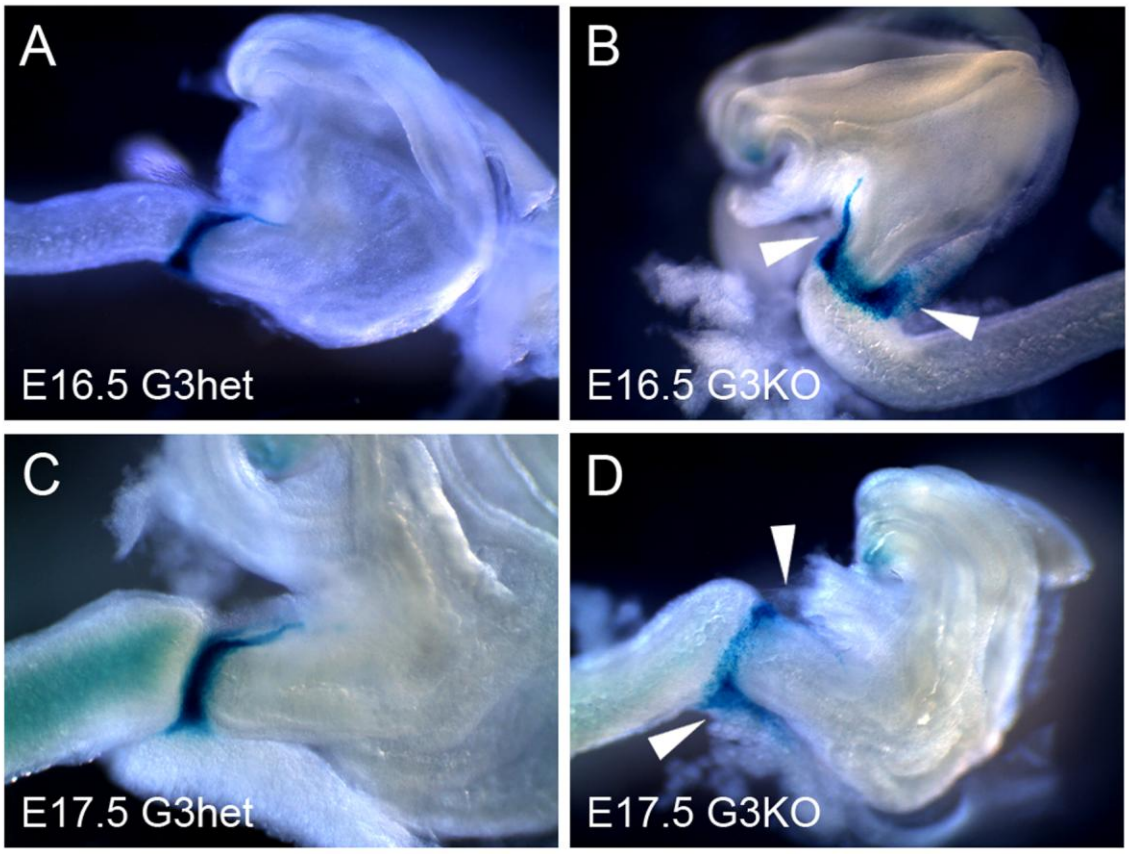

Figure III.8. Gata3 is required for the maintenance of the ventral pyloric cords. Whole mount X-gal staining of *Gata3* heterozygote (G3het) and *Gata3* null (G3KO) foreguts at E16.5 and E17.5: A) G3het at E16.5, B) G3KO at E17.5, C) G3het at E17.5, and D) G3KO at E17.5. In A) and C) blue X-gal staining represents *Gata3* expression, while in B) and D) it represents transcription from the endogenous *Gata3* locus. Notice the presence of blue staining in the ventral pyloric cords in B) but the absence of that staining in D) (white arrowheads). Also note the expansion of X-gal stain into pancreatic tissue in B) and D) (white arrowheads).

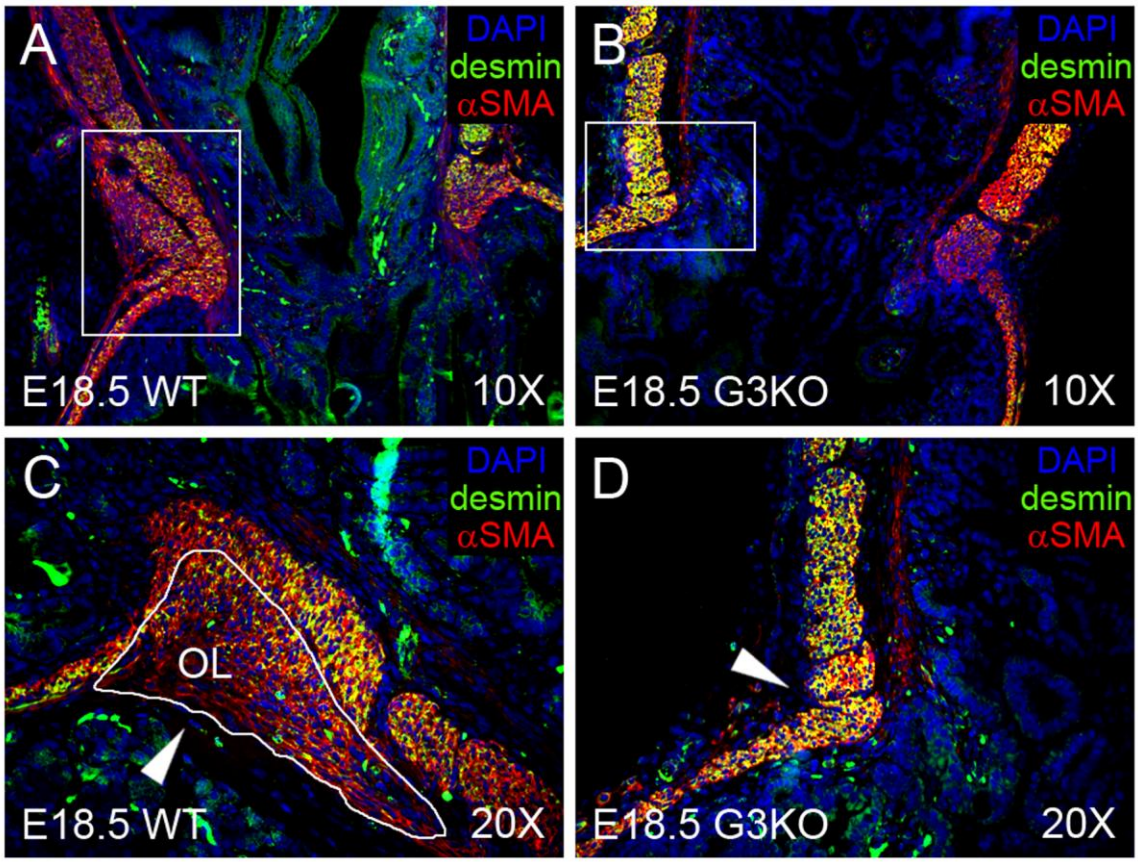

Figure III.9. Loss of Gata3 disrupts the outer longitudinal smooth muscle layer at the pylorus. Immunofluorescence for desmin and αSMA in sectioned wild type and *Gata3* null (G3KO) foregut at E18.5: A) WT at E18.5 (10X), B) G3KO at E18.5 (10X), C) WT at E18.5 (20X), and D) G3KO at E18.5 (20X). Green is desmin, red is αSMA, and nuclei are counterstained in blue with DAPI. Red staining is myofibroblasts, green staining is smooth muscle precursor cells, and yellow staining is differentiated smooth muscle. Notice the absence of yellow staining in the OL region in B) and D) (white arrowheads). The white boxes in A) and B) roughly correspond to the area of higher magnification in C) and D), respectively. OL = outer longitudinal smooth muscle layer. Note again the apparent attenuation of pyloric constriction and the difference in shape of the inner circular smooth muscle layer. Immunostaining by Mr. Ajay Prakash.

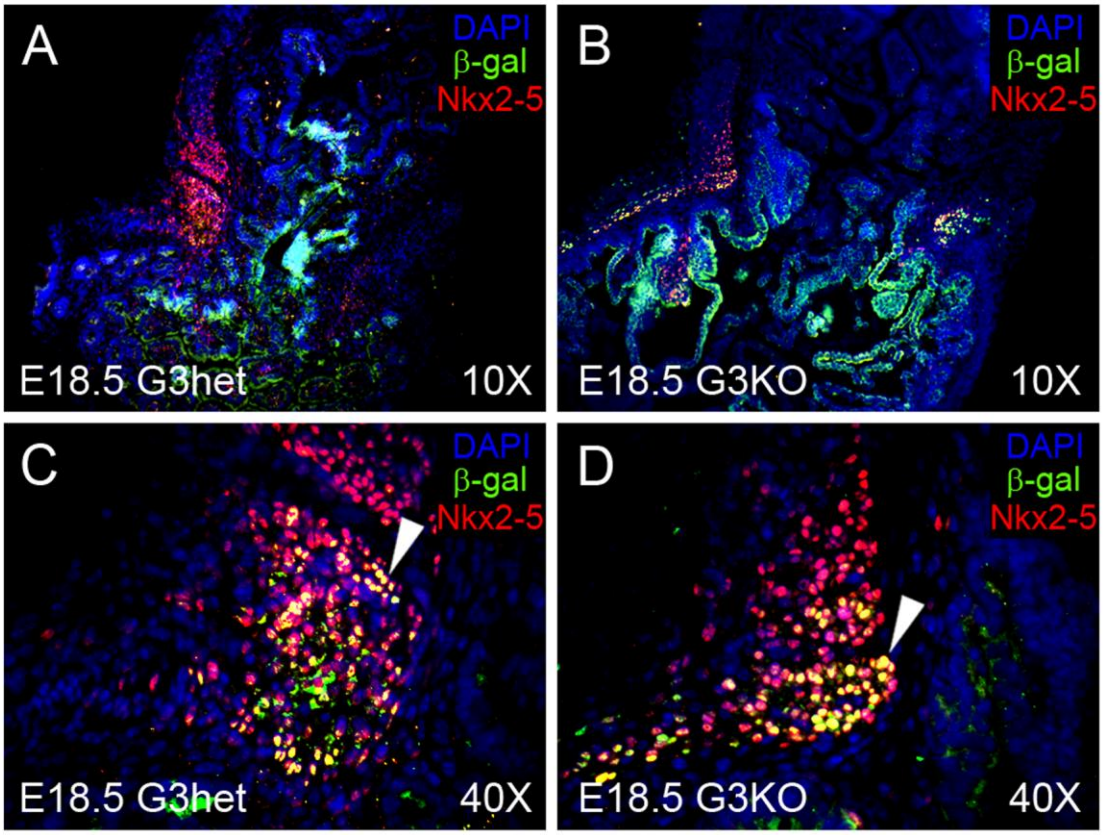

Figure III.10. Gata3 is not epistatic to Nkx2-5. Immunofluorescence for Nkx2-5 and β-gal in sectioned *Gata3* heterozygous (G3het) and *Gata3* null (G3KO) foregut: A) G3het at E18.5 (10X), B) G3KO at E18.5 (10X), C) G3het at E18.5 (40X), and D) G3KO at E18.5 (40X). Nkx2-5 is red, β-gal is green, and nuclei have been counterstained blue with DAPI. In A) and C) red staining is Nkx2-5 expression, green staining is Gata3 expression, and yellow staining is co-expression of Gata3 and Nkx2-5. In B) and D) red staining is Nkx2-5 expression, green staining is transcription from the endogenous *Gata3* locus, and yellow is co-expression of Nkx2-5 and the endogenous *Gata3* locus. Notice the presence of yellow nuclei in C) and D) (white arrowheads). The white boxes in A) and B) roughly correspond to the area of higher magnification in C) and D), respectively. Immunostaining by Mr. Ajay Prakash.

Table III.1. Protocols for immunostaining experiments.

Experiment	Protocol
Gata3 only (**Figure III.2**)	1. Mouse anti-Gata3 O/N at 4°C (1:100) 2. Biotinylated horse anti-mouse/rabbit/goat (universal) IgG for 30 min at RT (1:200) 3. Cy3-conjugated mouse anti-biotin for 30 min at RT (1:100)
Nkx2-5 only (**Figure III.4**)	1. Goat anti-Nkx2-5 O/N at 4°C (1:100) 2. Biotinylated horse anti-mouse/rabbit/goat (universal) IgG for 30 min at RT (1:200) 3. Cy3-conjugated mouse anti-biotin for 30 min at RT (1:100)
Nkx2-5, αSMA, and desmin (**Figures III5.A,C**)	1. Goat anti-Nkx2-5 O/N at 4°C (1:100) 2. Biotinylated horse anti-mouse/rabbit/goat (universal) IgG for 30 min at RT (1:200) 3. Cy5-conjugated mouse anti-biotin for 30 min at RT (1:100) 4. Rabbit anti-desmin O/N at 4°C (1:500) 5. Alexa Fluor 488 donkey anti-rabbit IgG for 30 min at RT (1:500) 6. Cy3-conjugated, mouse anti-aSMA for 30 min at RT (1:500)
Gata3 and desmin (**Figures III5B,D**)	1. Mouse anti-Gata3 O/N at 4°C (1:100) 2. Biotinylated horse anti-mouse/rabbit/goat (universal) IgG for 30 min at RT (1:200) 3. Cy3-conjugated mouse anti-biotin for 30 min at RT (1:100) 4. Rabbit anti-desmin O/N at 4°C (1:500) 5. Alexa Fluor 488 donkey anti-rabbit IgG for 30 min at RT (1:500)
αSMA and desmin (**Figures III.6,9**)	1. Rabbit anti-desmin O/N at 4°C (1:500) 2. Alexa Fluor 488 donkey anti-rabbit IgG for 30 min at RT (1:500) 3. Cy3-conjugated, mouse anti-aSMA for 30 min at RT (1:500)
Nkx2-5 and β-gal (**Figure III.10**)	1. Goat anti-Nkx2-5 O/N at 4°C (1:100) 2. Biotinylated horse anti-universal IgG for 30 min at RT (1:200) 3. Cy5-conjugated mouse anti-biotin for 30 min at RT (1:100) 4. Chicken anti-β-gal O/N at 4°C (1:1000)[1] 5. Biotinylated rabbit anti-chicken IgY for 30 min at RT (1:200) 6. Cy3-conjugated mouse anti-biotin for 30 min at RT (1:100)

[1]Sections were pre-blocked again with the Avidin/Biotin Blocking Kit prior to incubation with the chicken anti-β-gal primary antibody

III.5. Attribution

The first draft of this chapter was written by Aaron Udager. The text was subsequently revised with suggestions from Dr. Deborah Gumucio, Dr. Linda Samuelson, Dr. Scott Barolo, and Mr. Ajay Prakash.

III.6. Publication

We anticipate that text from this chapter will be revised to form a manuscript for publication.

III.7. References

1. Li, X., Udager, A.M., Hu, C., Qiao, X.T., Richards, N., and Gumucio, D.L. (2009). Dynamic patterning at the pylorus: formation of an epithelial intestine-stomach boundary in late fetal life. Dev Dyn *238*, 3205-3217.

2. Cai, W.Q., and Gabella, G. (1984). Structure and innervation of the musculature at the gastroduodenal junction of the guinea-pig. J Anat *139 (Pt 1)*, 93-104.

3. Daniel, E.E. (1992). Sphincters : normal function--changes in diseases, (Boca Raton, Fla.: CRC Press).

4. Horton, B.T. (1928). Pyloric musculature, with special reference to pyloric block. Am J Anat *41*, 197-226.

5. Panteli, C. (2009). New insights into the pathogenesis of infantile pyloric stenosis. Pediatr Surg Int *25*, 1043-1052.

6. Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K., and Yamamura, K. (2002). Region-specific gastrointestinal Hox code during murine embryonal gut development. Dev Growth Differ *44*, 77-84.

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

8. Nielsen, C., Murtaugh, L.C., Chyung, J.C., Lassar, A., and Roberts, D.J. (2001). Gizzard formation and the role of Bapx1. Dev Biol *231*, 164-174.

9. Smith, D.M., Grasty, R.C., Theodosiou, N.A., Tabin, C.J., and Nascone-Yoder, N.M. (2000). Evolutionary relationships between the amphibian, avian, and mammalian stomachs. Evol Dev *2*, 348-359.

10. Moniot, B., Biau, S., Faure, S., Nielsen, C.M., Berta, P., Roberts, D.J., and de Santa Barbara, P. (2004). SOX9 specifies the pyloric sphincter epithelium through mesenchymal-epithelial signals. Development *131*, 3795-3804.

