

Refining sea urchin developmental gene regulatory
network models by incorporating Wnt signaling and
information processed at the *hox11/13b* locus

Thesis by
Miao Cui

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy in Biology

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
(Defended May 6th, 2016)

© 2016

Miao Cui
ORCID: [0000-0003-3150-3930]

ACKNOWLEDGMENTS

First and most importantly, I would like to thank my thesis advisor, Dr. Eric Davidson, for giving me the opportunity to work as a PhD student in his laboratory over the past 6 years. His passion for science and his spirit had always inspired my quest for never-ending improvement and dedication towards making great strides for the scientific community. I am also grateful to him for providing guidance and unreserved support during my thesis research. None of this would be possible without him, and I hope he would be proud if he was still with us to witness this final section of my PhD thesis.

I also want to thank my academic foster parent, Dr. Marianne Bronner, for being such a generous person and for everything she did to help me to get through the most difficult few months of my PhD. Dr. Bronner has shown me how to be successful in an academic career, not only by doing good research but also equally important by being kind and generous to others.

Working in Davidson's lab has always been stimulating. I appreciate the collaboration I had with Dr. Isabelle Peter and Dr. Enhu Li. Dr. Peter has been a long-standing collaborator as well as a co-mentor to me in the past four years. I appreciate the time and effort she devoted to the projects as well as all the helpful discussions we had. I cannot be more grateful to Dr. Li's mentorship, friendship, and his patience in teaching me.

I also want to thank the past and the present colleagues from the Davidson lab: Andy Ransick for being my first mentor and teaching me how to do injection; Jongmin Nam, Julius Barsi, Smadar Ben-Tabou-De Leon, and Qiang Tu for teaching me how to do cis-reg analysis; my fellow graduate students Stefan Materna, Eric Erkenbrack, and Jonathon Valencia for their friendship and supports; Andy Cameron and Parul Kudtarkar for keeping the Spbase (now EchinoBase) running; Erika Vielmas, Julie Hahn, Ping Dong, Miki Yun for their technical support and friendships. Special thanks for Jane Rigg and Deanna Thomas for keeping the lab together and making sure everything in the lab runs smoothly. I want to acknowledge Brain Peng He, Say-Tar Goh, and Brain William from Dr. Barbara Wold's lab for their assistance, valuable advices, and friendship.

I would also like to thank the members of my thesis committee, Dr. Ellen Rothenberg, Dr. Paul Sternberg, and Dr. Angelike Stathopoulos, for their time, guidance, support, and incredible valuable advices on career development.

Finally I would like to thank my parents for their unconditional love and support and for believing me and understanding me, and my husband Xiaojian (JJ) Deng for his support and for the effort he made in proof-reading my manuscript and thesis. I am forever indebted to them because of that.

ABSTRACT

The availability of advanced GRN models for sea urchin development presents a unique opportunity to address the function of signaling interactions in cell fate specification at a system-wide level. Here we take a global approach to investigate the regulatory functions of the Wnt signaling system during pre-gastrular development. We examine the embryonic specification processes in order to determine in which embryonic lineages and at what time specific Wnt signals are required. We show a functional divergence among individual Wnt ligands despite their similar and partially overlapped spatial expression. By studying TF activators, we show that expression of *wnt* genes is tightly controlled and is correlated with their respective functions. In particular Wnt1 and Wnt16, which regulate endodermal specification, are activated by the endoderm regulator Hox11/13b. Motivated by these results and in an effort to further enhance our understanding of endodermal specification, we conducted a cis-regulatory analysis of the *hox11/13b* gene across a range of developmental stages up to 60 hours post-fertilization. We identify Ets, Eve, and Tcf as direct regulators of *hox11/13b*, and we show that their combinatorial control drives the endoderm-specific and dynamic early expression of *hox11/13b*. Furthermore, we show that its late expression in the hindgut is controlled by an inter-modular AND logic gate, in which two separate regulatory modules are both required but neither alone is sufficient.

PUBLISHED CONTENT AND CONTRIBUTIONS

Cui, M., Siriwon, N., Li, E., Davidson, E.H., Peter, I.S. (2014). "Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo." Proc Natl Acad Sci U S A **111**(47): E5029-5038. doi: 10.1073/pnas.1419141111

Miao Cui participated in the designing of the project, performed research, analyzed data, and participated in the writing of the manuscript.

Li, E., Cui, M., Peter, I.S., and Davidson, E.H. (2014). "Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo." Proc Natl Acad Sci U S A **111**(10): E906-913. doi: 10.1073/pnas.1323105111

Miao Cui participated in the conception of the project, performed research, analyzed data, and participated in the writing of the manuscript.

TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract	v
Published Content and Contributions.....	vi
Table of Contents.....	vii
List of Illustrations and/or Tables.....	x

Chapter 1: Introduction

Overview	1
Sea urchin embryogenesis: specification of cell fates.....	4
Understanding the architecture of gene regulatory networks	9
Understanding differential gene expression <i>via</i> cis-regulatory analysis	22
The endoderm gene regulatory networks.....	31
Theme of thesis.....	40

Chapter 2:

Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo

Abstract	54
Introduction.....	55
Results	
Spatial and temporal expression of <i>wnt</i> , <i>fzd</i> , <i>sfrp</i> , and <i>dkk</i> genes.....	59

System-wide perturbation of Wnt signaling by a Porcupine inhibitor.....	65
Wnt ligands execute distinct functions	73
Wnt-dependent spatial patterning function	79
Control of <i>wnt</i> gene expression by cell fate specification GRNs	81
Discussion	85
Methods.....	93
Supplemental figures and table	96

Chapter 3:

Endoderm-specific HOX gene expression mediated by temporally and spatially distinct interacting cis-regulatory modules