11. Roberts, D.J., Johnson, R.L., Burke, A.C., Nelson, C.E., Morgan, B.A., and Tabin, C. (1995). Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. Development *121*, 3163-3174.

12. Smith, D.M., Nielsen, C., Tabin, C.J., and Roberts, D.J. (2000). Roles of BMP signaling and Nkx2.5 in patterning at the chick midgut-foregut boundary. Development *127*, 3671-3681.

13. Smith, D.M., and Tabin, C.J. (1999). BMP signalling specifies the pyloric sphincter. Nature *402*, 748-749.

14. Theodosiou, N.A., and Tabin, C.J. (2005). Sox9 and Nkx2.5 determine the pyloric sphincter epithelium under the control of BMP signaling. Dev Biol *279*, 481-490.

15. Self, M., Geng, X., and Oliver, G. (2009). Six2 activity is required for the formation of the mammalian pyloric sphincter. Dev Biol *334*, 409-417.

16. Verzi, M.P., Stanfel, M.N., Moses, K.A., Kim, B.M., Zhang, Y., Schwartz, R.J., Shivdasani, R.A., and Zimmer, W.E. (2009). Role of the homeodomain transcription factor Bapx1 in mouse distal stomach development. Gastroenterology *136*, 1701-1710.

17. Elliott, D.A., Solloway, M.J., Wise, N., Biben, C., Costa, M.W., Furtado, M.B., Lange, M., Dunwoodie, S., and Harvey, R.P. (2006). A tyrosine-rich domain within homeodomain transcription factor Nkx2-5 is an essential element in the early cardiac transcriptional regulatory machinery. Development *133*, 1311-1322.

18. Hendriks, R.W., Nawijn, M.C., Engel, J.D., van Doorninck, H., Grosveld, F., and Karis, A. (1999). Expression of the transcription factor GATA-3 is required for the development of the earliest T cell progenitors and correlates with stages of cellular proliferation in the thymus. Eur J Immunol *29*, 1912-1918.

19. van Doorninck, J.H., van Der Wees, J., Karis, A., Goedknegt, E., Engel, J.D., Coesmans, M., Rutteman, M., Grosveld, F., and De Zeeuw, C.I. (1999). GATA-3 is involved in the development of serotonergic neurons in the caudal raphe nuclei. J Neurosci *19*, RC12.

20. Kaufman, C.K., Zhou, P., Pasolli, H.A., Rendl, M., Bolotin, D., Lim, K.C., Dai, X., Alegre, M.L., and Fuchs, E. (2003). GATA-3: an unexpected regulator of cell lineage determination in skin. Genes Dev *17*, 2108-2122.

21. Lim, K.C., Lakshmanan, G., Crawford, S.E., Gu, Y., Grosveld, F., and Engel, J.D. (2000). Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. Nat Genet *25*, 209-212.

22. Kolterud, A., Grosse, A.S., Zacharias, W.J., Walton, K.D., Kretovich, K.E., Madison, B.B., Waghray, M., Ferris, J.E., Hu, C., Merchant, J.L., et al. (2009). Paracrine Hedgehog signaling in stomach and intestine: new roles for hedgehog in gastrointestinal patterning. Gastroenterology *137*, 618-628.

23. Burn, S.F., Boot, M.J., de Angelis, C., Doohan, R., Arques, C.G., Torres, M., and Hill, R.E. (2008). The dynamics of spleen morphogenesis. Dev Biol *318*, 303-311.

24. Moriguchi, T., Takako, N., Hamada, M., Maeda, A., Fujioka, Y., Kuroha, T., Huber, R.E., Hasegawa, S.L., Rao, A., Yamamoto, M., et al. (2006). Gata3 participates in a complex transcriptional feedback network to regulate sympathoadrenal differentiation. Development *133*, 3871-3881.

25. Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev *9*, 1654-1666.

26. Brewer, A.C., Alexandrovich, A., Mjaatvedt, C.H., Shah, A.M., Patient, R.K., and Pizzey, J.A. (2005). GATA factors lie upstream of Nkx 2.5 in the transcriptional regulatory cascade that effects cardiogenesis. Stem Cells Dev *14*, 425-439.

27. Davis, D.L., Wessels, A., and Burch, J.B. (2000). An Nkx-dependent enhancer regulates cGATA-6 gene expression during early stages of heart development. Dev Biol *217*, 310-322.

28. Durocher, D., Charron, F., Warren, R., Schwartz, R.J., and Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. EMBO J *16*, 5687-5696.

29. Durocher, D., and Nemer, M. (1998). Combinatorial interactions regulating cardiac transcription. Dev Genet *22*, 250-262.

30. Jiang, Y., Drysdale, T.A., and Evans, T. (1999). A role for GATA-4/5/6 in the regulation of Nkx2.5 expression with implications for patterning of the precardiac field. Dev Biol *216*, 57-71.

31. Nishida, W., Nakamura, M., Mori, S., Takahashi, M., Ohkawa, Y., Tadokoro, S., Yoshida, K., Hiwada, K., Hayashi, K., and Sobue, K. (2002). A triad of serum response factor and the GATA and NK families governs the transcription of smooth and cardiac muscle genes. J Biol Chem *277*, 7308-7317.

32. Peterkin, T., Gibson, A., and Patient, R. (2003). GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation. EMBO J *22*, 4260-4273.

33. Searcy, R.D., Vincent, E.B., Liberatore, C.M., and Yutzey, K.E. (1998). A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. Development *125*, 4461-4470.

34. Sepulveda, J.L., Belaguli, N., Nigam, V., Chen, C.Y., Nemer, M., and Schwartz, R.J. (1998). GATA-4 and Nkx-2.5 coactivate Nkx-2 DNA binding targets: role for regulating early cardiac gene expression. Mol Cell Biol *18*, 3405-3415.

35. Shiojima, I., Komuro, I., Oka, T., Hiroi, Y., Mizuno, T., Takimoto, E., Monzen, K., Aikawa, R., Akazawa, H., Yamazaki, T., et al. (1999). Context-dependent transcriptional cooperation mediated by cardiac transcription factors Csx/Nkx-2.5 and GATA-4. J Biol Chem *274*, 8231-8239.

36. Zhang, Y., Rath, N., Hannenhalli, S., Wang, Z., Cappola, T., Kimura, S., Atochina-Vasserman, E., Lu, M.M., Beers, M.F., and Morrisey, E.E. (2007). GATA and Nkx factors synergistically regulate tissue-specific gene expression and development in vivo. Development *134*, 189-198.

37. Nagel, S., Meyer, C., Quentmeier, H., Kaufmann, M., Drexler, H.G., and MacLeod, R.A. (2008). MEF2C is activated by multiple mechanisms in a subset of T-acute lymphoblastic leukemia cell lines. Leukemia *22*, 600-607.

38. Jit, I. (1952). The development and the structure of the suspensory muscle of the duodenum. Anat Rec *113*, 395-407.

39. Jit, I., and Grewal, S.S. (1977). The suspensory muscle of the duodenum and its nerve supply. J Anat *123*, 397-405.

40. van der Zypen, E., and Revesz, E. (1984). Investigation of development, structure and function of the phrenicocolic and duodenal suspensory ligaments. Acta Anat (Basel) *119*, 142-148.

CHAPTER IV

GENOME-WIDE CLUSTERING ANALYSIS OF GLI/CI BINDING SITES ACROSS DIVERGENT DROSOPHILA SPECIES REVEALS HEDGEHOG-REGULATED ENHANCERS

Hh pathway activity regulates gene expression via interaction of Gli/Ci transcription factors with specific sequences, Gli/Ci binding sites, in genomic DNA. In *Drosophila* many previously characterized Hh-regulated enhancers contain clusters of Gli/Ci binding sites, and the relative clustering of these sites is preserved in divergent *Drosophila* species. In a genome-wide analysis we detected significant clustering of Gli/Ci sites in *D. melanogaster* and *D. pseudoobscura*, while a closely related set of Gli/Ci sites containing an *in silico* "mutation" that is predicted to prevent binding of Ci protein *in vitro* showed little clustering. We therefore determined the clustering topography surrounding every Gli/Ci site in the genome by calculating a clustering coefficient (CC) that reflected the degree of clustering for each site. To determine the significance of observed clustering, we used Monte Carlo modeling of randomly generated genomes with preserved guanine-cytosine (GC) topography. This analysis revealed a large number of predicted Hh-regulated enhancers. Comparison of clustering topography in multiple *Drosophila* species revealed conserved clusters, several of which were functionally tested. We show that novel intronic enhancers within the genes *inv* and *rdx* are active in Hh-responsive cells of the larval imaginal wing disc.

IV.1. Introduction

The core components of the Hh signaling pathway are highly conserved across evolution, as are the mechanisms of pathway activation and signal transduction [1]. In *Drosophila* extracellular ligand Hh binds to the transmembrane receptor Ptc, which allows another transmembrane protein, Smo, to act on intracellular components. Following a complex series of phosphorylation and proteolysis events, the Hh signal is finally transduced by the downstream transcription factor Ci. [The Gli family of transcription factors (i.e., Gli1, Gli2, and Gli3) transduce Hh signals in the mouse.] Nuclear translocation and binding of Gli/Ci proteins to specific DNA sequences, Gli/Ci binding sites, regulates Hh target gene expression.

In *Drosophila*, Hh signaling is essential for early embryonic segmentation and larval imaginal disc patterning [1]. In these tissues Hh acts as a morphogen: it elicits distinct transcriptional responses at discrete distances from the source of signal. For example, in the developing wing disc, Hh is secreted by cells in the posterior compartment and activates target gene expression in anterior compartment cells close to the A-P compartment boundary. The larval wing disc Hh target genes *ptc* and *decapentaplegic* (*dpp*) are expressed in distinct but overlapping domains. *ptc* is activated in a stripe of cells along the A-P compartment boundary [2]. *dpp* is expressed in a wider domain anteriorly but is excluded from cells that abut the A-P compartment boundary [3]. A few Hh target genes have been identified in other tissues including *wingless* (*wg*) in the early embryo and *hairy* (*h*) in the larval leg disc [4, 5]. However, the paucity of known direct

Hh targets (and by extension Hh-regulated enhancer sequences) is an impediment to a more complete understanding of the roles for Hh in development and homeostasis.

The identification of enhancers is central to distinguishing direct from indirect transcriptional targets of signaling pathways such as that directed by Hh [6]. A number of *in vivo* experimental techniques are available for this purpose [i.e., enhancer traps, chromatin immunoprecipitation (ChIP), etc.] [7, 8]. However, the abundance of high-quality sequence data and annotation resources has spurred the development of *in silico* methods to find enhancers. Many of these approaches analyze the relative spacing (i.e., clustering) of transcription factor binding sites [9-11]. Other methods consider evolutionary conservation as well, including two for identifying Hh-regulated enhancers in mice [12, 13]. In these latter two approaches, evolutionary conservation of sequence and/or transcription factor binding site arrangement is used to limit the amount of sequence that will be searched for binding sites. Much of the genome is therefore ignored from the start. However, it is clear that enhancers evolve, and this evolution appears to depend on selection for function (phenotypic character) rather than for sequence, order or spacing of component binding sites [14].

Rebeiz et al. described a genome-wide approach to identify enhancers regulated by the Notch signaling pathway in *Drosophila* [9]. This method, known as SCORE (Site Clustering Over Random Expectation), detects statistically significant clustering of binding sites for the Notch-regulated transcription factor Suppressor of Hairless [Su(H)]. SCORE, however, focuses primarily on predicted high affinity binding sites that are most

similar to the consensus binding sequence, and information regarding potentially functional low affinity sites is lost. As illustrated by the enhancers for *dpp* and *wg* in *Drosophila*, clustering of non-optimal Gli/ci sites is a property of some Hh-regulated enhancers [3, 5]. Thus, a narrow focus on high-affinity sites is likely to net only a subset of functional enhancer elements. An additional limitation of SCORE is that its method of background clustering estimation does not account for the effect of local variation in genomic GC-content on the probability of clustering, an important consideration for GC-rich binding sites, such as that of Gli/Ci.

In the present study we describe a novel approach that analyzes optimal consensus as well as non-optimal sites, models background clustering to account for genomic topography in GC-content, and uses conservation as a final filter rather than an initial selector. Using this tool, Topo-TX, we show that in the genomes of two distantly related *Drosophlia* species (*D. melanogaster* and *D. pseudoobscura*) there is statistically significant local clustering of Gli/Ci sites. These clusters of Gli/Ci sites occur at frequencies well above the level of clustering predicted in random genomes. In contrast, when the search parameters are slightly modified to include a single nucleotide *in silico* "mutation" in the Gli/Ci binding matrix, which abrogates ci binding *in vitro*, the clustering profile resembles that of the random genome set [12]. We therefore identify potential Hh-regulated enhancers genome-wide by assessing the topography of clustered Gli/Ci sites in *D. melanogaster* and *D. pseudoobscura*. Topo-TX predictions were assessed by examining the ability of putative enhancer sequences to activate reporter gene expression in transgenic flies. We describe novel enhancers within the genes

roadkill (*rdx*) and *invected* (*inv*) that are co-expressed in Hh-responsive cells in the larval imaginal wing disc.

IV.2. Materials and methods

Identification of predicted Gli/Ci binding sites in genomic sequence

A nucleotide distribution matrix derived from *in vitro* DNA binding assays with recombinant Ci protein was obtained from the Genomatix database (www.genomatix.de) [12]. The consensus index vector for a matrix reflects the degree of nucleotide preference at each position [15]. The first 9 bp of the Ci matrix had the highest consensus index vector values; therefore we focused on those positions in our analysis. Custom Perl scripts were designed to analyze the matrix and define a set of 9-mers that passed minimum thresholds for core region (75%) and matrix (81%) similarity to an optimal consensus Gli/ci site [15]. A closely-related set of 9-mers ("mutant" Gli/ci sites) that do not bind Ci protein *in vitro* was defined by changing the invariant cytosine at position 6 of the nucleotide distribution matrix to a guanine (C6G) [12].

Repeat-masked, whole genome FASTA files for *D. melanogaster* (dm3) and *D. pseudoobscura* (dp3) were downloaded from the UCSC Genome Bioinformatics website (genome.ucsc.edu). Custom Perl scripts were designed to catalog the coordinates of Gli/Ci binding sites in both genomes. This list of site positions was annotated using custom Perl scripts and a refFlat transcript annotation file (available from UCSC) to mask potential sites in coding exons.

Generation of scrambled genomic sequence and random genomic distributions of binding sites

One hundred artificial genomes were generated by scrambling bases but maintaining GC-content topography (e.g., A becomes A or T, C becomes C or G, etc) at each position in dm3 or dp3. In this manner the genome-wide topography of GC-content is preserved between real and scrambled genomes. Potential Gli/Ci binding sites were then annotated in these scrambled genomes. Using the total population of positions from these artificial genomes 1,000 independent, random genomic distributions of Gli/Ci binding site positions were generated using a Monte Carlo simulation approach: site positions were pooled and then selected randomly without replacement. These distributions preserve the overall characteristics of sites dm3 or dp3 (i.e., number, affinity, and orientation). This method of random genome generation and modeling ensures that GC topography, the total number of Gli/Ci sites, and the distribution of relative affinities of Gli/Ci sites are similar in random genomes and dm3 or dp3.

Statistical analysis

For the analysis of number and relative distribution of Gli/Ci binding sites in real and scrambled genomic sequences we used the Microsoft Excel function NORMDIST. For Topo-TX, we numerically ordered inter-site distances (cluster sizes) for clusters of *n* adjacent sites in either dm3 or dp3 and calculated the cumulative frequency of each size ("observed value"). These data were plotted as an empirical distribution function. We then compared observed values to the average cumulative frequency ("expected value") in random binding site distributions generated by the Monte Carlo simulation. We

calculated 95% confidence intervals associated with the expected value and determined whether the observed value was within this range [16]. For observed values that were not within the 95% confidence intervals, we calculated the clustering coefficient: observed frequency divided by expected frequency.

Mapping orthologous regions of Gli/Ci clustering

We compared lists of highly clustered Gli/Ci binding sites in either dm3 or dp3 using nearest gene annotation. If an orthologous gene was associated with highly clustered sites in both genomes, we retrieved the genomic coordinates of clustered sites and looked for areas of overlap: batches of genomic coordinates from dm3 were converted to the coordinates of the orthologous region in dp3 by the liftOver tool from the UCSC Genome Bioinformatics website, and custom Perl scripts identified overlapping coordinate domains.

Generation of transgenic reporter flies

Putative enhancers were PCR amplified from wild type *D. melanogaster* genomic DNA and cloned into an eGFP reporter vector (pGanesh-G2) [17]. Plasmid DNA was microinjected into embryos and adults were screened for transgene integration. Larval wing discs were dissected for imaging with whole mount confocal microscopy. Embryos were analyzed by in situ hybridization for eGFP and imaged with light microscopy. Results are representative of multiple (>3) independent lines.