Abstract	123
Introduction.....	124
Results	
Expression dynamics of <i>hox11/13b</i> and identification of active regulatory modules.....	129
Early regulatory module <i>E</i> contains two functionally distinct sub-modules.....	135
Dual functionality of TCF is responsible for the specific spatial pattern of <i>hox11/13b</i> expression	139
ETS and EVE work sequentially to activate early <i>hox11/13b</i> expression	142

A distal regulatory module <i>D</i> contributes to robust <i>hox11/13b</i> activation.....	147
Late expression in the hindgut requires two separate cis-regulatory modules	149
Identification of binding sites required for late <i>hox11/13b</i> expression in the hindgut	154
Discussion	157
Methods.....	164
Supplemental figures and table	171
Chapter 4: Conclusion	191
Wnt signaling is not required for the specification of mesoderm lineages and the early expression of endoderm genes due to the presence of maternal beta-catenin	192
The earliest function of Wnt signaling is to restrict the neurogenic apical plate	193
Functional specificity of Wnt signaling	194
The regulatory logic for <i>hox11/13b</i> expression	196
The establishment of the endoderm/ectoderm boundary is controlled at the <i>hox11/13b</i> locus	198
Context-dependent multi-functionality of binding sites in the <i>hox11/13b</i> regulatory modules	199

LIST OF ILLUSTRATIONS AND/OR TABLES

	<i>Page</i>
Chapter 1	
Figure 1.....	9
Figure 2.....	24
Figure 3.....	32
Chapter 2	
Figure 1.....	64
Figure 2.....	72
Figure 3.....	78
Figure 4.....	85
Supplement Figures and tables	96-114
Chapter 3	
Figure 1.....	133
Figure 2.....	134
Figure 3.....	138
Figure 4.....	141
Figure 5.....	154
Figure 6.....	157
Figure 7.....	158
Supplement Figures and tables	171-181

*Chapter 1***INTRODUCTION****Overview**

Development is a unidirectional process that transforms a fertilized egg from a single cell into a multicellular organism composed of multiple tissues with distinct functions. Two of the most interesting and defining features of development are its ability to continuously increase organismal complexity and the fact that it is highly reproducible among individuals and across generations (Peter and Davidson 2015). This makes it clear that the biological program governing development must be extremely robust and well-conserved over time. It is now understood that this program is encoded within the genome in two different components (Davidson 2006). First, the coding sequences of regulatory genes are transcribed and translated in order to produce regulatory proteins that mediate gene expression by binding to target DNA sites in a sequence-specific manner. Second, regulatory regions that are often located in noncoding stretches of DNA can control the expression of genes by integrating multiple regulatory inputs in a single given cell. Therefore, genomic DNA directs development by determining which genes will be expressed and how they will drive development forward within each cell and at every point in time.

Every cell in an organism contains an identical genome, with a few specialized

exceptions, and yet different cells can express different sets of genes and adopt distinct cellular fates and functions. The relationship between the genomic DNA sequence and gene expression activity is therefore not linear (Peter and Davidson 2015). Gene regulatory networks (GRNs) control how a collection of expressed regulatory genes interact at the DNA level to determine the function and identity of each spatial domain in the organism at every stage of development. GRNs explain how cells can function differently despite having the same genomic information, and why the complexity of established regulatory states increases as development proceeds.

Specification, by which cells attain different regulatory states, is the basic process responsible for organizing development. It is continuous, progressive, and irreversible, and it is a direct functional output of developmental GRNs. Specification drives development by partitioning embryos into distinct spatially-organized domains that pre-program the fate and function of the cells which descend from them. To reach a given specification state, cells go through various regulatory states that lead from one to the next. The process often involves cell-autonomous reconstruction of GRNs, which typically necessitates changing patterns of gene expression, and also often requires inductive signaling interactions between regulatory domains. Embryonic specification determines the structure and complexity of animal body plans and is precisely regulated throughout developmental time and space. Understanding how specification is regulated by studying the underlying GRN structure can offer a causal explanation of how

development is controlled.

As a model system, the sea urchin offers many advantages for the study of the molecular mechanisms underlying developmental events. These advantages include a relatively slow growth rate which allows for the detailed capture of developmental processes in high resolution, transparent embryos that are readily accessible in large numbers year round, fewer copies of homologous genes due to the lack of whole genome duplication, which eases functional genetic studies, and straightforward and efficient gene transfer systems that do not rely on transgenic animal lines. Over the past 20 years, this system has been further advanced by novel molecular tools, genome sequencing, transcriptomics data, and improved morphological understanding of cell fate specification. At present, maps of early sea urchin development are highly comprehensive and comprise the specification of nearly all the embryonic domains up to the pre-gastrular stage. The best known and experimentally determined, large-scale embryonic GRN models are those of the specification of endoderm and mesoderm in the embryos of the sea urchin *Strongylocentrotus purpuratus* up to gastrulation (Davidson 2006, Peter and Davidson 2010, Peter and Davidson 2011, Oliveri et al, 2002, Materna et al, 2013). This GRN encompasses about half of the embryo, covering precursors of two mesodermal lineages (skeletogenic mesoderm and non-skeletogenic mesoderm) and two endodermal lineages (anterior endoderm, also known as veg2 endoderm, and posterior endoderm, which is also known as the veg1 endoderm). Major progress has

recently been made on the GRN linkages within other major territories including oral and aboral ectoderm, the neurogenic ciliated band, and the later oral and aboral mesoderm (Su, Li et al. 2009, Li, Materna et al. 2012, Li, Materna et al. 2013, Materna, Ransick et al. 2013, Barsi, Li et al. 2015). Networks for neurogenic apical plate specification and for the later development of gut are being rapidly solved (R. Feuda and I. Peter, unpublished data)

Since there has been such a success in decoding the genomic program for sea urchin embryonic development, the purpose of this introduction is to summarize three essential aspects that necessitate our understanding about these developmental processes: i) Where and when embryonic lineage specification occurs; ii) How the underlying GRNs are structured (network construction); iii) How differential gene expression is being regulated (cis-regulatory analysis). The last part of the introduction focuses specifically on the endoderm, with particular attention paid to network architecture that specifies both the anterior and posterior endoderm.

SEA URCHIN EMBRYOGENESIS: SPECIFICATION OF CELL FATES

Embryogenesis in sea urchin *S. purpuratus* embryos has been reviewed extensively (Davidson, Cameron et al. 1998). It exhibits radial holoblastic cleavage. The first two cleavages pass through the animal and vegetal poles and produce 4 macromeres, both of which are essentially developmentally equivalent and totipotent at this stage. The third cleavage is perpendicular to the first two cleavage

planes and generates 8 cells of equal size but of different developmental potential. The 4 cells in the animal hemisphere of the embryo will ultimately specify to ectodermal tissues and the 4 cells in the vegetal hemisphere will form endodermal and mesodermal tissues. The fourth cleavage divides the cells of the animal tier into eight mesomeres, each with the same volume; the vegetal tier, however, undergoes an unequal cleavage in order to produce four large cells, the macromeres, and four smaller micromeres at the vegetal pole. The fifth cleavage is also unequal and is additionally desynchronized. The animal mesomeres and the vegetal macromeres first undergo an equal cleavage followed by an asymmetrical division of micromeres, which produces 4 small micromeres that remain at the vegetal pole and 4 large micromeres that lie directly above. By the seventh cleavage, cells have undergone additional rounds of division both laterally and equatorially. The embryos now contain 128 cells with a number of distinct cell types having already been specified. The cells in the animal hemisphere are specified into either neurogenic apical plate or oral/aboral/lateral ectoderm. The vegetal hemisphere macromeres are further divided into veg1 and veg2 cells. The veg1 layer cells will later specify into vegetal ectoderm and posterior endoderm, whereas the veg2 layer cells will give rise to the anterior endoderm and the non-skeletogenic mesoderm (NSM). The large micromeres are specified to the skeletogenic mesenchyme (SM), and the small micromeres will give rise to embryonic germ cells. A diagram summarizing the lineage specification processes of sea urchin embryos including these early stages of cleavages is shown in Figure 1.

The cellular fates of most lineages are achieved through conditional specification. The only cells whose fates are determined autonomously are the SMs, meaning that these micromeres will still form skeletal spicules if they are isolated from the embryo and cultured in an *in vitro* condition (Okazaki 1975). The specification of these SMs occurs as early as the 4th cleavage and is triggered by the vegetal localization of the maternal nuclear β -catenin, a nuclear effector of the canonical Wnt pathway, in the micromeres. This asymmetric distribution of nuclear β -catenin is mediated by the maternal protein Disheveled that protects cytoplasmic β -catenin from degradation in the vegetal cortex of the oocyte (Leonard and Ettensohn 2007, Kumburegama and Wikramanayake 2008). Cytoplasmic β -catenin then translocates to the nucleus and forms a complex with T-cell factor (Tcf), converting it from a repressor to an activator. At the same time, maternal Otx also enters the nucleus in the vegetal pole (Kenny, Kozlowski et al. 1999). The combination of β -catenin nuclear localization and maternal Otx directly activate the expression of *pmar1* in the large micromeres, which in turn initiates the specification of the SM lineage (Oliveri, Carrick et al. 2002, Oliveri, Tu et al. 2008). Between the 7th and 9th cleavage, the SMs located in the central vegetal plate of the embryo express Delta ligands that can interact with the Notch receptors expressed in the adjacent veg2 cells. The Delta-Notch receptor signaling converts the inner layer of veg2 cells to NSMs, which later give rise to pigment cells, immunocytes, and muscle cells. The outer layer of veg2 cells further from the vegetal pole are not exposed to this Delta-Notch signaling and instead initiate the expression of some early endodermal genes including *wnt* ligands, due to the activating function of the maternal TCF/ β -catenin

complex (Peter and Davidson 2010). The expression of *wnt* ligands in the veg2 endodermal cells then induces the accumulation of zygotic nuclear β -catenin in adjacent veg1 cells, which are consequently specified into the posterior endoderm and the vegetal perianal ectoderm lineage (Cui et al. 2015) There is thus a cascade wherein the asymmetric distribution of maternal nuclear β -catenin in the vegetal pole micromeres induces large micromeres to become SMs. These SMs in turn induce the specification and the fate determination between NSM and veg2 endoderm within the veg2 cells. Finally, veg2 induces the cells spatially located above them to assume a veg1 endodermal fate. In the animal half of the embryo, cells of the ectodermal lineages are depleted of nuclear β -catenin. In fact, ectopic expression of nuclear β -catenin in the animal hemisphere by lithium treatment prevents the development of the ectodermal lineage, suggesting an antagonistic function of nuclear β -catenin in specifying the animal ectodermal fates. Cells located in the animal pole form the neurogenic apical plate which gives rise to larval neurons, and the cells residing between the apical plate and the veg1 macromeres specify into either oral (ventral) or aboral (dorsal) ectoderm, which will compose the majority of the larval body skin (Li, Cui et al. 2014). The partition of the apical plate and the oral/aboral ectoderm along the animal-vegetal axis is mediated by Wnt8 signaling, possibly through non-canonical pathways or via a cascade of repressive mechanisms (Range, Angerer et al. 2013). Therefore, lineage specification of sea urchin embryos along the animal-vegetal axis is largely dependent on the asymmetric distribution of the maternal nuclear β -catenin and the vegetal-specific activation of the Wnt signaling.