IV.3. Results

Identification of putative Gli/Ci binding sites from in vitro binding data

The relative affinity of recombinant Ci protein for specific DNA sequences has been assayed *in vitro* via measurement of independent, competitive DNA binding reactions between a labeled optimal consensus Gli/Ci binding site (GACCACCCA) and unlabeled, non-optimal binding sites that carry a single nucleotide substitution [12, 18]. Ci-DNA binding data from these experiments can be conveniently represented as a mononucleotide distribution [or position weight (PW)] matrix with four rows, one for each of the potential nucleotides (A, C, G or T), and *n* columns (where *n* is equal to the number of positions in the binding site). To evaluate potential Gli/Ci binding sites with this matrix, we utilized the MatInspector method, which calculates the weighted similarity to the optimal consensus site ("% matrix similarity") [15]. With reference to the *in vitro* binding data, % matrix similarity can be used as an estimate of relative affinity [15, 18].

The MatInspector-defined optimized threshold, which was designed to minimize false positive Gli/Ci binding sites but likely also misses some low affinity sites, is 81% matrix similarity. Forty-one 9-mer binding sites (of distinct sequence) meet or exceed this threshold (Table IV.1). Changing the minimum threshold alters the balance between sensitivity and specificity of binding site selection: 236 and six unique sites pass a minimum threshold of 70% or 90% matrix similarity, respectively (Figure IV.1). The scores for non-optimal Gli/Ci binding sites in the *wg*, *dpp*, and *hairy* Hh-regulated enhancers range from 68.7% to 91.5% matrix similarity (Table IV.2). Thus an 81%

matrix similarity threshold establishes a comprehensive set of Gli/Ci binding sites for computational analysis, though some potentially functional Gli/Ci binding sites are missed.

Non-coding regions of the *D. melanogaster* genome are significantly enriched in Gli/Ci binding sites

Using an 81% matrix similarity threshold, we identified a total of 35,707 putative Gli/Ci binding sites in the *D. melanogaster* genome. Of those, 9,542 (26.7%) are found in coding exons of annotated genes, 2,879 (8.1%) are in located in repeat sequence, and 23,509 (65.8%) are in non-repeat, non-coding DNA (i.e., promoters, untranslated regions, introns, intergenic regions). We then generated a cohort of 100 scrambled *D. melanogaster* genomes that have preserved GC-content topography. Using the same matrix similarly threshold, we identified Gli/Ci binding sites in these scrambled *D. melanogaster* genomes. Compared to the scrambled genomes the real *D. melanogaster* genome has more total sites, 35,707, as opposed to an average of 26,330 in the scrambled genomes (Figure IV.2A). This represents a 1.36-fold increase ($p < 0.01$) in total sites, and the number of sites in non-repeat, non-coding sequence, repeat sequence, and exonic sequence is 1.45-, 1.58-, and 1.12-fold enriched ($p < 0.01$), respectively (Figures IV.2B-D).

Beyond the total number of sites, the relative distribution of sites in the various sequence types (i.e., non-repeat, non-coding, repeat, and exonic) is significantly different. Compared to the scrambled genomes the percentage of sites in non-repeat, non-coding and repeat sequence in the real *D. melanogaster* genome is 1.06- and 1.16-fold greater (p

< 0.01), respectively (Figures IV.2.E-F). In contrast, the occurrence of sites in exons is 1.20-fold less ($p < 0.01$) than in the scrambled genomes (Figure IV.2.G).

Significant genome-wide clustering of Gli/Ci binding sites in divergent *Drosophila* species

The *ptc*, *dpp*, *wg*, and *hairy* enhancers all contain multiple Gli/Ci binding sites, which suggests that tight clustering of Gli/Ci binding sites may be a property of many Hh-regulated enhancers [2-5]. To identify other regions in the genome that contain clusters of Gli/Ci binding sites, we determined the clustering profile of all Gli/Ci sites genome-wide. Using a modification of the method described previously by Rebeiz et al., we recorded the distances between all possible clusters of *n* consecutive sites in the real *D. melanogaster* genome ([9]; see Chapter IV.2). Using these values we determined the observed frequency distribution of inter-site distances (i.e., cluster interval sizes) for clusters of *n* sites, for *n* = 2-10. We then used a Monte Carlo method to model the expected frequency distribution of cluster interval sizes for clusters of *n* sites (see Materials and Methods). Finally we compared the observed and expected frequencies for each of these cluster interval sizes and, on a site by site basis, calculated a clustering coefficient (CC) that describes the probability that each identified binding site is part of a cluster of Gli/Ci sites [CC = (observed frequency of clustering at a given interval size) / (expected frequency of clustering at the same interval size)]. A CC greater than 1 indicates that the observed frequency of a cluster interval size is greater than expected.

After excluding Gli/Ci sites in coding exons and/or repeat sequences, we performed independent analyses on clusters of varying number of *n* sites, for *n* = 2-10. Significant

clustering is observed for all *n* examined, and the maximum CC of cluster interval sizes ranges from 4.929 (for *n* = 2 sites) to 74 (for *n* = 3 sites) (Figure IV.3A). Genome-wide 28% (6,295) and 3% (653) of sites had a maximum CC of greater than two or ten, respectively. Similar results were observed when this analysis was applied to Gli/Ci binding sites in the genome of the distantly related species *D. pseudoobscura*. The range of maximum CC, however, was lower for all *n* examined (Figure IV.3B).

Clustered Gli/Ci binding sites are associated with known Hh-regulated enhancers

To provide initial validation that our method can identify Hh-regulated enhancers *in silico*, we examined the clustering coefficient for each of the Gli/Ci binding sites in the *ptc*, *dpp*, *wg*, and *hairy* enhancers. The three optimal consensus sites in the *ptc* enhancer are part of a larger cluster of four sites in 502 bp, each having a maximum CC of 14.2. At the selected % matrix similarity threshold (81%) of our analysis, only one of the three known non-optimal sites in the *dpp* enhancer is identified (GGCCACCTA), but it is part of a large cluster of nine sites in 8,060 bp, each having a maximum CC of 3.739. These sites span a large cis-regulatory region that directs *dpp* expression in Hh-responsive patterns in larval imaginal discs and the early embryo [19]. Similarly, in the *wg* enhancer, only a single previously identified site (GACCTCCCA) is detected in our analysis, but it is part of a larger cluster of seven sites in 7,799 bp, each having a maximum CC of 2.007. In the *hairy* enhancer there are two previously identified Gli/Ci binding sites, one of which (GACCTCCCA) is detected by our method. This site is part of a larger cluster of ten sites in 9,293 bp, each having a maximum CC of 4.027. In addition, two previously unrecognized predicted sites in the *hairy* enhancer are identified by our analysis

(GACCGCCAA and GACCCCACA): these sites are 8 bp apart and each has a maximum CC of 4.196. Thus Topo-TX detects all four of these previously identified enhancers because they are each associated with a statistically significant cluster of Gli/ci sites.

A set of mutant Gli/ci binding sites shows little clustering in D. melanogaster

To assess whether the significant clustering observed for Gli/Ci binding sites was a specific property of functional Gli/Ci sites, we repeated the analysis in *D. melanogaster* using a set of mutant Gli/Ci sites. The analysis was done using the same binding matrix characteristics and threshold as was used for functional Gli/Ci sites, except that the sequence of the binding matrix contained a single nucleotide substitution [cytosine to guanine at position 6 (mut6CG)], which is predicted from *in vitro* DNA binding data to prevent Ci protein binding while preserving GC-content topography. For cluster sizes n = 2-10 the distribution of CC for mutC6G sites was dramatically different than wild type (Figure IV.4; data not shown). Whereas CC for wild type sites increased significantly at smaller interval sizes, mutC6G CC did not. Furthermore there was no significant association between clustered mutC6G sites and the *ptc*, *dpp*, *wg*, or *hairy* enhancers (data not shown). This suggests that the high clustering coefficients observed for wild type Gli/Ci sites in *D. melanogaster* are a functionally relevant property.

Conservation of Gli/ci binding site clustering between divergent Drosophila species identifies known and potential hh-regulated enhancers

From the Topo-TX analysis, we identified many significant clusters of Gli/ci sites, which we classified as potential Hh-regulated enhancers. One approach to further enrich for functional vs. non-functional enhancers is to examine the conservation of binding site

clustering in two distantly related Drosophila species, *D. melanogaster* and *D. pseudoobscura* [20]. Table IV.3 lists 25 such regions that exhibit significant Gli/Ci clustering in both genomes. The *ptc* and *hairy* enhancers are identified by this analysis. In contrast, neither the *dpp* nor *wg* enhancers are detected. However, a part of the large *dpp* cis-regulatory element just flanking the *dpp* enhancer is found [19].

Though some functional enhancers are overlooked by the application of the conservation filter, this may be an appropriate way to decrease the false discovery rate. Among the regions identified by this analysis are potential intronic enhancers within the genes *inv* and *rdx* (Table IV.3). Both of these genes are expressed in the larval imaginal wing disc [21, 22]. The predicted *inv* enhancer has nine Gli/ci sites in 2,549 bp and these sites have maximum clustering coefficients that vary between 21 and 27 (Table IV.4); this is one of the most Gli/Ci site-enriched areas of the genome. The *inv* enhancer contains two optimal consensus Gli/ci sites and the other seven sites differ from the optimal consensus by a single nucleotide. The predicted *rdx* enhancer has very different properties. In 1,567 bp it contains three sub-optimal Gli/Ci sites, each of which has a maximum CC of 2.7 (Table IV.4).

Predicted Hh-regulated enhancers are active in Hh-responsive cells of the Drosophila wing disc

To functionally test the putative intronic Hh enhancers identified in *inv* and *rdx*, we generated transgenic flies carrying reporter constructs driven by these sequences and examined their expression in spatial domains known to be regulated by Hh in the imaginal wing disc. The *inv* enhancer drives GFP expression in a narrow stripe,

perpendicular to the A-P axis (Figure IV.5A). Similar to the *ptc* enhancer the *inv-GFP* expression domain is in the center of the wing disc, near the A-P compartmental boundary. Its dorsal and ventral boundaries, however, do not extend beyond the wing pouch [2]. In contrast, the *rdx* enhancer drives GFP expression extending along the entire D-V axis of the wing disc near the A-P compartmental boundary (Figure IV.5D).

We next examined whether *inv-GFP* and *rdx-GFP* are expressed in Hh-responsive cells of the wing disc. To accomplish this, we took advantage of a transgene reporter driven by a synthetic *dpp* enhancer [*dpp(ci-HiAf)-RFP*], which contains optimal consensus Gli/Ci sites in place of its three non-optimal sites; this transgene drives RFP expression in Hh-responsive cells in the wing disc (Scott Barolo, University of Michigan, Department of Cell and Developmental Biology, personal communication). The expression domain of *dpp(ci-HiAf)-RFP* is shown in Figures IV.5B and E. Co-expression of *dpp(ci-HiAf)-RFP* and *inv-GFP* is observed within the wing pouch but not in more dorsal or ventral domains of the wing disc (Figure IV.5C). Similarly, there is significant co-expression between *dpp(ci-HiAf)-RFP* and *rdx-GFP* along the entire D-V axis (Figure IV.5F). However, whereas the *inv-GFP* expression domain is narrower than *dpp(ci-HiAf)-RFP*, *rdx-GFP* is slightly wider. These differences in expression domain might reflect the difference in affinity of Ci protein for the sites within the two enhancers. Indeed, the *rdx* enhancer and the native *dpp* enhancer share a wide expression domain and predicted low-affinity, non-optimal Gli/Ci sites, while the *inv* enhancer, the *ptc* enhancer and the synthetic *dpp(ci-HiAf)* transgene contain high-affinity Gli/Ci sites and a narrower expression domain that abuts the A-P boundary. The association of high-affinity Gli/Ci

sites with expression close to the morphogen source has also been noted and extensively investigated in Dr. Barolo's laboratory.

IV.4. Discussion

In this study we examined the genome-wide clustering properties of Gli/Ci sites in divergent *Drosophila* species (*D. melanogaster* and *D. pseudoobscura*). We report that there is significantly greater clustering of these sites than expected by random chance, and clustered Gli/Ci sites are associated with known Hh-regulated enhancers. In contrast, a set of highly similar sites that cannot bind Ci protein *in vitro* shows little overall clustering. By comparing orthologous regions of clustered Gli/Ci sites, we identified a number of putative Hh-regulated enhancers. Finally, we show that two of these novel intronic enhancers, located within the genes *rdx* and *inv*, are active in Hh-responsive cells of the larval imaginal wing disc. Thus we propose that Hh-regulated enhancers can be identified *in silico* by comparative evolutionary analysis of genome-wide Gli/Ci site clustering.

The two novel enhancers identified and functionally tested here, *inv* and *rdx*, are both located in intron 1, but they have different properties with respect to their component Gli/Ci sites and drive GFP expression in spatially distinct, but overlapping domains. *Inv* contains highly clustered Gli/Ci sites in both *D. melanogaster* and *D. pseudoobscura*. *inv* and *engrailed* (*en*) are linked genes in extant *Drosophila* species, and *inv* is thought to have arisen by a duplication of the ancestral *en* gene [21]. While inv and en share

significant homology within a 117 amino acid region that includes the homeodomain, the *inv* and *en* genes are very different. The *inv* locus spans more than 30 kb, while the *en* locus is only about 4 kb. In particular, intron 1 of *inv* is much larger than the corresponding *en* intron 1 (27,659 bp vs. 1,133 bp), which contains regulatory elements that direct expression in the early embryo [21, 23, 24]. There are only subtle differences in *inv* and *en* expression patterns in early embryos; these differences are mediated by tightly linked regulatory elements [21, 25]. The extent to which the *inv* enhancer controls *en* versus *inv* expression is not clear. In the wing disc, *inv* and *en* are expressed throughout the posterior compartment, where they repress anterior compartment cell fate and regulate *hh* expression [26]. In late larval stages, *inv* and/or *en* are activated by Hh in a narrow domain of anterior compartment cells within the wing pouch, along the A-P compartmental boundary [27-30]. The correlation between this Hh-dependent expression pattern and the expression domain of *inv-GFP* suggests that the *inv* enhancer directs *inv* and/or *en* expression specifically in the cells of the anterior compartment.

The *rdx* enhancer is located in the large intron 1 of the *rdx* isoforms B/C/D and contains clustered Gli/Ci sites in both *D. melanogaster* and *D. pseudoobscura* [22]. *rdx* was recently identified as a negative regulator of the Hh signaling pathway [22, 31]. It encodes a small adaptor protein involved in ubiquitination and degradation of full-length Ci protein. Interestingly, *rdx* expression is activated by Hh signaling, similar to the canonical negative Hh pathway regulator *ptc*. In the wing disc, *rdx* is expressed in a broad domain of anterior compartment cells near the A-P compartmental boundary along the length of D-V axis [22]. This pattern is very similar to *rdx-GFP* expression, which

indicates that *rdx* is a likely a direct target of Hh signaling in the wing disc. In addition to *ptc* (and now *rdx*) in *Drosophila* and *Ptch1*, *Gli1* and *Hhip1* in mouse, there is recent evidence of transcriptional regulation of other Hh pathway components by Hh pathway activity [32]. This illustrates the complex cellular mechanisms that are involved in establishing discrete domains of transcriptional response to Hh morphogen signaling. For example, in the wing disc, the negative Hh pathway regulators *ptc* and *rdx* are expressed in overlapping domains, but act on different parts of the pathway [22]. Finally SPOP, the human ortholog of Rdx, can regulate Gli protein levels, and *SPOP* has been linked to renal cell cancer [31, 33].

The expression domains of the *inv* and *rdx* enhancers reveal an apparent correlation between the stripe patterns in the wing disc (i.e., A-P width, D-V length, proximity to the A-P compartmental boundary) and the composition of Gli/Ci sites in Hh-regulated enhancers. The enhancer that contains optimal consensus Gli/Ci sites (*inv*) has a narrow expression domain. The cells in this region directly abut the A-P compartmental boundary, are closest to the source of Hh, and thus, experience the highest levels of Hh signaling. In contrast, the enhancer that contains non-optimal Gli/Ci sites (*rdx*) has a broader expression domain. Similar findings are seen for the *ptc* and *dpp* enhancers, which have high- and low-affinity Gli/Ci sites, respectively. These findings suggest that the type of Gli/Ci binding sites (i.e., optimal consensus vs. non-optimal) in a putative Hh-regulated wing disc enhancer may have predictive value for its expression pattern. In addition, these data agree well with a proposed "reverse French flag" model of Hh morphogen signaling, wherein opposing A-P gradients of Ci repressor (Ci^R) and activator

(Ci^A) expression mediate distinct transcriptional responses (unpublished data, Dr. Scott Barolo, Department of Cell and Developmental Biology, University of Michigan). In this model, optimal consensus sites direct expression in a narrow region close to the Hh source: high protein binding affinity for both Ci^A and Ci^R renders optimal consensus sites acutely sensitive the decreasing ratio of Ci^A to ci^R expression at increasing distances from the Hh source. Mutagenesis of Gli/Ci sites in the *inv* and *rdx* enhancers will be necessary to test this hypothesis.