Non-apical ectoderm and NSM are further divided into oral and aboral compartments (Li, Materna et al. 2012, Materna, Ransick et al. 2013). It has been shown that Nodal signaling is necessary for the fate specification along the oral-aboral axis for both ectoderm and NSM through activation of a homeodomain repressor *not* (Li, Materna et al. 2012, Materna, Ransick et al. 2013). The expression of *not* in the oral NSM leads to the repression of the aboral regulatory state in the oral NSM and the expression of a specific set of regulatory genes that constitute the oral NSM GRN. Similarly, the expression of *not* in the oral ectoderm represses aboral ectoderm genes and contributes to the bilateral spatial organization of the embryonic oral ectoderm. Nodal signaling also indirectly activates aboral genes by inducing the expression of *bmp2/4*, which is then transduced to the aboral ectoderm to activate the aboral ectoderm specification program. The asymmetric activation of *nodal* in the oral face is initially mediated by Bzip transcription factors, the activity of which is redox sensitive and later maintained by its antagonist Lefty in the aboral side (Nam, Su et al. 2007, Range and Lepage 2011). Its later expression in the vegetal lineages requires a Wnt signaling-mediated de-repression mechanism. Thus, lineages patterning along the animal-vegetal axis and oral-aboral axis are interconnected and temporally coordinated. As summarized above, cell fate specification in sea urchin embryos is a coherent process initiated by asymmetrically distributed maternal inputs and signaling transduction.

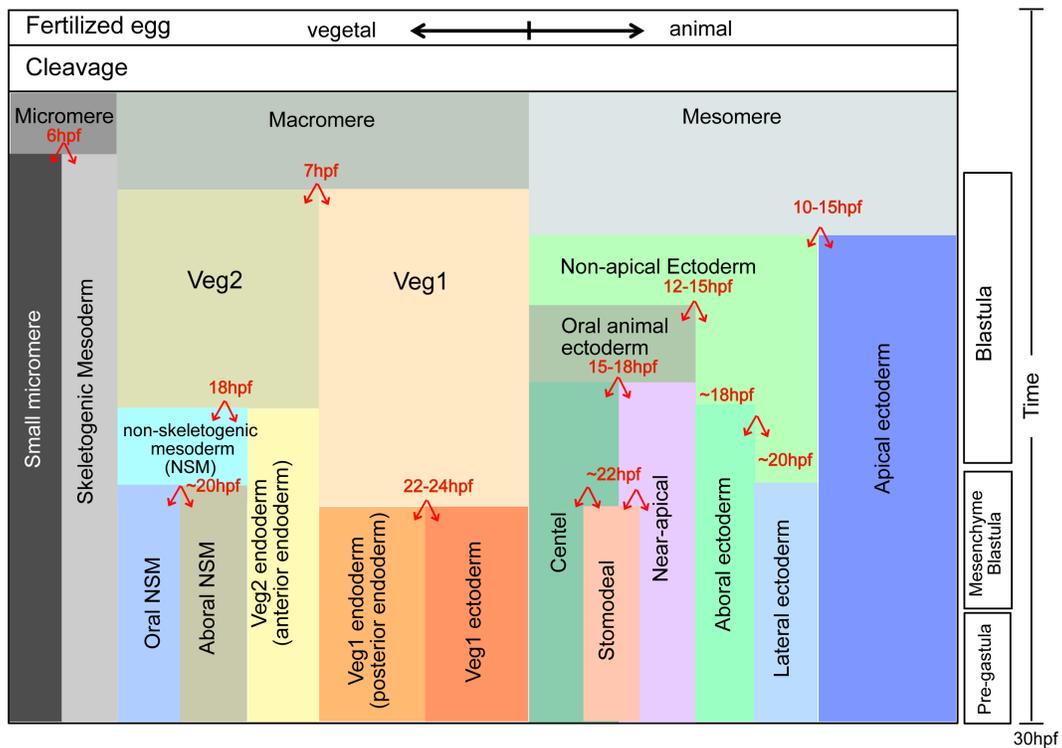


Figure 1. Process diagram of lineage specification in sea urchin embryos. The specification processes of all embryonic lineages during pre-gastular development of sea urchin embryos are shown in respect to time and the radial position along the animal-vegetal axis. Red arrows indicate separation of cellular fates; the timing of events is specified. Figure adapted from Li, Cui et al. 2014 and Peter and Davidson 2010.

UNDERSTANDING THE ARCHITECTURE OF GENE REGULATORY NETWORKS

GRNs are composed of regulatory genes such as transcription factors (TF) and signaling mediators, as well as the interactions between them (Davidson 2006). They consist of multiple modular entities, each of which is composed of the regulatory interactions necessary for a discrete developmental task, e.g. response to

signaling induction, exclusion of another cell fate, or the activation of differentiation effector genes. Experimentally establishing the structure of a GRN involves the building of a logic map that reveals cause-effect linkage of the regulatory factors involved (Davidson, Rast et al. 2002). Understanding how GRNs are structured offers mechanistic insights into developmental events such as specification and differentiation. The approach to constructing GRN models has been successfully applied to sea urchin embryonic development, and had previously been described thoroughly (Materna and Oliveri 2008). Here I summarize the key steps of GRN models construction, including insights gained from recent studies and from my own thesis research.

Before starting to map GRNs, the embryological information of the cell lineages await for GRN models construction need to be understood. This includes when during development and from where in the embryo these lineages emerge. Other important considerations include the types of adjacent groups of cells, and an understanding of changing cell-cell contacts as a result of cellular movement. Answering these questions is essential in order to lay the groundwork that GRNs models will be built upon. The position of cells in respect to the overall orientation of the embryos and the neighboring cell lineages is also important, as this may determine the initial inputs that start a specification program. Complexity within embryos increases as development proceeds, and a single embryonic domain in a given developmental stage can be further divided into multiple cell lineages during subsequent stages. For practical reasons, it is common to constrain network model

construction to a specific time and space. One example is the GRN governing the specification of sea urchin endomesoderm (Peter and Davidson 2010, Peter and Davidson 2011). The network study was subdivided to first focus on the period from 9-18h post-fertilization, before the separation of the endoderm and the mesoderm, as well as the period 18-30h post-fertilization, when both lineages have undergone divergent regulatory programs.

Another piece of information required before initiating GRN construction is the characterization of regulatory states, although this is sometimes acquired in parallel with the identification of network components as discussed below. Regulatory domains transiently represent groups of identical cells that transcribe the same sets of transcription factors. Characterization of these regulatory states during a particular set of developmental events is important for network construction for two reasons: 1) it reveals how the establishment of a cell lineage or subdivision of embryonic domains occurs progressively with both spatial and temporal resolution; 2) this progressive establishment of regulatory states is the output of GRNs and can be later used to validate the structures and logic of constructed networks.

Identification of candidate network components

Ever since the sea urchin genome was sequenced in 2006, it has provided invaluable information into sea urchin biology. The genome encodes about 23000 genes and is roughly 800 megabases in size (Sodergren, Weinstock et al. 2006). It

contains nearly all bilaterian transcription factor families and a majority of known signal transduction genes. All the genes that are active during regulatory processes of interest can be identified by whole transcriptome analysis (RNA-seq) (Tu, Cameron et al. 2012, Tu, Cameron et al. 2014). A complete list of network components can be obtained in this way, and their relative level of expression can also be quantified at the same time. The transcriptome of 10 different sea urchin embryonic developmental stages is already available. Recently, these analyses have been extended to individual cell populations, including presumptive pigment cells, presumptive neurogenic cells, presumptive skeletogenic cells, cells from the stomodeal region of the oral ectoderm, ciliated band cells, and veg1 cells (Barsi, Tu et al. 2014, Barsi, Tu et al. 2015). These data have greatly narrowed the search range for regulatory components and expedited the pace at which GRN models can be assembled.

Detailed temporal expression profiles for a majority of transcription factors have been established from 0 to 48h post-fertilization at 1-hour intervals (Materna, Nam et al. 2010). These profiles offer a wealth of information regarding the timing of transcription factor activation as well as the kinetics of their consequent gene regulation. Accurate characterization of the spatial expression of active transcription factors is essential for the identification of network components. The spatial gene expression patterns of many key transcription factor families have been measured, including zinc finger factors, Ets family factors, Homeodomain proteins, and nuclear mediators of some signaling pathways (Howard-Ashby, Materna et al.

2006, Materna, Howard-Ashby et al. 2006, Rizzo, Fernandez-Serra et al. 2006, Tu, Brown et al. 2006). These analyses make it possible to estimate the composition of regulatory inputs expressed in nearly all cell types and developmental stages in the early sea urchin embryo.

Regulatory genes that are specifically expressed in the embryonic domain of interest at during a time of interest are obvious GRN candidates. These genes do not, however, necessarily constitute a complete list of GRN components. Genes that are expressed more broadly and genes specifically excluded from the region of interest should also be considered. Although genes expressed broadly or even ubiquitously may not contribute to the spatial regulation of a particular developmental process, they may still be important components for the execution of GRN programs. For example, many genes in the endomesoderm GRN are activated by ubiquitously expressed activators while being spatially restricted by repressive mechanisms in other domains.

Another approach that has been successfully used to identify candidate genes involved in the specification of vegetal or animal lineages is the usage of chemical treatments that can transform embryos into either animalized or vegetalized compartments by interfering with specific signaling pathways. It has long been known that zinc treatment causes expansion of animal ectodermal domains while repressing the vegetal endomesodermal lineage and the opposite effect is observed upon lithium treatment (Poustka, Kuhn et al. 2007). By measuring changes in the

expression of regulatory genes in embryos treated with either chemical, genes exhibiting significant up- or down-regulation can be identified as candidate components for networks governing the specification of animal or vegetal domains. This type of approach to identify candidate genes can potentially be applied to any lineage known to undergo conditional/inductive specification processes. Candidate genes identified using this approach are discovered as a functional output of morphological changes, in the sense similar to perturbation analyses except for being less specific, and is not based on temporal or spatial associations.

Examination of cause-effect linkages through perturbation analyses

Perturbation of the transcription or function of regulatory genes followed by quantitatively and spatially monitoring the consequences of this disruption is an essential step in building experimentally-derived GRN models. It is the only way to convert temporal and spatial information into a functional understanding that is composed of cause-effect linkages between regulatory genes. Perturbation analyses need to be carried in a systematic order wherein the sequence of genes being perturbed is dependent upon their temporal profiles. It is generally helpful to group the genes by the timeframe when they are first expressed in a particular domain of interest. Genes that are expressed first are often located at the top of GRN hierarchy, and they are more likely to regulate genes expressed during later periods of time. Therefore, it makes sense to first perturb early-expressed genes. This principle is typically applied during network construction for sea urchin

embryogenesis.

Perturbation experiments can be designed to alter either gene transcription or protein function. In either case, a successful perturbation experiment needs to be demonstrably effective and specific to the target gene of interest. Perturbation approaches are often divided into “loss-of-function” and “gain-of-function” analyses. “Loss-of-function” analyses are the most efficient and commonly used method for discovering target genes. Common experimental techniques for these analyses include gene knockdown, which can be directed through morpholino anti-sense nucleotides (commonly used in the sea urchin system), shRNA, or genetic knock-out (only feasible in systems with genetic tools available such as mice or *Drosophila*). The efficacy of gene knockdown experiments can be assessed via immunostaining or western immunoblot.

Other approaches designed to interfere with the function of a protein of interest can also be effective means of mapping regulatory wiring. These approaches may rely on RNA cloning of synthetic proteins, such as proteins with truncated sequences or which have been engineered to contain heterologous active or repressive domains. When introduced into embryos, these synthetic proteins compete with the endogenous proteins to either cause a functional loss of the endogenous protein or to forcibly alter its role in the activation or repression of downstream genes. In some cases these approaches are preferable to gene knock-down experiments, despite relying on mRNA overexpression. For example, sea urchin *otx* is a highly

maternally transcribed gene, whose function cannot be easily disrupted through morpholino perturbation. By creating a recombinant Otx protein containing its DNA-binding domain fused with the repressor domain of *Drosophila* Engrailed, this synthetic protein will now bind to its usual target genes but it will now repress their transcription instead. Other successful examples include the use of truncated forms of signaling receptors that lack intracellular domains to outcompete functional endogenous receptors for available ligand, thus disrupting signaling transduction. It is worth noting that introducing synthetic mRNA into the embryos does cause the expression of genes in cells that do not normally express their endogenous counterparts. This may cause off-target experimental artifacts due to non-specific binding or effects occurring in unrelated domains. Therefore, the amount of mRNA injected needs to be carefully controlled and tested.