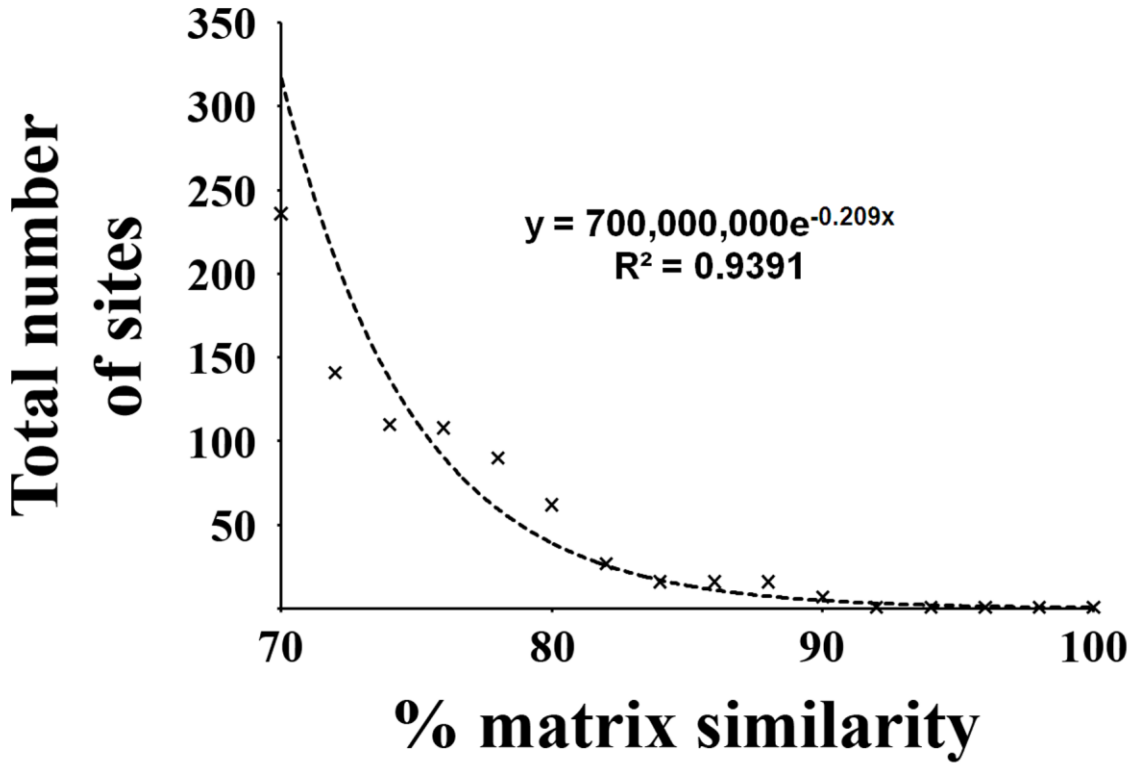

Figure IV.1. The number of predicted Gli/Ci binding sites increases exponentially as a function of % matrix similarity. An X-Y scatter plot of the total number of Gli/Ci binding sites (plotted on the y-axis) for a given % matrix similarity threshold (plotted on the x-axis). The data are plotted as "x" marks. The inverse exponential function ($y = 700{,}000{,}000e^{-0.209x}$) is plotted as a dashed line. R^2 is the coefficient of determination. It is an estimator of goodness of fit between data and underlying models. For example an R^2 of 1 indicates that the model is a perfect fit. In our case the model is the inverse exponential function and the R^2 indicates that it fits the data reasonably well.

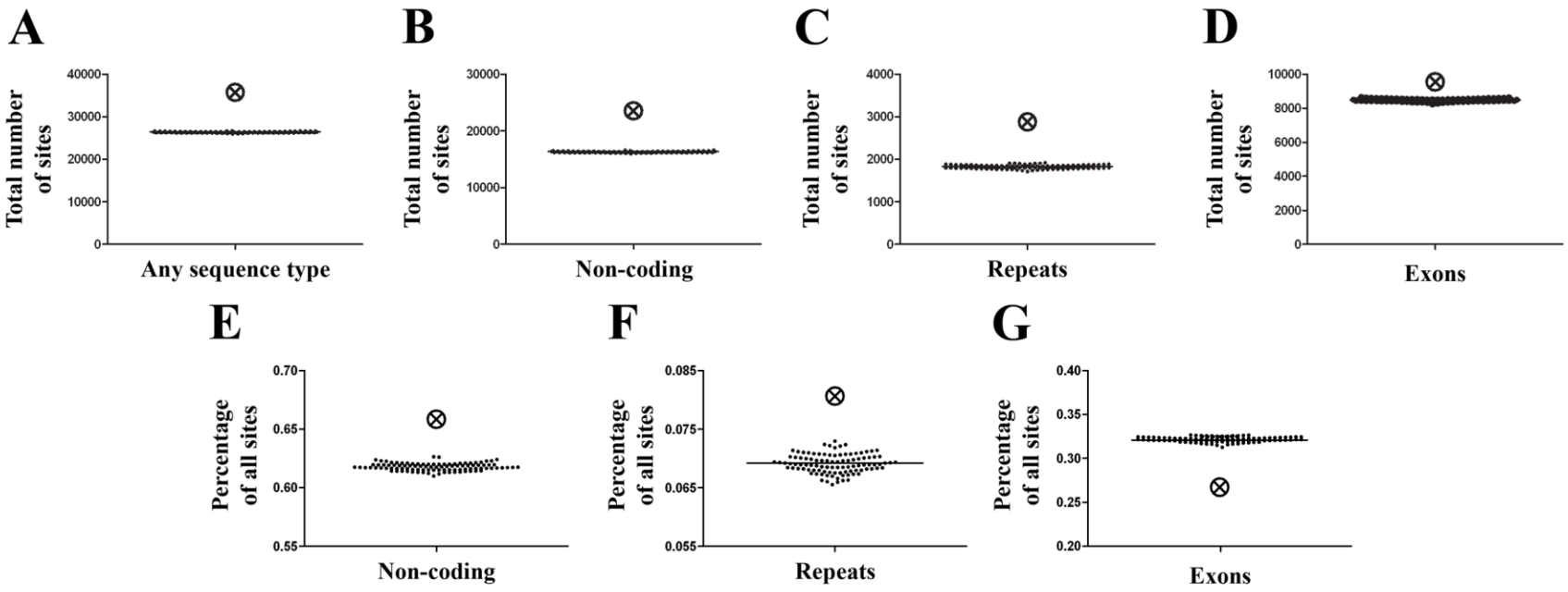

Figure IV.2. Non-coding regions in D. melanogaster are significantly enriched in Gli/Ci binding sites. Vertical scatter plot comparing the total number or relative percentage of Gli/Ci sites (plotted on the y-axis) for specific sequence types in real and 100 scrambled *D. melanogaster* genomes: A) total number of sites in the genome, B) total number of sites in non-repeat, non-coding regions, C) total number of sites in repeat sequence, D) total number of sites in coding exons, E) percentage of all sites in non-coding regions, F) percentage of sites in repeat sequence, and G) percentage of sites in coding exons. The population of data from the scrambled genomes was approximately normally distributed. Notice that there is significant enrichment of Gli/Ci sites in A), B), C), and D) relative to the scrambled genomes ($p<0.01$). Also note that while the relative percentage of Gli/Ci sites in E) and F) is significantly greater than expected, it is significantly less than expected for Gli/Ci sites in G) ($p<0.01$). Small circles = scrambled genomes; large circle with "x" in the middle = real genome.

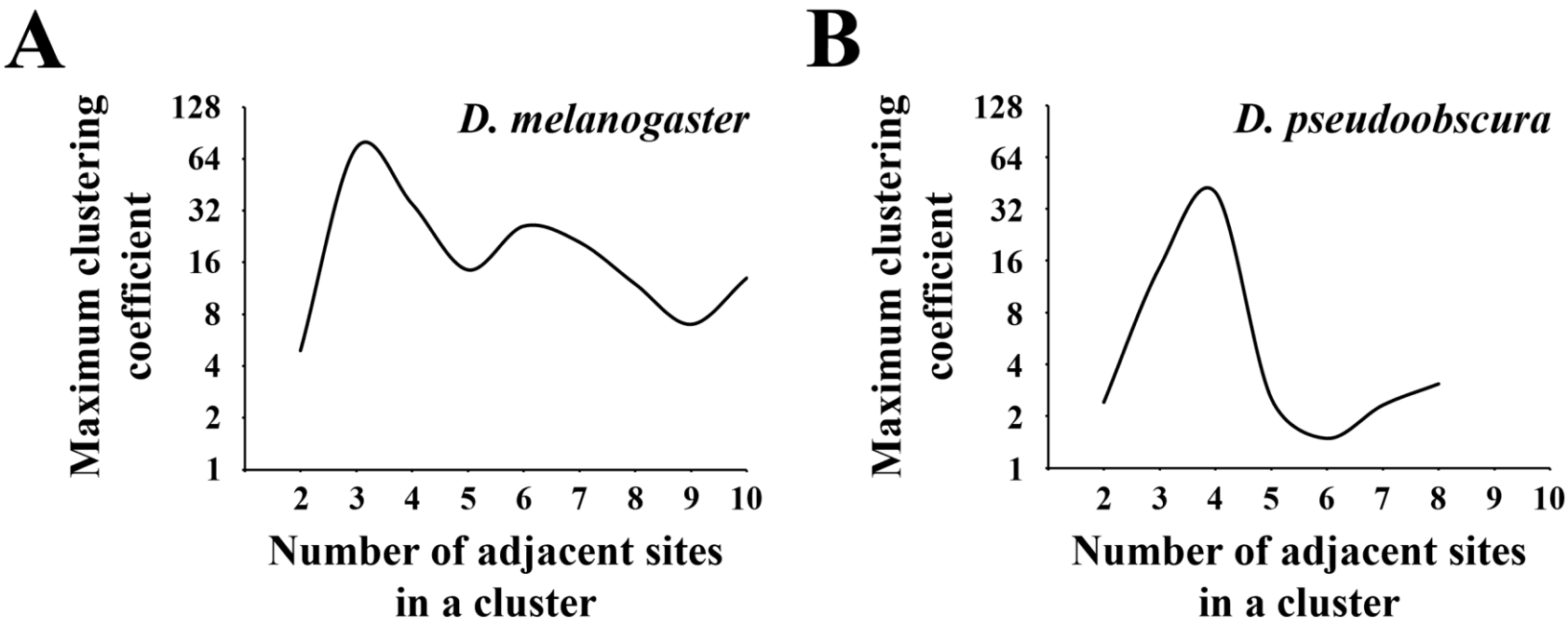

Figure IV.3. Gli/Ci binding sites are significantly clustered in two divergent Drosophila species. X-Y scatter plots of the maximum clustering coefficient (CC; plotted on the y-axis) for a given number of adjacent Gli/Ci sites in a cluster (plotted on the x-axis): A) *D. melanogaster*, for n = 2-10 sites and B) *D. pseudoobscura*, for n = 2-8 sites. The data are connected by a smoothed line. The scale of the y-axis is log base 2. Notice that both A) and B) show large maximum CC, with a peak between three and four sites. Also note that A) has larger values than B) for all number of adjacent sites in a cluster.

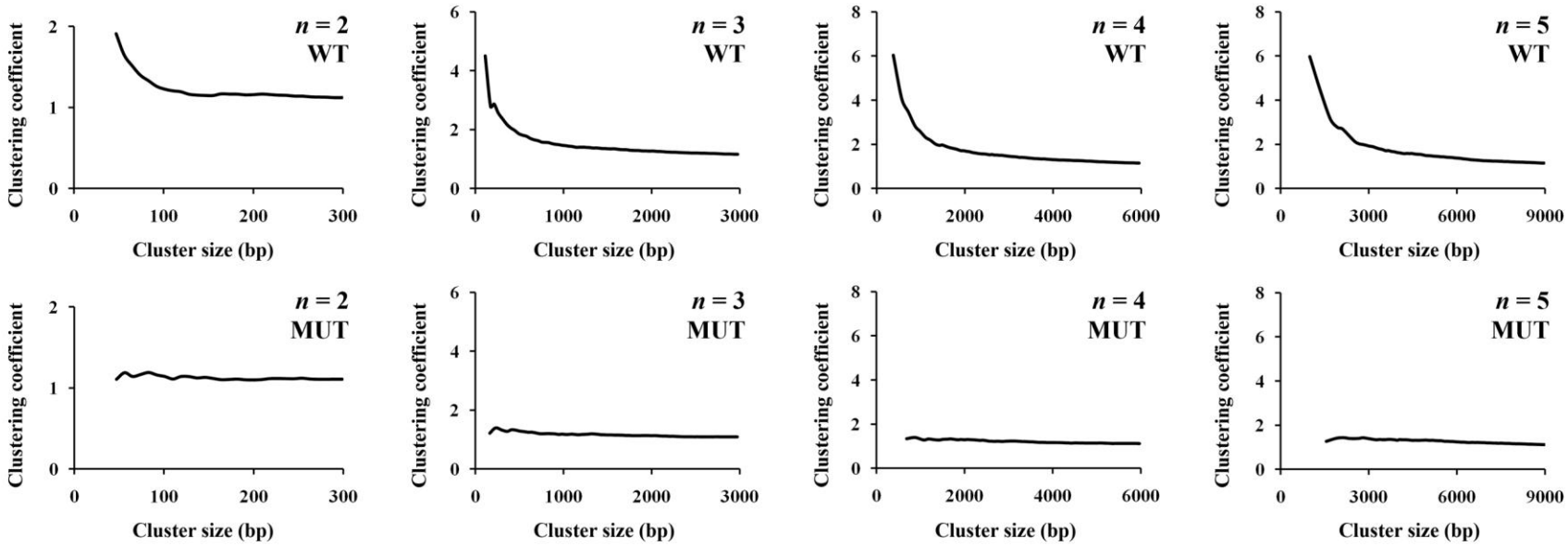

Figure IV.4. Significant clustering of likely functional Gli/Ci binding sites and lack of clustering of a closely related set of mutant Gli/Ci sites that cannot bind Ci protein. X-Y scatter plots of the clustering coefficient (CC; plotted on the y-axis) for various cluster sizes (in bp; plotted on the x-axis) and number *n* of adjacent sites in a cluster. Top = wild type (WT) Gli/Ci binding sites; Bottom = mutant (MUT) Gli/Ci binding sites. Data are connected by a smoothed line. For a given number *n* of adjacent sites in a cluster the WT and MUT data are paired and plotted on the same axes. Notice that for WT sites the CC increases dramatically as cluster size decreases, CC changes little for MUT sites.

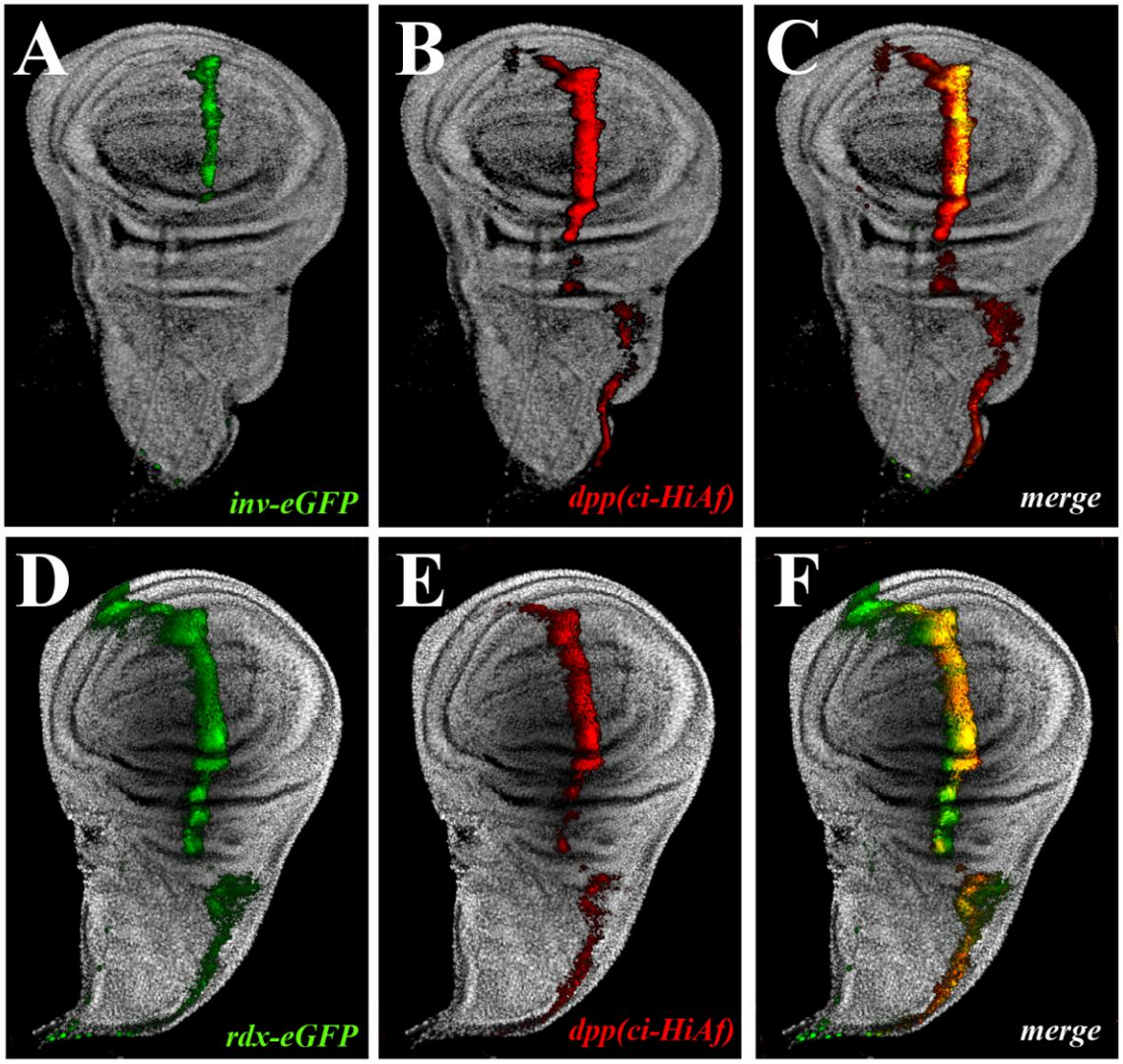

Figure IV.5. Novel predicted enhancers are active in Hh-responsive cells of the Drosophila larval imaginal wing disc. Fluorescent microscopy of *inv-GFP* or *rdx-GFP* and *dpp(ci-HiAf)-RFP* expression patterns in *D. melanogaster* larval imaginal wing discs: A) *inv-GFP*, B) *dpp(ci-HiAf)-RFP*, C) *inveGFP* and *dpp(ci-HiAf)-RFP* merge, D) *rdx-GFP*, E) *dpp(ci-HiAf)-RFP*, and F) *rdx-GFP* and *dpp(ci-HiAf)-RFP* merge. In A) and C) green is *inv-GFP* expression, while in D) and F) green is *rdx-GFP* expression. In B), C), D), and E), red is *dpp(ci-HiAf)-RFP* expression. In C) and F) yellow is co-expression between either *inv-GFP* or *rdx-GFP* and *dpp(ci-HiAf)-RFP*. The nuclei are gray. In all panels anterior is to the left, ventral is at the top, posterior is to the right, and dorsal is at the bottom. *dpp(ci-HiAf)-RFP* is just along the A-P compartmental boundary. Notice the green staining along the A-P compartmental boundary in A) and D). Also, note the yellow staining in C) and F). Transgenic fly experiments were performed by Dr. Dave Parker in Dr. Scott Barolo's laboratory (Department of Cell and Developmental Biology, University of Michigan).

Table IV.1. Predicted Gli/Ci binding sites used in the Topo-TX analysis.

Site sequence	Nucleotide substitutions[1]	% core similarity	% matrix similarity
GACCACCCA	0	100	100
GACCCCCCA	1	78.347	91.841
GACCTCCCA	1	77.755	91.618
GGCCACCCA	1	100	91.502
GCCCACCCA	1	100	91.492
GACCGCCCA	1	77.06	91.356
GACCACACA	1	100	90.674
GACCACCTA	1	100	89.686
GACCACCAA	1	100	89.663
GACCACCGA	1	100	89.519
GACCACCCC	1	100	88.859
GACCACCCT	1	100	88.859
GACCACCCG	1	100	88.778
AACCACCCA	1	100	88.214
TACCACCCA	1	100	88.202
CACCACCCA	1	100	88.105
GGCCCCCCA	2	78.347	83.343
GCCCCCCCA	2	78.347	83.333
GGCCTCCCA	2	77.755	83.12
GCCCTCCCA	2	77.755	83.11
GGCCGCCCA	2	77.06	82.858
GCCCGCCCA	2	77.06	82.848
GACCCCACA	2	78.347	82.515
GACCTCACA	2	77.755	82.292
GGCCACACA	2	100	82.177
GCCCACACA	2	100	82.167
GACCGCACA	2	77.06	82.03
GACCCCCTA	2	78.347	81.526
GACCCCCAA	2	78.347	81.504
GACCCCCGA	2	78.347	81.36
GACCTCCTA	2	77.755	81.303

Site sequence	Nucleotide substitutions[1]	% core similarity	% matrix similarity
(Table IV.1 continued)			
GACCTCCAA	2	77.755	81.281
GGCCACCTA	2	100	81.188
GCCCACCTA	2	100	81.178
GGCCACCAA	2	100	81.166
GCCCACCAA	2	100	81.156
GACCTCCGA	2	77.755	81.137
GACCGCCTA	2	77.06	81.041
GGCCACCGA	2	100	81.021
GACCGCCAA	2	77.06	81.019
GCCCACCGA	2	100	81.011

[1]From the optimal consensus.

Table IV.2. Analysis of Gli/Ci binding sites in four known Hh-regulated enhancers.

Site	Sequence	Nucleotide substitutions	% core similarity	% matrix similarity	Annotation
ptc1	**GACCACCCA**	**0**	**100**	**100**	**Strongest *in vitro* binding[1]**
ptc2	**GACCACCCA**	**0**	**100**	**100**	**Strong *in vitro* binding[1]**
ptc3	**GACCACCCA**	**0**	**100**	**100**	**Strong *in vitro* binding[1]**
dpp1	**GGCCACCTA**	**2**	**100**	**81.2**	**Weak *in vitro* binding[1]**
dpp2	GACCGCCCG	2	81.7	80.1	Moderate *in vitro* binding[1]
dpp3	GACCTCCCC	2	82.2	80.5	Little or no *in vitro* binding[1]
wg1	GTCCACGCT	3	100	69.7	Little or no *in vitro* binding[1]
wg2	**GACCTCCCA**	**1**	**82.2**	**91.6**	**Weak *in vitro* binding[1]**
wg3	GAGCAGCCA	2	46.3	74.7	Little or no *in vitro* binding[1]
wg4	GTTCACGCA	3	74.2	68.7	Moderate *in vitro* binding[1]
hairy1	GACCACCAT	2	100	78.5	Weak *in vivo* function [4]
hairy1	**GACCTCCCA**	**1**	**82.2**	**91.6**	**Strong *in vivo* function [4]**

[1]Personal communication, Dr. Scott Barolo, Cell and Developmental Biology, University of Michigan

Table IV.3. Orthologous regions of significant Gli/Ci clustering in two divergent Drosophila species. Bold regions are previously known Hh-regulated enhancers, while underlined regions are novel Hh-regulated enhancers confirmed in this study.

Nearest annotated gene	*D. melanogaster* coordinates (dm3)	*D. pseudoobscura* coordinates (dp3)
18w	chr2R:15,942,280-15,944,765	chr3:13,041,031-13,045,198
ac	chrX:271,178-273,582	chrXL_group1e:4,244,990-4,248,028
Actn	chrX:1,927,416-1,927,424	chrXL_group1a:2,534,873-2,535,008
beat-IV	chr3R:19,406,564-19,407,130	chr2:2,857,832-2,858,310
CG17341	chr2L:14,138,121-14,138,129	chr4_group4:5,505,850-5,507,590
CG5036	chr2R:13,680,860-13,681,982	chr3:16,462,330-16,462,891
CG8701	chr2R:4,269,955-4,272,576	chr3:2,232,782-2,236,006
Dot	chr2L:3,612,212-3,614,251	chr4_group3:5,111,713-5,113,669
dpp[1]	chr2L:2,467,642-2,470,099	chr4_group4:1,742,911-1,745,729
Frq2	chrX:18,086,163-18,086,189	chrXL_group3a:55,279-55,299
gsb	chr3:10,036,940-10,037,990	chr2R:20,952,508-20,953,535
heph	chr3R:27,705,343-27,708,738	chr2:14,965,397-14,968,428
hth	chr3R:6,416,903-6,417,440	chr2:7,012,120-7,013,078
htl	chr3R:13,885,074-13,885,097	chr2:28,877,667-28,877,695
inv	chr2R:7,379,044-7,381,376	chr3:4,545,120-4,547,216
Mrtf	chr3L:2,716,608-2,717,636	chrXR_group5:94,297-95,278
Or65a	chr3L:6,291,390-6,293,089	chrXR_group3a:329,548-331,399
osp	chr2L:14,684,404-14,685,171	chr4_group1:3,152,528-3,154,008
Pk	chr2R:3,055,985-3,061,029	chr3:1,332,897-1,338,014
ps	chr3R:5,268,610-5,270,365	chr2:3,111,760-3,113,925
ptc	**chr2R:4,533,170-4,536,527**	**chr3:13,642,912-13,646,107**
Ptx1	chr3R:26,734,929-26,734,937	chr2:431,845-432,440
rdx	chr3R:9,815,395-9,816,961	chr2:13,535,966-13,537,231
SrpRbeta[2]	**chr3L:8,700,707-8,700,731**	**chrXR_group8:8,871,522-8,871,546**
tara	chr3R:12,070,577-12,074,288	chr2:25,218,494-25,224,094

[1]This region near *dpp* is located in a well-characterized domain of cis-regulatory sequence [19].

[2]This region is the *hairy* enhancer.

Table IV.4. Analysis of predicted Gli/Ci binding sites in two novel Hh-regulated enhancers.

Site	*D. melanogaster* coordinates (dm3)	Sequence	Nucleotide substitutions	% core similarity	% matrix similarity	*D. melanogaster* maximum CC
***inv* intronic enhancer (chr2R:7,378,778-7,381,476)**						
inv1	chr2R:7,378,828-7,378,836	GACCACCCA	0	100	100	27
inv2	chr2R:7,379,044-7,379,052	GACCACACA	1	100	90.7	27
inv3	chr2R:7,379,218-7,379,226	GGCCACCCA	1	100	91.5	27
inv4	chr2R:7,379,440-7,379,448	TACCACCCA	1	100	88.2	27
inv5	chr2R:7,380,187-7,380,195	GACCACACA	1	100	90.7	27
inv6	chr2R:7,380,383-7,380,391	GACCACCTA	1	100	89.7	27
inv7	chr2R:7,380,446-7,380,454	GACCACCCT	1	100	88.9	27
inv8	chr2R:7,381,338-7,381,346	GACCACCCA	0	100	100	21
inv9	chr2R:7,381,368-7,381,376	GACCACACA	1	100	90.7	21
***rdx* intronic enhancer (chr3R: 9,815,295-9,817,061)**						
rdx1	chr3R:9,815,395-9,815,403	GGCCACACA	2	100	82.2	2.7
rdx2	chr3R:9,816,122-9,816,130	GGCCACCAA	2	100	81.2	2.7
rdx3	chr3R:9,816,953-9,816,961	AACCACCCA	1	100	88.2	2.7

IV.5. Attribution

The first draft of this chapter was written by Aaron Udager. The text was subsequently revised with suggestions from Dr. Deborah Gumucio, Dr. Linda Samuelson, and Dr. Scott Barolo.

IV.6. Publication

We anticipate that text from this chapter will be revised to form a manuscript for publication.

IV.7. References

1. Jiang, J., and Hui, C.C. (2008). Hedgehog signaling in development and cancer. Dev Cell *15*, 801-812.

2. Alexandre, C., Jacinto, A., and Ingham, P.W. (1996). Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. Genes Dev *10*, 2003-2013.

3. Muller, B., and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. Development *127*, 2999-3007.

4. Kwon, C., Hays, R., Fetting, J., and Orenic, T.V. (2004). Opposing inputs by Hedgehog and Brinker define a stripe of hairy expression in the Drosophila leg imaginal disc. Development *131*, 2681-2692.

5. Von Ohlen, T., and Hooper, J.E. (1997). Hedgehog signaling regulates transcription through Gli/Ci binding sites in the wingless enhancer. Mech Dev *68*, 149-156.

6. Arnone, M.I., and Davidson, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development *124*, 1851-1864.

7. Bellen, H.J., O'Kane, C.J., Wilson, C., Grossniklaus, U., Pearson, R.K., and Gehring, W.J. (1989). P-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev *3*, 1288-1300.

8. Johnson, K.D., and Bresnick, E.H. (2002). Dissecting long-range transcriptional mechanisms by chromatin immunoprecipitation. Methods *26*, 27-36.

9. Rebeiz, M., Reeves, N.L., and Posakony, J.W. (2002). SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. Proc Natl Acad Sci U S A *99*, 9888-9893.

10. Markstein, M., Markstein, P., Markstein, V., and Levine, M.S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. Proc Natl Acad Sci U S A *99*, 763-768.

11. Berman, B.P., Nibu, Y., Pfeiffer, B.D., Tomancak, P., Celniker, S.E., Levine, M., Rubin, G.M., and Eisen, M.B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc Natl Acad Sci U S A *99*, 757-762.

12. Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell *124*, 47-59.

13. Vokes, S.A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

14. Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. PLoS Biol *3*, e93.

15. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res *23*, 4878-4884.

16. Manly, B.F.J. (2007). Randomization, bootstrap, and Monte Carlo methods in biology, 3rd Edition, (Boca Raton, FL: Chapman & Hall/ CRC).

17. Swanson, C.I., Hinrichs, T., Johnson, L.A., Zhao, Y., and Barolo, S. (2008). A directional recombination cloning system for restriction- and ligation-free construction of GFP, DsRed, and lacZ transgenic Drosophila reporters. Gene *408*, 180-186.

18. Hallikas, O., and Taipale, J. (2006). High-throughput assay for determining specificity and affinity of protein-DNA binding interactions. Nat Protoc *1*, 215-222.

19. Blackman, R.K., Sanicola, M., Raftery, L.A., Gillevet, T., and Gelbart, W.M. (1991). An extensive 3' cis-regulatory region directs the imaginal disk expression of decapentaplegic, a member of the TGF-beta family in Drosophila. Development *111*, 657-666.

20. Berman, B.P., Pfeiffer, B.D., Laverty, T.R., Salzberg, S.L., Rubin, G.M., Eisen, M.B., and Celniker, S.E. (2004). Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome Biol *5*, R61.

21. Coleman, K.G., Poole, S.J., Weir, M.P., Soeller, W.C., and Kornberg, T. (1987). The invected gene of Drosophila: sequence analysis and expression studies reveal a close kinship to the engrailed gene. Genes Dev *1*, 19-28.

22. Kent, D., Bush, E.W., and Hooper, J.E. (2006). Roadkill attenuates Hedgehog responses through degradation of Cubitus interruptus. Development *133*, 2001-2010.

23. Kassis, J.A. (1990). Spatial and temporal control elements of the Drosophila engrailed gene. Genes Dev *4*, 433-443.

24. Drees, B., Ali, Z., Soeller, W.C., Coleman, K.G., Poole, S.J., and Kornberg, T. (1987). The transcription unit of the Drosophila engrailed locus: an unusually small portion of a 70,000 bp gene. EMBO J *6*, 2803-2809.

25. Gustavson, E., Goldsborough, A.S., Ali, Z., and Kornberg, T.B. (1996). The Drosophila engrailed and invected genes: partners in regulation, expression and function. Genetics *142*, 893-906.

26. Hidalgo, A. (1998). Growth and patterning from the engrailed interface. Int J Dev Biol *42*, 317-324.

27. Guillen, I., Mullor, J.L., Capdevila, J., Sanchez-Herrero, E., Morata, G., and Guerrero, I. (1995). The function of engrailed and the specification of Drosophila wing pattern. Development *121*, 3447-3456.

28. Blair, S.S. (1992). Engrailed expression in the anterior lineage compartment of the developing wing blade of Drosophila. Development *115*, 21-33.