“Gain-of-function” analyses are used to examine the effect of increased expression and/or of ectopic expression of a particular gene. This is typically carried out via the overexpression of *in vitro* transcribed mRNA into the embryos. While this approach can be used to test the effect of increased candidate gene expression, it is more often used to confirm regulatory wiring that has been already discovered. With some exceptions, it is generally expected that if an effect is observed with gene knockdown, the opposite effect should also be observed when this gene is overexpressed. When the goal of a gain-of-function experiment is to test the effects of ectopic gene expression, it is designed to test the sufficiency of the examined gene to activate or represses its target genes in cells that do not normally express

this gene. In this type of experiment, the spatial effects of target genes need to be carefully monitored.

Gain-of-function analyses are particularly useful when they are combined with gene knockdown in a single experiment. These combinatorial experiments allow a researcher to test the specificity of the perturbation by examining whether overexpression of mRNA is able to rescue the effect of gene knockdown, and to simultaneously discover intermediate regulators that can clarify network structure. For the first purpose, mRNA of the same gene is introduced into the embryos – note that it is important to be sure that this mRNA is “immune” to the gene knockdown approach being used. For the second purpose, mRNA of different genes is overexpressed. For example, functional disruption of gene A may decrease the expression of both gene B and gene C; however, this effect may be rescued by overexpression of B but not of C. These results collectively suggest a linear regulatory pathway, in which A activates B and B then activates C. Precautions similar to those discussed above must be taken when introducing synthetic RNA in order to minimize the production of experimental artifacts. Several concentrations of mRNA should be utilized, and the morphology of the injected embryos should be carefully monitored before analyzing any molecular effects.

In the sea urchin model, perturbative agents are most commonly delivered to fertilized eggs, and are thus ubiquitously present in all domains and at all stages of development so long as they remain effective. This can be problematic for genes

that are expressed in multiple domains and/or exhibit multiphasic expression. In other system like mouse and *Drosophila*, this problem can be resolved via conditional knockdown (e.g. tamoxifen-inducible gene knockout), but this still remains a prominent issue for sea urchins. In the case of signal transduction, time-specific perturbation can be achieved by treating embryos with small-molecule inhibitors that can penetrate embryos during any desired developmental window. A potentially useful method that may help to circumvent this problem is the creation of tissue-specific reporter constructs such as BAC recombinants to drive the expression of dominant-negative form of genes in a timely and spatially controlled manner. In sea urchins these expression constructs clonally integrate into the genome resulting in mosaic expression patterns, but it is nevertheless a useful technique, particularly when using fluorescence-based cell sorting to isolate a population of reporter positive cells in order to assess a regionalized effect of the loss of gene function.

For network construction purposes, any perturbation analysis needs to be subject to an accurate evaluation of its effects on gene expression in order to establish regulatory linkages between genes. The effect of perturbation can be monitored quantitatively using technique that allows for RNA quantification, such as qPCR, Nanostring nCounter, RNA microarrays, or the spatial use of whole mount *in situ* hybridization. Quantitative measurement is in some cases favored over spatial regulation, but the opposite is often true whereby spatial measurement plays a more dominant role. Quantitative measurement is typically faster and easier to scale-up,

and it is therefore often used for the initial screening for target genes. *In situ* hybridization, on the other hand, takes several days to complete and has a limited ability to undergo multiplexing; nonetheless it is an invaluable and highly sensitive method of monitoring spatial changes of gene expression.

Negative control experiments are extremely important for any perturbation analysis. They provide a reference to which gene expression levels in the perturbed embryos can be normalized. The choice of negative control depends on the perturbation methods being used. For morpholino injection, a random mixture is often a good control. For mRNA overexpression, mRNA of non-functional exogenous gene, such as GFP, can be used. For chemical inhibition, the solvent solution, e.g. DMSO and ethanol, is suitable as a negative control. In all cases, wild-type embryos should be always included as a negative control to rule out any potential intrinsic issues within the embryos.

Network models assembly

Network models assembly is the process of extracting and analyzing the perturbation data and presenting them in the form of regulatory interactions between network components. This process can be repetitive and is commonly combined with perturbation analyses. Careful interpretation of perturbation experiments is essential to minimize the numbers of false links. Biological repeats for perturbation analyses are required, especially for the systems that have non-

inbred genetic backgrounds such as sea urchins.

Although the goal of network assembly is to include all possible regulatory interactions, it is more practical to start by focusing on smaller portions of the network. Such subnetwork units are sometimes referred to as modules, wherein genes are connected functionally. A provisional module can be identified directly through perturbation analyses by identifying a set of genes that share a common regulatory factor. Revisions and combinations of modules can then be made through the use of further perturbation analyses.

Some computational tools are available to aid in the analysis of the perturbation data and the visualization of results, with the most common tool used for sea urchin GRNs being BioTapestry. It is a tool that can represent the structure and dynamic properties of GRNs while also being able to interpret perturbation data and suggest alternative network architecture that may be tested by additional experiments. It is also useful for data storage, which makes it convenient for later reexamination and modification of network structures. Other tools including ARACNE, GENIE3, and GeneMANIA, all of which are made to infer GRN structures from large expression profiles such as microarray datasets. These tools are often used in mammalian systems and are designed to scale up the complexity of regulatory networks with algorithms that predict interactions based on functional association, gene co-expression, and protein co-localization.

Perturbation analyses can indicate where regulatory linkages may exist; however, they are limited in their ability to distinguish between direct and indirect interactions. An example is a coherent feedforward loop, where gene A activates gene B and both A and B are needed for the activation of gene C. Perturbation analyses may fail to discover this sub-circuit, and will instead interpret it as a sequential activation where A activates B and B then activates C. In other cases, perturbation analyses may fail to predict regulatory interactions regardless of whether they are direct or indirect. An example of this is an incoherent feedforward loop, where gene A activates both gene B and gene C, while B represses C. Perturbation of A can disrupt the activation of C that is mediated by its own product, but it will also disrupt the repression regulated by B. The expression change of C therefore may be undetectable due to the loss of opposing regulators. In this case, the interactions can only be resolved by cis-regulatory analysis of gene C.

UNDERSTANDING DIFFERENTIAL GENE EXPRESSION VIA CIS-REGULATORY ANALYSIS (CRA)

Sea urchin zygotic gene transcription starts during early cleavages. Differential gene expression can be already observed by the 5th cleavages, approximately 6-7 hour post fertilization. At this time, *foxq2* is specifically expressed in the 8 mesomeres at the animal hemisphere, while *eve* is expressed in the 4 macromeres and 4 micromeres at the vegetal half of the embryos (Li, Cui et al. 2014) This differential expression of *foxq2* and *eve* is regulated by maternal β -catenin (Cui et al. 2015). These two genes are likely the earliest known zygotic regulatory markers for animal and vegetal fates, respectively.

The genome contains all the instructions necessary to process the maternal anisotropies, signaling transduction, and TF inputs and to transform them into differential patterns of gene expression. The sequences that provide these instructions for gene expression are called regulatory sequences, and are often referred to as cis-regulatory elements. They function as integrated TF binding platforms, and are recognized both by major lineage specifiers and DNA binding effectors of signaling pathways, allowing them to determine where in embryos and when during development to activate or repress the expression of a given gene. Characterized by their functions, these elements are classified as “promoters” which

are proximal to the genes and are essential for transcription initiation, “repressors” which can inhibit the expression of genes, “insulators” which obliterate enhancer–promoter interaction when present between them preventing non-cognate enhancer–promoter crosstalk, and “enhancers” that can be bound by regulatory factors to interact with promoters therefore to enhance the transcription of genes (Matharu and Ahituv 2015). Among these elements, enhancers exhibit wide functional diversity between tissues, suggesting that these elements have an important role in determining tissue specific gene expression (Heintzman, Hon et al. 2009). Identifying cis-regulatory elements and annotating functional TF binding sites within them provides crucial insights into how specific patterns of gene expression are achieved. This is also an invaluable approach for confirming, revising, and also constructing network linkages. Below is a summary of the essential steps involved in cis-regulatory analysis, with a focus on the methods applied in sea urchins in addition to some discussion of approaches in mouse and *Drosophila* models. A summary of experimental procedures is illustrated in Figure 2.

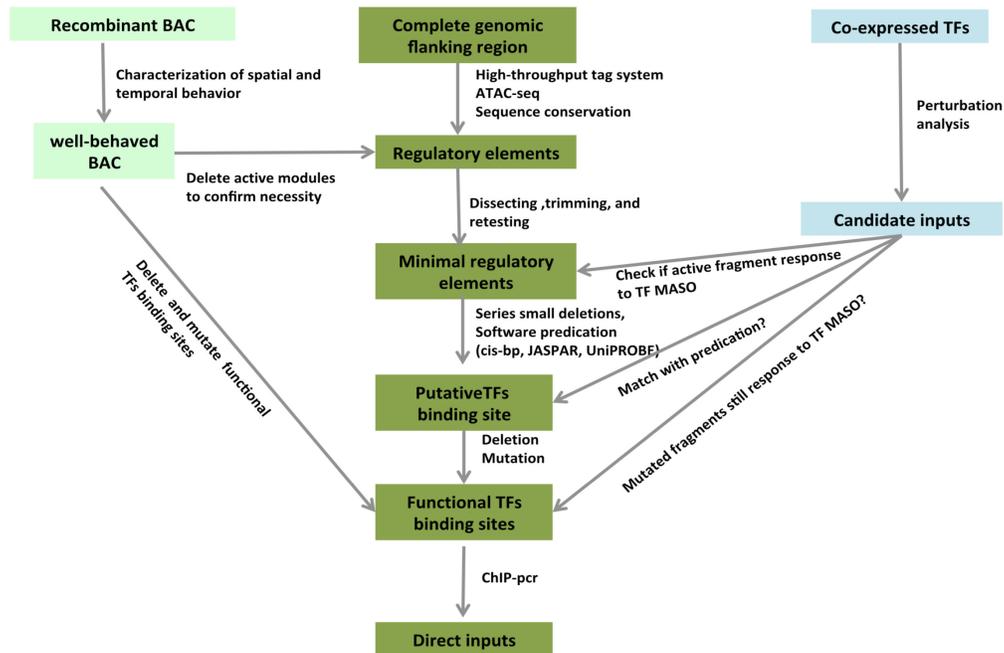


Figure 2. A pipeline for cis-regulatory analysis.

This flowchart summarizes the experimental design for cis-regulatory analysis. Boxes represent each step in the analysis and the lines indicate the experimental procedures to be conducted in order to proceed to the next step. Three paths are carried in parallel that each starts with a BAC reporter construct (light green), the genomic sequence of the gene (dark green), and the co-expressed TFs (blue), respectively.

Identification of cis-regulatory elements

A large bacterial artificial chromosome (BAC) library with average of 100kb of sequence per clone is a helpful resource to initiate CRA. A BAC covering the gene of interest can be identified by northern blot from the BAC library using gene specific probes. Homologous recombination is then used to generate a BAC reporter construct by knocking a reporter gene in frame into the first exon. When

injected into sea urchin embryos, the BAC reporter can give quantitative as well as spatial information of the regulatory functions of the sequence contained therein. This information is subsequently compared to endogenous patterns of gene expression in order to confirm whether the BAC sequence is sufficient to generate correct gene expression. Once confirmed, the BAC reporter will highlight a genomic region that can be screened for smaller cis-acting elements as well as a sequence platform for the testing of the necessity of these elements.