29. Blair, S.S., and Ralston, A. (1997). Smoothened-mediated Hedgehog signalling is required for the maintenance of the anterior-posterior lineage restriction in the developing wing of Drosophila. Development *124*, 4053-4063.

30. Strigini, M., and Cohen, S.M. (1997). A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. Development *124*, 4697-4705.

31. Zhang, Q., Zhang, L., Wang, B., Ou, C.Y., Chien, C.T., and Jiang, J. (2006). A hedgehog-induced BTB protein modulates hedgehog signaling by degrading Ci/Gli transcription factor. Dev Cell *10*, 719-729.

32. Tenzen, T., Allen, B.L., Cole, F., Kang, J.S., Krauss, R.S., and McMahon, A.P. (2006). The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. Dev Cell *10*, 647-656.

33. Liu, J., Ghanim, M., Xue, L., Brown, C.D., Iossifov, I., Angeletti, C., Hua, S., Negre, N., Ludwig, M., Stricker, T., et al. (2009). Analysis of Drosophila segmentation network identifies a JNK pathway factor overexpressed in kidney cancer. Science *323*, 1218-1222.

CHAPTER V

CELL-SPECIFIC GENE EXPRESSION: MECHANISMS AND CONSEQUENCES

In this thesis, I have presented data from my investigations of the mechanisms of pylorus morphogenesis and the identification of Hh-regulated enhancers. Here, I will expand on the significance of these findings and outline potential future areas of research.

V.1. Evolutionary change and genetic redundancy in the GI tract

In Chapter II, I described the antral, pyloric, and duodenal transcriptomes at two critical time periods during pylorus development (E14.5 and E16.5). I showed that, at E14.5, global gene expression is largely similar in the antrum and duodenum, while two days later, at E16.5, there is a dramatic upregulation of hundreds of genes, specifically in duodenal epithelium. We call this event *intestinalization* because these genes are associated with the identity and function of mature intestinal epithelium. It is remarkable that this maturation of the tubular gut occurs so late in gut development, well after the accessory organs, such as liver and pancreas, are well established. It is also striking that this event is limited to the intestine; the stomach does not show an equivalent up-regulation of factors associated with stomach identity until 24-48 hours later.

Concomitant with intestinalization, I identified several duodenal epithelial transcription factors (i.e., *Tcfec*, *Creb3l3*, and *Hnf4γ*) that are upregulated dramatically at E16.5 and may be involved in the intestinalization event. I found that levels of Hh but not Wnt signaling are significantly altered in intestinal mesenchyme during the intestinalization event, suggesting that down-regulation of Hh signaling is important for intestinalization. Finally, I described a unique expression domain centered at the pylorus and identified several novel pyloric genes, including the zinc finger transcription factor *Gata3* and the TGF-β modulator *nephrocan*.

The further study of these pyloric-specific genes identified in this study may improve our understanding of clinical syndromes such as Infantile Hypertrophic Pyloric Stenosis and duodenogastric reflux. Such studies are well underway, beginning with the analysis of *Gata3* function in Chapter III. Another interesting aspect of our transcriptional analysis, however, is its implications for our appreciation of the events involved in the evolution of the GI tract. The transcriptome of an organ reflects the transitory culmination of evolutionary co-adaptation to changing organismal requirements and environmental stimuli, and therefore, it is of interest to consider the ontogeny of an organ within its evolutionary history.

The earliest and most primitive "gut" within the *Animalia* kingdom was formed by the invagination of epithelial cells in primitive multicellular sponges, which created a sealed pocket for extracellular digestion [1]. The first tubular gut (i.e., a digestive tract with both mouth and anus) appeared in the *Bilateria* subregnum and, thus, predates the split

between vertebrates and invertebrates. The accessory organs of the vertebrate GI tract were acquired later, and interestingly, the stomach also appears to be a relatively recent evolutionary adaptation [2]. Indeed, our studies indicate that the development of stomach identity within the gut tube lags behind that of the intestine. In Chapter II, I showed that between E14.5 and E16.5, hundreds of genes are simultaneously up-regulated in intestinal epithelium, and these genes are associated with the adult function of the intestine. In contrast, relatively few expression changes are seen in the stomach during this time; the stomach transcriptome remains relatively "primitive" at E16.5.

At some point after E16.5, however, the stomach will also up-regulate genes associated with adult stomach function and identity. I speculated in Chapter II that the stomach region may have been "carved out" of the intestinal tube with the help of a repressor. This repressor may have insulated the stomach region of the GI tube from further development at the time of major transcriptional change in the intestine, while allowing it to be plastic enough to respond to later changes that make it uniquely a stomach. Of course, this evolutionary path seems to require two changes: (1) the expression of a repressor in the stomach domain; and (2) the activation of stomach-specific genes in this region. As a first step in this evolutionary path, it is much easier to imagine the evolutionary advantage of (2) than (1). It will be of interest to carefully examine the transcriptomes of the distal foregut in organisms that lack a stomach to determine whether, indeed, a rudimentary pattern for stomach-like gene expression is present. If so, this may suggest that the later evolution of an intestinal repressor expression domain in this territory could help specialize the function of the rudimentary stomach.

Another striking finding not discussed in Chapter II is that the intestinal transcriptome seems to exhibit a tremendous genetic redundancy. This may reflect the fact that the intestine, as the primary site for nutrient absorption and one of the few common organs of the GI tract among *Bilateria*, is essential for the survival of an organism [1]. As such, there is likely strong evolutionary pressure to preserve intestinal function, and this may explain the genetic redundancy of transcription factors and signaling molecules involved in intestinal development. Paralogous genes commonly arise by duplication of a common ancestral gene, and therefore, it is not unusual for their expression patterns to be similar [3]. However, many paralogs eventually evolve unique cis-regulatory elements and distinct tissue- or cell-specific gene expression patterns. The intestine is unusual in this manner because many paralogs share similar expression patterns. Two of the three members of the paralogous group of Cdx transcription factors (i.e., *Cdx1*, *Cdx2*, and *Cdx4*), for example, have similar, overlapping expression domains in intestinal epithelium [4]. Additionally, several other families of transcription factors (i.e., Hox, Fox, Gata, and Hnf4) and signaling molecules (i.e., Hh, Bmp, Notch, and Wnt) have paralogous members with similar expression patterns in the intestine ([5-12]; unpublished data from our laboratory).

Single mutants for many of the intestinal genes in these paralogous groups show little obvious effect on intestinal development [13-17]. In contrast, pancreas development is completely and specifically perturbed in single mutants for *Pdx1* [18]. This is intriguing because the pancreas develops precisely within a narrow domain of *Pdx1* expression, and

Pdx1 does not have any paralogs. Also, the pancreas is a more recently acquired GI tract organ than the intestine. Together, these observations suggest that the duplication of critical intestinal developmental genes during *Bilateria* evolution may be a mechanism to preserve essential intestinal functions during GI tract development.

V.2. The establishment and maintenance of differentiated gastric epithelium

Thus far, I have focused primarily on the dramatic changes that take place in intestinal epithelium at E16.5. As mentioned above, dramatic changes in the intestinal transcriptome appear to promote intestinal identity, but a similar large change in the stomach transcriptome does not occur. However, concomitant with the intestinalization event described in Chapter II, gastric *Sox2* expression regresses anteriorly, such that at E18.5 it is excluded from the future glandular epithelium of the stomach [19]. Prior to E16.5, mutually repressive interactions between Cdx2 and Sox2 form a soft boundary between presumptive intestinal and gastric epithelium at the pylorus. However, *Cdx2* expression does not expand anteriorly with *Sox2* at E16.5; rather, its anterior expression boundary remains at the pylorus, where a single cell-thick boundary forms between intestinal and gastric epithelium (i.e., the epithelial pyloric border) ([20]; unpublished data from our laboratory). What then accounts for the anterior regression of *Sox2* between E14.5 and E18.5?

One possibility is that *Sox2* expression is regulated directly by specific signaling pathways, and changes in activity levels of these pathways along the A-P axis of the

stomach influence the *Sox2* expression domain. In this respect, it is interesting to note that inhibition of Wnt signaling plays an important role in epithelial patterning of the stomach, and that interference with inhibition of Wnt signaling (*Barx1* null mice) shows an anterior shift of Cdx2 expression [21, 22]. However, these foregut phenotypes are apparent prior to intestinalization, as early as E11.5 [21]. Furthermore, in Chapter II, I reported that there is no gradient of canonical Wnt pathway activity between the antrum and duodenum at E16.5. This suggests that inhibition of Wnt signaling in the stomach may have an indirect rather than direct effect on *Sox2* and *Cdx2* expression boundaries.

Besides Wnt, other signaling pathways (i.e., BMP and Fgf) may control *Sox2* and *Cdx2* expression domains in the posterior foregut. Loss of *Bapx1*, for example, is associated with epithelial patterning defects in the antrum [23]. Bapx1 is expressed in posterior stomach mesenchyme, where it prevents the activation of *Bmp4* by endodermal Hh signaling in this region of the stomach [24]. In the chicken, epithelial patterning of the pylorus is controlled by the Bmp antagonist *gremlin* [25]. Intriguingly, in Chapter II, I showed that at E16.5 the pyloric epithelium expresses *nephrocan*, an inhibitor of TGF-β signaling. These data suggest that inhibition of TGF-β superfamily signaling may regulate epithelial patterning in the distal stomach and/or pylorus by directly repressing *Sox2*. Thus, it would be interesting to determine whether sustained, ectopic expression of TGF-β superfamily ligands (i.e., *Bmp4*) prevents *Sox2* regression.

Abnormal Fgf signaling also disrupts normal stomach development [26]. Fgf10 is expressed throughout the mesenchyme of the glandular stomach and signals to fibroblast

growth factor receptor 2 (Fgfr2) in antral epithelium. Forced expression of *Fgf10* in the *Pdx1*-positive epithelial domain of the posterior foregut alters the epithelial architecture and relative distribution of enteroendocrine cell types; it is also associated with decreased levels of *Sox2* gene expression. Interestingly, in Chapter II, I reported that the intracellular Fgf inhibitor *sprouty-related, EVH1 domain containing 2* (*Spred2*) is specifically enriched in pyloric tissue at E16.5. This suggests that inhibition of Fgf signaling may control epithelial patterning at the pylorus, while perhaps Fgf10 is required to downregulate *Sox2* in the distal stomach. It would be interesting to determine whether endoderm-specific loss of *Fgfr2* prevents *Sox2* regression.

Another possibility is that an unknown transcriptional repressor of *Sox2* is activated in the future epithelial domain of the glandular stomach, and intriguingly, this repressor may also be involved in limiting the anterior expression of *Cdx2* to the epithelial pyloric border. The data I presented in Chapter II raises the possibility that transcription factors are enriched specifically in the glandular epithelium of the stomach at E16.5. In this respect the expression of Sox21, a transcriptional repressor, is noteworthy. *Sox2* and *Sox21* are expressed in overlapping domains of chicken foregut endoderm and have antagonistic, counterbalancing roles in neurogenesis [27, 28]. While *Sox2* expression in undifferentiated neural cells suppresses neurogenesis, forced *Sox21* expression promotes neurogenesis [27]. In addition, *Sox21* expression at the midbrain-hindbrain boundary is dependent on Fgf signaling [29]. I have not investigated the expression pattern of *Sox21* during intestinalization in these studies, and furthermore, it is unclear whether Sox21 can directly regulate *Sox2* expression. However, it is tempting to speculate that Sox21 is

downstream of Fgf10 signaling and may be involved in establishing the future domain of gastric epithelium, perhaps by repressing *Sox2*. It would be interesting to examine *Sox21* null mice for defects in epithelial patterning of the stomach.

Cdx2 and *Sox2* are excluded from the epithelium of the glandular stomach in mice. This suggests that they may be actively repressed in stomach epithelial cells, and furthermore, that this repression may be critical for gastric epithelial homeostasis. Intestinal metaplasia (IM) of the stomach, which is an intermediate step in the pathogenesis of some types of gastric cancer, involves the patchy conversion of gastric epithelium to an intestine-like epithelium [30]. Ectopic expression of *Cdx2* is associated with IM of the stomach in humans, and forced expression of *Cdx2* in gastric epithelium (FoxA3-Cdx2) in mice upregulates intestinal epithelial genes [31, 32]. Interestingly, in Chapter II, I identified several transcription factors that are upregulated specifically in duodenal epithelium during the intestinalization event at E16.5. In particular, it would be interesting to determine whether, similar to *Cdx2*, forced expression of *Tcfec* or *Hnf4*γ in gastric epithelium is able to activate intestinal gene expression.

While *Sox2* expression is clearly restricted from glandular stomach epithelium in mice, it may be expressed in normal gastric mucosa of humans and downregulated in IM of the stomach [33, 34]. This suggests that a novel *Sox2*-independent mechanism of stomach epithelial differentiation may have evolved specifically in rodents after the split from primates. To examine whether this species-specific difference is functionally relevant, it would be interesting to ectopically express *Sox2* throughout the endoderm. This

experiment would provide data for two key questions. Does ectopic *Sox2* expression in the glandular epithelium of the stomach prevent the establishment of gastric epithelial identity? What is the consequence of *Cdx2* and *Sox2* co-expression for epithelial identity in the intestine? The expectation is that ectopic *Sox2* would be sufficient to drive esophageal identity in the glandular stomach.

V.3. Is the pylorus a tissue organizer?

The pylorus is a unique structure in the posterior foregut of the GI tract: the posterior foregut is the source for hepatic, pancreatic and splenic precursors; the pylorus forms a complex smooth muscle sphincter; and after intestinalization, the pylorus is the site of the epithelial pyloric border, a one cell-thick boundary between duodenal and antral epithelium. How does such a small region contribute to so many diverse structures across a number of embryonic time points? One hypothesis is that the pylorus is a tissue organizer. In Chapter II, I describe a novel domain of epithelial and mesenchymal gene expression at the pylorus at E14.5 and E16.5. Interestingly, the epithelial pyloric border exists within this unique expression domain, which for some genes extends into both the antrum and the duodenum (i.e., *Pdx1* and *nephrocan*). In this way, the pylorus is strikingly similar to another border between two organs, the isthmus between the midbrain and hindbrain. The parallels here are interesting and worth a few paragraphs of further explanation, since the isthmus is known to have organizer properties.

During development, the brain is divided into three primary vesicles (from anterior to posterior): prosencephalon (forebrain), mesencephalon (midbrain), and rhombencephalon (hindbrain). The midbrain region eventually forms the tectum, while the hindbrain domain becomes the cerebellum. The isthmus is a region of unique gene expression that contains a one cell-thick boundary (i.e., the isthmic border) separating the midbrain and hindbrain territories [35, 36]. Expression of the homeodomain transcription factors engrailed 1 (En1) and En2 extends across the isthmic border in a decreasing P-A gradient (expression is highest posteriorly and lowest anteriorly). Within this En1/2 expression domain, two other homeodomain transcription factors, orthodenticle homolog 2 (Drosophila) (Otx2) and gastrulation brain homeobox 2 (Gbx2), pre-pattern the midbrain and hindbrain; *Otx2* is expressed in the midbrain, while *Gbx2* is expressed in the hindbrain. The expression domains of *Otx2* and *Gbx2* initially overlap at the isthmic border. However, subsequent mutual repressive interactions refine the expression domains, such that they abut but no longer overlap. In addition, the isthmus secretes Fgf8, which via activation of the Ras-Erk signaling pathway both stimulates *En1/2* expression and restricts *Otx2* expression anteriorly, to position the isthmic border.

Transplantation experiments have demonstrated that the isthmus itself is a tissue organizer [35, 36]. When isthmic tissue is transplanted either anteriorly, into the forebrain, or posteriorly, into the hindbrain, it retains its normal fate. In contrast, when forebrain tissue is transplanted next to the isthmus, the transplanted tissue changes fate according to its position relative to the isthmus. These are the classic properties of a tissue organizer [37]. Interestingly, exogenous application of Fgf8-coated beads to

forebrain tissue replicates the consequences of isthmus transplantation, which strongly suggests that Fgf8 is a primary organizing molecule at the isthmus [38].

Interestingly, as a tissue organizer, the isthmus is essential for its own development, as well as proper patterning of the tissue surrounding it. Germline loss of *Otx2* is associated with aberrant forebrain, midbrain, and anterior hindbrain structures, while loss of *Gbx2* leads to abnormal development of the midbrain and anterior hindbrain; loss of either gene altered the expression domain of *Fgf8* at the isthmus [39, 40]. As expected, *Fgf8* hypomorphs show substantial defects in midbrain-hindbrain patterning [41]. However, targeted inactivation of *Fgf8* also eliminated *Gbx2* expression, indicating the feedback regulatory role of the isthmic organizer [42].

Many similarities exist between the pylorus and the isthmus: (1) Transcription factors that specify distinct tissues (i.e., *Otx2*/*Gbx2*, *Cdx2*/*Sox2*) are initially expressed in overlapping domainsl; (2) the expression overlap eventually resolves into a one-cell-thick tissue boundary via mutually repressive interactions; and (3) signaling factors are uniquely expressed within a narrow domain that includes the tissue boundary. This suggests that the pylorus may be a tissue organizer. Transplantation studies, analogous to those performed in the brain, have not been conducted in the GI tract to determine the tissue organizing potential of the mouse pylorus. Here, I will explore some of these experiments and the expected outcomes. The most straightforward approach would be tissue transplantation between cultured gut explants *in vivo*. Tissue from E10.5 to E14.5 ROSA26 (i.e., β-gal-marked) embryos could be transplanted into stage-matched wild

type guts, in order to lineage trace the cells from the transplanted tissue. These explants could then be cultured together for several days before analysis (developmental growth slows down in culture).

First, after the pylorus is removed from a wild type gut, the antrum and duodenum could be placed together and cultured without an intervening pylorus. Would the *Sox2*/*Cdx2* pattern be perturbed? Would an epithelial pyloric border form? Would *Nkx2-5* and *Gata3* expression be lost? Would a pyloric sphincter develop? If the pylorus is a tissue organizer, the expectation is that the epithelial pyloric border would not form without it. Similarly, unless pyloric sphincter morphogenesis was independent of the organizing activity of the pylorus, it would be unlikely for the gut to develop a pyloric sphincter if the pyloric region was extirpated.

Second, a narrow region of pyloric tissue could be transplanted in the opposite orientation, such that *Sox2*-positive cells are posterior to *Cdx2*-positive cells at the pylorus. Where would the epithelial pyloric border form? Would there be more than one border? If the pylorus is a tissue organizer, the expectation may be that the *Cdx2*- and *Sox2*-positive cells resolve by themselves into a single epithelial pyloric border, similar to anterior and posterior compartment cells in the *D. melanogaster* larval imaginal wing disc [43]. In that case, it would be interesting to examine what happens to the cells that were formerly *Sox2*-positive. Would they die? Would they stop expressing *Sox2* and upregulate *Cdx2*? Would they migrate anteriorly past the *Cdx2*-positive cells? The questions of *Sox2*-positive cell lineages could be examined with transplanted tissue from

a *Sox2-Cre* and ROSA26R intercross. Alternatively, based on the mutual repressive interactions between Sox2 and Cdx2, it is possible that three borders may form in this experiment, such that there are reciprocal, isolated *Sox2*- and *Cdx2*-positive domains. In that case, it would be interesting to determine what types of epithelial cells were specified between them. For example, would *Sox2*-positive cells between two *Cdx2*-positive domains become stomach, esophagus, etc? In that case, the expectation is that the *Sox2*-positive cells would become esophagus.

Third, that same region of pyloric tissue could be transplanted at increasing posterior distance from the pylorus, in the correct orientation. Would there be extra epithelial pyloric borders? Would another pyloric sphincter develop? Is there a posterior limit to either border formation or sphincter development? If the pylorus is a tissue organizer, the expectation is that multiple epithelial pyloric borders and an extra pyloric sphincter would be formed from the transplanted tissue. Will there be a posterior limit to the organizing activity of the pylorus (a region of competency to conform to organizer instructions)?

Finally, in lieu of tissue transplantation, gut explants could be cultured with beads soaked in signaling ligands (i.e., Bmp and Fgf) and/or infected with viral expression vectors for intracellular transcription factors (i.e., *Nkx2-5*). After removal of the pylorus, would exogenous Bmp4 and/or *Nkx2-5* be sufficient for border formation and sphincter development? What about in the posterior intestine? Does Fgf8 have organizing activity in the GI tract? If the pylorus is a tissue organizer, the expectation is that some

combination of exogenous ligand and forced transcription factor expression should be able to recapitulate its organizing activity and generate of epithelial pyloric border and sphincter.

V.4. Transcriptional control of pyloric Nkx2-5 and Gata3 expression

Thus far, I have focused primarily on the descriptive analysis of *Nkx2-5* and *Gata3* expression at the pylorus. However, elucidating the regulatory relationships of these factors is necessary to more fully understand the function of *Gata3* during pylorus morphogenesis. This includes knowledge of how *Gata3* and *Nkx2-5* interact with one another, in addition to how each factor relates to the other genes of the pylorus developmental network (see Chapter I.3). An impediment to this goal, however, is the lack of a rigorous descriptive analysis of cell populations within the posterior foregut. In Chapter III, I described the immunohistological analysis of pyloric smooth muscle cell populations, and then I investigated the co-localization of Nkx2-5 and Gata3 with these cells. However, there are a variety of other cell types at the pylorus, and few previous studies of pylorus development have focused on defining cell type-specific expression patterns for pyloric genes. With this caveat in mind, I will describe the potential regulatory relationships for *Gata3* and *Nkx2-5* at the pylorus.

Given its early onset and narrow domain, the pyloric expression of *Nkx2-5* is likely to be tightly regulated by a combination of transcription factors, including Hox proteins, and Bmp signaling from the surrounding tissues. Given this, the expression of the related

gene *Nkx2-3* is interesting. *Nkx2-3* is expressed throughout the loose mesenchyme of the intestine. At its anterior expression boundary, which lies just short of the pylorus, it abuts the *Nkx2-5* expression domain [44]. This observation raises the possibility of a regulatory relationship between *Nkx2-3* and *Nkx2-5*. Intriguingly, misexpression of a synthetic Nkx2-5 fusion protein carrying a strong repressor domain downregulates endogenous *Nkx2-5* expression at the pylorus and shifts the *Nkx2-3*/*Nkx2-5* expression boundary anteriorly [45]. This suggests that Nkx2-5 may be able to autoregulate its own expression and, additionally, that the boundaries of the *Nkx2-3* domain may be set by Nkx2-5. Does Nkx2-5 control *Nkx2-3* expression? Does Nkx2-3 regulate *Nkx2-5* expression? To probe these potential regulatory relationships, it would be interestingly to examine *Nkx2-5* expression in *Nkx2-3* null embryos (and *vice versa*). In the absence of *Nkx2-3*, what is the posterior boundary of *Nkx2-5* expression? Will the anterior boundary of *Nkx2-3* expression shift anteriorly without *Nkx2-5*, as suggested by the Smith study? Alternatively, it would be interesting to explore the consequences of ectopic *Nkx2-3* expression at the pylorus. Would *Nkx2-5* expression be extinguished?

It is not clear what factors may limit the *Nkx2-5* domain on the stomach side. Loss of either *Bapx1* or *Six2* results in the absence of the pyloric constriction, which places these factors near the top of the hierarchy of pylorus development. *Bapx1* and *Six2* are expressed in apparently overlapping domains with *Nkx2-5* at the pylorus [46, 47]. In contrast to *Nkx2-5*, the expression of *Six2* and *Bapx1* extends anteriorly into the posterior stomach mesenchyme [2, 46, 47]. These patterns suggest that *Bapx1* and/or *Six2* may specify a zone for pylorus development in the posterior foregut. Interestingly, Bapx1 and

Six2 inhibit *Bmp4* expression in the posterior stomach [24, 46]. Thus, in contrast to surrounding tissue, part of the *Bapx1/Six2* expression domain does not express *Bmp4*. Instead, it contains cells that are competent to respond to Bmp4 signals and activate *Nkx2-5* [45, 48]. Intriguingly, neither the loss of *Bapx1* nor *Six2* extinguishes the expression of *Nkx2-5* [23, 46, 47]. These results are surprising because the pyloric phenotype of *Bapx1* and *Six2* null embryos is not due to a complete loss of *Nkx2-5* expression. As discussed in Chapter III, these observations may reflect an effect on *Gata3*, downstream of *Nkx2-5*. Alternatively, it raises the possibility that *Bapx1* and *Six2* are partially functionally redundant. It would be interesting to examine pylorus development and *Nkx2-5* expression in *Bapx1*/*Six2* compound null embryos. What is the pyloric phenotype of these embryos? Would *Nkx2-5* expression be extinguished? How is the expression of *Bmp4* affected?

In Chapter III, I presented several lines of circumstantial evidence for a regulatory relationship between *Gata3* and *Nkx2-5*. The expression patterns of these genes are highly similar, with only small differences in the timing of onset (i.e., *Nkx2-5* is earlier) and breadth of expression domain (i.e., *Nkx2-5* is broader). Gata3 and Nkx2-5 are both expressed in specific smooth muscles at the pylorus, including the outer longitudinal muscle, dorsal nodule, and ventral pyloric cords. I showed that *Gata3* is not epistatic to *Nkx2-5*, but that Nkx2-5 co-localizes with active transcription from the endogenous *Gata3* locus. Taken together, these data strongly suggest that *Nkx2-5* is epistatic to *Gata3*. Germline *Nkx2-5* null embryos die prior to *Gata3* expression at the pylorus, but it would be interesting to examine pylorus development and *Gata3* expression after

conditional inactivation of *Nkx2-5* [49]. What is the pyloric phenotype? Would *Gata3* expression be extinguished?

In Chapter III, I also noted that a large number of Nkx2-5-positive cells do not co-express Gata3. It seems likely, therefore, that Nkx2-5 functions coordinately with other regulatory inputs to control Gata3 expression. As discussed in Chapter III, Bmp4 signaling may also have direct regulatory input on *Gata3* expression. However, if *Gata3* and *Nkx2-5* are both direct targets of Bmp4 signaling, it is unlikely that Bmp4 signaling alone can account for the differences in their expression patterns. Another potential input on *Gata3* expression is Sox9, which is expressed in a similar domain to *Nkx2-5* and *Gata3* at the pylorus [25, 46, 50]. Interestingly, while Bmp4 signaling directly activates *Sox9* expression, it does so independently of *Nkx2-5* [25]. Although the cell type-specific expression of *Sox9* has not been examined, it's tempting to speculate that Sox9 could restrict *Gata3* expression by actively repressing its transcription in a subset of *Nkx2-5*-positive cells. It would be interesting to examine the co-localization of *Nkx2-5*, *Sox9*, and *Gata3* at the pylorus. Are *Gata3* and *Sox9* co-expressed? Are *Nkx2-5* and *Sox9* co-expressed? Alternatively, it would be interesting to explore the consequences of ectopic *Nkx2-5* expression in posterior foregut mesenchyme. Can Nkx2-5 activate *Gata3* expression in other mesenchymal domains? Finally, while germline *Sox9* null embryos die prior to *Gata3* expression at the pylorus, it would be interesting to examine pylorus development and *Gata3* expression after conditional inactivation of *Sox9* [51]. What is the pyloric phenotype? Would *Nkx2-5* expressed be altered? Would *Gata3* expression be affected?

The most significant difference between *Nkx2-5* and *Gata3* pyloric expression involves splenic precursor cells (SPCs). Between E12.5 and E14.5, SPCs migrate anteriorly from the pylorus to populate the spleen [47]. *Nkx2-5*, but not *Gata3*, is expressed in these migrating cells. That pyloric Gata3 expression begins at approximately the same time as SPC migration might suggest that Gata3 acts as a lineage determination factor (i.e., switch) between SPCs and smooth muscle cells of the pylorus (i.e., $Nkx2\text{-}5^{+}/Gata3^{neg}$ cells become SPCs, while $Nkx2\text{-}5^{+}/Gata3^{+}$ cells become pyloric smooth muscle). This hypothesis predicts that loss of *Gata3* increases SPCs at the expense of smooth muscle cells, which may lead to an abnormally large spleen. However, I observed no obvious gross morphological changes in the spleen of *Gata3* null embryos. Another possibility that would be interesting to explore is that *Gata3* must be excluded from SPCs. In that case, persistent, ectopic *Gata3* expression in *Nkx2-5*-positive cells may disrupt spleen morphogenesis.

V.5. Evolution of Gli/Ci binding sites and identifying Hh-regulated enhancers

In Chapter IV, I identified novel Hh-regulated enhancers *in silico* by examining the conservation of relative Gli/Ci binding site clustering between divergent *Drosophila* species. Unlike previous computational methods for identifying Hh regulated enhancers in vertebrates, our method, Topo-TX, does not limit the area searched to regions of high inter-species sequence conservation, nor does it impose restrictions on the number, arrangement, or relative affinity of transcription factor binding sites [52, 53]. There are

only two requirements for binding sites in Topo-TX: they must have a minimum predicted affinity; and they must show significant clustering relative to other sites (on a genome-wide level). In this way, Topo-TX is very tolerant to variations in binding site number, spacing, and relative affinity within enhancer sequences. Given the nature of binding site evolution, this tolerance may increase the sensitivity of Topo-TX predictions [54]. However, even with these permissive standards, our method is likely to miss some Hh-regulated enhancers. Since relative clustering with other sites is a prerequisite for Topo-TX, enhancers that contain a single, non-clustered binding site will not be detected. In addition, enhancers with many low affinity sites may not be identified. The advent of chromatin immunoprecipitation (ChIP) strategies to study Hh enhancers provides another method to survey for transcription factor binding sites [53, 55, 56]. The merging of Topo-TX and ChIP data is likely to help us isolate and functionally dissect a large number of enhancers in the near future. This will undoubtedly teach us much about the rules by which enhancers function and evolve.

The selection of transcription factor binding sites has a significant impact on any *in silico* binding site analysis. In Chapter IV, I used a mono-nucleotide distribution matrix that was derived from *in vitro* DNA binding data for recombinant Ci protein [52, 57]. The data for this matrix was obtained from measurements of competitive protein-DNA binding reactions between oligomers containing the optimal consensus Gli/Ci site (GACCACCCA) and a similar site carrying *n* nucleotide substitution (*n* = 1-9). While for Gli/Ci there are 262,144 total 9-mers that could be tested *in vitro*, only oligomers with a single nucleotide substitution were examined (there are 28 unique mono-nucleotide

substitutions for Gli/Ci) [52, 57, 58]. The most significant advantage of using this matrix is that it is based on experimental binding data. However, it is unclear how well *in vitro* DNA binding of short oligomers with single nucleotide substitutions translates to *in vivo* DNA binding. For example, some proteins have preferences for specific di-nucleotide combinations in a binding site [58]. Thus, it would be interesting to determine Ci *in vitro* DNA binding affinity for di-nucleotide-substituted oligomers (there 121 unique di-nucleotide substitutions for Gli/Ci). The di-nucleotide distribution matrix derived from these data may allow more precision in distinguishing between 9-mers that bind or do not bind Ci protein *in vivo*. This, in turn, may improve the sensitivity and/or specificity of Topo-Tx predictions. Of course, other issues not considered in such an analysis include the possible contribution of nucleotides outside of the 9-mer core to binding affinity and the contribution of surrounding (cooperative) binding interactions. Modeling of these parameters will be assisted by the identification and functional analysis of additional enhancers.

The *Drosophila* genus has a well-developed set of phylogenomic resources for evolutionary analysis. In addition to *D. melanogaster*, the genomes of eleven related species in the *Drosophila* genus have been sequenced [59]. These twelve species (including *D. melanogaster*) can be organized into several phylogeny groups, spanning different evolutionary time periods (million years ago; MYA) [60]: *melanogaster* subgroup (6-15 MYA), *Sophophora* subgenus group (20-30 MYA), and *Drosophila* genus group (~40 MYA). Thus, the genomes of these species represent a comprehensive sample of evolution within the *Drosophila* genus, allowing assessment of both closely

and distantly related *Drosophila* species. As discussed below, while most extant tools for the identification of functional enhancers in genomic DNA utilize these rich genomic resources to extract and analyze areas of sequence conservation, other considerations (e.g., binding site clustering, binding site affinity and binding cooperativity) may be equally important.

As fundamental units of enhancers, binding sites for transcription factors are a primary source of both stability and change during evolution [54]. Indeed, evolutionary change within an enhancer can affect several properties of transcription factor binding sites: total number, spatial arrangement, and protein binding affinity. Previous studies have examined the evolution of transcription factor binding sites in related species of yeast (*Saccharomyces*) and *Drosophila* [61, 62]. In yeast, functionally important positions within a binding site had lower rates of evolution, while in *Drosophila*, there was substantial loss and gain (i.e., turnover) of binding sites, even between closely related extant species. Interestingly, despite significant changes in genomic composition and structure, gene expression patterns are often preserved between divergent *Drosophila* species [54]. This suggests that the function, but perhaps not the exact sequence, of many cell-specific enhancers are under positive selective pressure during evolution. Unpublished data from Dr. Scott Barolo's laboratory (Department of Cell and Developmental Biology, University of Michigan) indicates that the predicted protein binding affinities of sites within an enhancer are critical for its correct spatiotemporal expression domain, and for some sites, the relative affinity is preserved among related species. Importantly, maintenance of binding affinity does not require strict maintenance

of sequence. It is instructive, therefore, to examine known enhancers to determine whether indeed sequence conservation is a characteristic of most functional enhancers.

In studies not described in Chapter IV, I investigated the conservation of Gli/Ci binding site sequence in the context of four known Hh-regulated enhancers: *ptc*, *dpp*, *wg*, and *hairy* [63-66]. Fifty percent (4/8) of the binding sites in these enhancers were sequence conserved across the *Drosophila* genus group, 75% (6/8) were conserved across the *Sophophora* subgenus group, and 100% (8/8) were conserved across the *melanogaster* subgroup. If this analysis is extended to the *rdx* and *inv* enhancers identified in Chapter IV, the percentages of sequence conserved sites are: 30% (6/20) across the *Drosophila* genus group, 50% (10/20) across the *Sophophora* subgenus group, and 90% (18/20) across the *melanogaster* subgroup. In addition, each of these individual enhancers is associated with at least one Gli/Ci sites that is sequence conserved across the *melanogaster* subgroup. Although this is a small sample size of sequences, these preliminary results suggest that Gli/Ci binding sites in Hh-regulated enhancers are nearly always sequence conserved across the *melanogaster* subgroup and are often sequence conserved across the *Sophophora* sub-genus or *Drosophila* genus groups. This analysis suggests that sequence conservation of Gli/Ci binding sites across the *melanogaster* subgroup may be a useful filter for Topo-TX results.

While the use of sequence conservation as a single criterion across the entire genus may miss 70% of the functional enhancers, it is nevertheless of great interest to examine the sites that *can* be detected by such a strict filter. Genome-wide in *D. melanogaster*, there

are 154 sites that are sequence conserved across the *Drosophila* genus group. Use of the Topo-TX tool will allow us to determine which of these sites are associated with significant relative clustering; such an analysis is ongoing.

Thus far, I have focused on the application of Topo-TX to identifying Hh-regulated enhancers in *D. melanogaster*. However, our original impetus for developing this method was to identify direct targets of Hh signaling in the mouse intestine. The most significant obstacle for applying Topo-TX to the mouse genome, in the manner described in Chapter IV, is the lack of appropriate phylogenomic resources. The nearest evolutionary species with a sequenced genome is the rat, and the split between the *Rodentia* and *Primates* orders was about 80 MYA [67]. On the other hand, there are rich genomic resources for human and the rest of the order *Primates*: the Great Apes (15-20 MYA), the Old World Monkeys (25-30 MYA), and the New World Monkeys (~40 MYA) [68]. It would be interesting to compare the relative clustering of Gli/Ci binding sites in each of these Primate genomes to identify orthologus regions of significant Gli/Ci sites clustering. These putative Hh-regulated enhancers could then be tested for reporter activity in transgenic mice. Would these regions correspond to functional enhancers *in vivo*? Are direct Hh targets in primates also direct Hh targets in rodents? It is clear that the further exploration of all of these questions would provide an interesting and exciting chapter to future theses on cell-specific genes expression in GI development and homeostasis.

V.6. Attribution

The first draft of this chapter was written entirely by Aaron Udager. The text was revised with suggestions from Dr. Deborah Gumucio.

V.7. Publication

Some text from this chapter has been revised by Aaron Udager, Mr. Ajay Prakash, and Dr. Deborah Gumucio and included in a manuscript for publication as a chapter in "Molecular Biology of Digestive Organs," a forthcoming volume in the Elsevier series *Progress in Molecular Biology and Translational Science*.

V.8. References

1. Nielsen, C. (2008). Six major steps in animal evolution: are we derived sponge larvae? Evol Dev *10*, 241-257.

2. Smith, D.M., Grasty, R.C., Theodosiou, N.A., Tabin, C.J., and Nascone-Yoder, N.M. (2000). Evolutionary relationships between the amphibian, avian, and mammalian stomachs. Evol Dev *2*, 348-359.

3. Eichler, E.E., and Sankoff, D. (2003). Structural dynamics of eukaryotic chromosome evolution. Science *301*, 793-797.

4. Guo, R.J., Suh, E.R., and Lynch, J.P. (2004). The role of Cdx proteins in intestinal development and cancer. Cancer Biol Ther *3*, 593-601.

5. Drewes, T., Senkel, S., Holewa, B., and Ryffel, G.U. (1996). Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. Mol Cell Biol *16*, 925-931.

6. Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K., and Yamamura, K. (2002). Region-specific gastrointestinal Hox code during murine embryonal gut development. Dev Growth Differ *44*, 77-84.

7. Sekimoto, T., Yoshinobu, K., Yoshida, M., Kuratani, S., Fujimoto, S., Araki, M., Tajima, N., Araki, K., and Yamamura, K. (1998). Region-specific expression of murine Hox genes implies the Hox code-mediated patterning of the digestive tract. Genes Cells *3*, 51-64.

8. Friedman, J.R., and Kaestner, K.H. (2006). The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci *63*, 2317-2328.

9. Burch, J.B. (2005). Regulation of GATA gene expression during vertebrate development. Semin Cell Dev Biol *16*, 71-81.

10. Kolterud, A., Grosse, A.S., Zacharias, W.J., Walton, K.D., Kretovich, K.E., Madison, B.B., Waghray, M., Ferris, J.E., Hu, C., Merchant, J.L., et al. (2009). Paracrine Hedgehog signaling in stomach and intestine: new roles for hedgehog in gastrointestinal patterning. Gastroenterology *137*, 618-628.

11. Sancho, E., Batlle, E., and Clevers, H. (2004). Signaling pathways in intestinal development and cancer. Annu Rev Cell Dev Biol *20*, 695-723.

12. Theodosiou, N.A., and Tabin, C.J. (2003). Wnt signaling during development of the gastrointestinal tract. Dev Biol *259*, 258-271.

13. Subramanian, V., Meyer, B.I., and Gruss, P. (1995). Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. Cell *83*, 641-653.

14. Molkentin, J.D., Tymitz, K.M., Richardson, J.A., and Olson, E.N. (2000). Abnormalities of the genitourinary tract in female mice lacking GATA5. Mol Cell Biol *20*, 5256-5260.

15. Gerdin, A.K., Surve, V.V., Jonsson, M., Bjursell, M., Bjorkman, M., Edenro, A., Schuelke, M., Saad, A., Bjurstrom, S., Lundgren, E.J., et al. (2006). Phenotypic screening of hepatocyte nuclear factor (HNF) 4-gamma receptor knockout mice. Biochem Biophys Res Commun *349*, 825-832.

16. Kaestner, K.H., Katz, J., Liu, Y., Drucker, D.J., and Schutz, G. (1999). Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression in vivo. Genes Dev *13*, 495-504.

17. Kaestner, K.H., Hiemisch, H., and Schutz, G. (1998). Targeted disruption of the gene encoding hepatocyte nuclear factor 3gamma results in reduced transcription of hepatocyte-specific genes. Mol Cell Biol *18*, 4245-4251.

18. Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., and Wright, C.V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development *122*, 983-995.

19. Que, J., Okubo, T., Goldenring, J.R., Nam, K.T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development *134*, 2521-2531.

20. Braunstein, E.M., Qiao, X.T., Madison, B., Pinson, K., Dunbar, L., and Gumucio, D.L. (2002). Villin: A marker for development of the epithelial pyloric border. Dev Dyn *224*, 90-102.

21. Kim, B.M., Buchner, G., Miletich, I., Sharpe, P.T., and Shivdasani, R.A. (2005). The stomach mesenchymal transcription factor Barx1 specifies gastric epithelial identity through inhibition of transient Wnt signaling. Dev Cell *8*, 611-622.

22. Kim, B.M., Miletich, I., Mao, J., McMahon, A.P., Sharpe, P.A., and Shivdasani, R.A. (2007). Independent functions and mechanisms for homeobox gene Barx1 in patterning mouse stomach and spleen. Development *134*, 3603-3613.

23. Verzi, M.P., Stanfel, M.N., Moses, K.A., Kim, B.M., Zhang, Y., Schwartz, R.J., Shivdasani, R.A., and Zimmer, W.E. (2009). Role of the homeodomain

transcription factor Bapx1 in mouse distal stomach development. Gastroenterology *136*, 1701-1710.

24. Nielsen, C., Murtaugh, L.C., Chyung, J.C., Lassar, A., and Roberts, D.J. (2001). Gizzard formation and the role of Bapx1. Dev Biol *231*, 164-174.

25. Moniot, B., Biau, S., Faure, S., Nielsen, C.M., Berta, P., Roberts, D.J., and de Santa Barbara, P. (2004). SOX9 specifies the pyloric sphincter epithelium through mesenchymal-epithelial signals. Development *131*, 3795-3804.

26. Nyeng, P., Norgaard, G.A., Kobberup, S., and Jensen, J. (2007). FGF10 signaling controls stomach morphogenesis. Dev Biol *303*, 295-310.

27. Sandberg, M., Kallstrom, M., and Muhr, J. (2005). Sox21 promotes the progression of vertebrate neurogenesis. Nat Neurosci *8*, 995-1001.

28. Uchikawa, M., Kamachi, Y., and Kondoh, H. (1999). Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech Dev *84*, 103-120.

29. Jukkola, T., Lahti, L., Naserke, T., Wurst, W., and Partanen, J. (2006). FGF regulated gene-expression and neuronal differentiation in the developing midbrain-hindbrain region. Dev Biol *297*, 141-157.

30. Tsukamoto, T., Mizoshita, T., and Tatematsu, M. (2006). Gastric-and-intestinal mixed-type intestinal metaplasia: aberrant expression of transcription factors and stem cell intestinalization. Gastric Cancer *9*, 156-166.

31. Almeida, R., Silva, E., Santos-Silva, F., Silberg, D.G., Wang, J., De Bolos, C., and David, L. (2003). Expression of intestine-specific transcription factors, CDX1 and CDX2, in intestinal metaplasia and gastric carcinomas. J Pathol *199*, 36-40.

32. 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.

33. Li, X.L., Eishi, Y., Bai, Y.Q., Sakai, H., Akiyama, Y., Tani, M., Takizawa, T., Koike, M., and Yuasa, Y. (2004). Expression of the SRY-related HMG box protein SOX2 in human gastric carcinoma. Int J Oncol *24*, 257-263.

34. Tsukamoto, T., Inada, K., Tanaka, H., Mizoshita, T., Mihara, M., Ushijima, T., Yamamura, Y., Nakamura, S., and Tatematsu, M. (2004). Down-regulation of a gastric transcription factor, Sox2, and ectopic expression of intestinal homeobox genes, Cdx1 and Cdx2: inverse correlation during progression from gastric/intestinal-mixed to complete intestinal metaplasia. J Cancer Res Clin Oncol *130*, 135-145.

35. Nakamura, H., Sato, T., and Suzuki-Hirano, A. (2008). Isthmus organizer for mesencephalon and metencephalon. Dev Growth Differ *50 Suppl 1*, S113-118.

36. Joyner, A.L. (1996). Engrailed, Wnt and Pax genes regulate midbrain--hindbrain development. Trends Genet *12*, 15-20.

37. Meinhardt, H. (2006). Primary body axes of vertebrates: generation of a near-Cartesian coordinate system and the role of Spemann-type organizer. Dev Dyn *235*, 2907-2919.

38. Crossley, P.H., Martinez, S., and Martin, G.R. (1996). Midbrain development induced by FGF8 in the chick embryo. Nature *380*, 66-68.

39. Matsuo, I., Kuratani, S., Kimura, C., Takeda, N., and Aizawa, S. (1995). Mouse Otx2 functions in the formation and patterning of rostral head. Genes Dev *9*, 2646-2658.

40. Wassarman, K.M., Lewandoski, M., Campbell, K., Joyner, A.L., Rubenstein, J.L., Martinez, S., and Martin, G.R. (1997). Specification of the anterior hindbrain and establishment of a normal mid/hindbrain organizer is dependent on Gbx2 gene function. Development *124*, 2923-2934.

41. Meyers, E.N., Lewandoski, M., and Martin, G.R. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. Nat Genet *18*, 136-141.

42. Chi, C.L., Martinez, S., Wurst, W., and Martin, G.R. (2003). The isthmic organizer signal FGF8 is required for cell survival in the prospective midbrain and cerebellum. Development *130*, 2633-2644.

43. Blair, S.S. (2003). Lineage compartments in Drosophila. Curr Biol *13*, R548-551.

44. Buchberger, A., Pabst, O., Brand, T., Seidl, K., and Arnold, H.H. (1996). Chick NKx-2.3 represents a novel family member of vertebrate homologues to the Drosophila homeobox gene tinman: differential expression of cNKx-2.3 and cNKx-2.5 during heart and gut development. Mech Dev *56*, 151-163.

45. Smith, D.M., Nielsen, C., Tabin, C.J., and Roberts, D.J. (2000). Roles of BMP signaling and Nkx2.5 in patterning at the chick midgut-foregut boundary. Development *127*, 3671-3681.

46. Self, M., Geng, X., and Oliver, G. (2009). Six2 activity is required for the formation of the mammalian pyloric sphincter. Dev Biol *334*, 409-417.

47. Burn, S.F., Boot, M.J., de Angelis, C., Doohan, R., Arques, C.G., Torres, M., and Hill, R.E. (2008). The dynamics of spleen morphogenesis. Dev Biol *318*, 303-311.

48. Smith, D.M., and Tabin, C.J. (1999). BMP signalling specifies the pyloric sphincter. Nature *402*, 748-749.

49. Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes Dev *9*, 1654-1666.

50. Theodosiou, N.A., and Tabin, C.J. (2005). Sox9 and Nkx2.5 determine the pyloric sphincter epithelium under the control of BMP signaling. Dev Biol *279*, 481-490.

51. Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., and de Crombrugghe, B. (2001). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci U S A *98*, 6698-6703.

52. Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E., and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell *124*, 47-59.

53. Vokes, S.A., Ji, H., McCuine, S., Tenzen, T., Giles, S., Zhong, S., Longabaugh, W.J., Davidson, E.H., Wong, W.H., and McMahon, A.P. (2007). Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development *134*, 1977-1989.

54. Borok, M.J., Tran, D.A., Ho, M.C., and Drewell, R.A. Dissecting the regulatory switches of development: lessons from enhancer evolution in Drosophila. Development *137*, 5-13.

55. Vokes, S.A., Ji, H., Wong, W.H., and McMahon, A.P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes Dev *22*, 2651-2663.

56. Lee, E.Y., Ji, H., Ouyang, Z., Zhou, B., Ma, W., Vokes, S.A., McMahon, A.P., Wong, W.H., and Scott, M.P. Hedgehog pathway-regulated gene networks in cerebellum development and tumorigenesis. Proc Natl Acad Sci U S A *107*, 9736-9741.

57. Hallikas, O., and Taipale, J. (2006). High-throughput assay for determining specificity and affinity of protein-DNA binding interactions. Nat Protoc *1*, 215-222.

58. Stormo, G.D., and Fields, D.S. (1998). Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci *23*, 109-113.

59. Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.

60. Powell, J.R. (1997). Progress and prospects in evolutionary biology : the Drosophila model, (New York: Oxford University Press).

61. Moses, A.M., Pollard, D.A., Nix, D.A., Iyer, V.N., Li, X.Y., Biggin, M.D., and Eisen, M.B. (2006). Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput Biol *2*, e130.

62. Moses, A.M., Chiang, D.Y., Kellis, M., Lander, E.S., and Eisen, M.B. (2003). Position specific variation in the rate of evolution in transcription factor binding sites. BMC Evol Biol *3*, 19.

63. Alexandre, C., Jacinto, A., and Ingham, P.W. (1996). Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. Genes Dev *10*, 2003-2013.

64. Kwon, C., Hays, R., Fetting, J., and Orenic, T.V. (2004). Opposing inputs by Hedgehog and Brinker define a stripe of hairy expression in the Drosophila leg imaginal disc. Development *131*, 2681-2692.

65. Muller, B., and Basler, K. (2000). The repressor and activator forms of Cubitus interruptus control Hedgehog target genes through common generic gli-binding sites. Development *127*, 2999-3007.

66. Von Ohlen, T., and Hooper, J.E. (1997). Hedgehog signaling regulates transcription through Gli/Ci binding sites in the wingless enhancer. Mech Dev *68*, 149-156.

67. Kumar, S., and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. Nature *392*, 917-920.

68. Horvath, J.E., and Willard, H.F. (2007). Primate comparative genomics: lemur biology and evolution. Trends Genet *23*, 173-182.