There are two general approaches used to identify cis-regulatory elements: candidate approach and unbiased systematic approach. In the candidate approach, phylogenetic footprinting is a useful means of finding regulatory sequences by comparing these sequences in different sea urchin species. The rationale behind this is that the regulatory sequences are more likely to be constrained during evolution relative to other DNA sequences that do not contribute to gene expression. Another candidate approach that has just been recently adopted is using ATAC-seq data. ATAC-seq analysis is a transposon-based method that measures chromatin accessibility (Buenrostro, Giresi et al. 2013, Buenrostro, Wu et al. 2015). DNA sequences where the transposase integrates are indicated as regions of open chromatin, which often correlate with the functional sequences. ATAC-seq data for 7 stages of sea urchin *Strongylocentrotus purpuratus* development covering 18 to 70hr post-fertilization are now available at the EchinoBase (<http://www.echinobase.org/Echinobase/>). This comprehensive dataset is useful for predicting candidate regulatory sequences. However, this method is not a direct

measurement of DNA regulatory activity, and as such any region that binds protein complexes and prevent chromatin occupancy, including gene bodies and insulators, would show up as positive results. To overcome this caveat, looking for regions that show dynamic chromatin accessibility across many development stages (it is known that regulatory elements drive gene expression in a time dependent manner) may allow for the elimination of false positive results.

In other systems such as mouse and *Drosophila* models, a variety of other candidate approaches have been applied towards the discovery of regulatory elements. These include epigenetic profiling of histone signatures such as H3K27ac and H3K4me1 which mark enhancers, DNase-seq/MNase-seq/FAIRE-seq that also detect open chromatin regions, and chromosome conformation capture based methods (3C, 4C, 5C, Hi-C, and ChIA-PET) that utilize nuclear proximity ligation to assess physical interaction between enhancer and promoter. The sea urchin has not yet adapted to these methods due to the difficulty of obtaining sufficient homogenous cell population as well as a lack of good antibodies. Nonetheless, it is believed that these methods will become feasible in the sea urchin in the future and will greatly accelerate the speed of regulatory element discovery.

The potential regulatory regions identified by the above described candidate approaches are all based on indirect measurement of enhancer activity, and thus must undergo functional validation *in vivo*. In sea urchins this validation is carried out using a reporter assay in which candidate sequences are placed in front of either

a gene endogenous basal promoter or a heterologous basal promoter fused to a reporter gene – the basal promoters of *endo16*, *gatae*, and *nodal* are often used in this assay. The reporter constructs are then introduced into embryos via injection at the 1-cell stage, and they then concatenate and incorporate into the genome. The regulatory activity of the potential element is then characterized by measuring the temporal and spatial expression of the reporter.

As for an unbiased systematic approach, every single nucleotide sequence of the BAC is subjected to testing for regulatory capability. The establishment of a high-throughput reporter system by Nam *et al.* (Nam, Dong et al. 2010, Nam and Davidson 2012), in which each reporter gene is coded with an unique barcode that can be distinguished by either qPCR primers or Nanosting nCounter probes, has greatly increased the scale of these analyses in sea urchins. By pooling many barcoded reporter constructs together, it is possible to simultaneously measure the regulatory activity of up to 129 different DNA sequences. Similar reporter systems with even higher throughput benefitting from next generation sequencing technologies have also been developed in mouse and *Drosophila* models. STARR-seq is an approach first developed in *Drosophila* which is able to assess enhancer activity for millions of DNA fragments during a single experiment (Arnold, Gerlach et al. 2013). Its unique design relies upon DNA sequences that are placed downstream of a minimal promoter, allowing active enhancers to transcribe themselves, serving as their own barcodes. The activity level of each enhancer is reflected by its richness among cellular RNAs. Massive Parallel Reporter Assay is

another approach that was developed in the mammalian system (Smith, Taher et al. 2013). Its design principle resembles the high-throughput barcode reporter assay in sea urchins. The main difference is that it utilizes a randomly synthesized sequence for tagging the luciferase reporter, which then can be read out by RNA sequencing. Both STARR-seq and Massive Parallel Reporter Assay rely on viral infection for delivery. The reporter constructs introduced into cells remain in an extrachromosomal form during development and do not incorporate into the genome, which needs to be taken into consideration when interpreting the results generated using these approaches.

The reporter-based screening approach is undoubtedly an efficient means of discovering enhancers. However, it has two caveats: 1) it can only identify enhancers and not other regulatory elements such as repressors; 2) it only measures the regulatory capacity of DNA fragments in an isolated form out of its genomic context, and it thus does not test the necessity of nor capture the interactions among regulatory elements. To overcome these shortcomings, studying a regulatory element in its native genomic environment is particularly helpful. Deletions of regulatory elements in the genome can be achieved in mice, *Drosophila*, and some other model systems through recombination systems such as Cre/lox and FLP/FRT, or through genome editing techniques such as CRISPR/cas9 and TALEN. These methods have not been well established in the sea urchin, and an alternative approach is the use of a BAC library, which is a useful system to conduct sequence manipulation (i.e. deletions and mutations) that potentially mimics the sequence

changes in the genomic context.

Identification of functional binding sites and associated trans-acting factors

After the discovery of active regulatory elements, a series of deletions need to be conducted at both ends of this sequence to search for the smallest fragment which is necessary and sufficient for the reporter activity called the “minimal element”. The length of minimal elements can vary from 200bp to 1kb; however, the smaller the minimal sequence isolated, the easier it is to identify transcription factor binding sites located therein. The next step is to locate positions of functional sequences, which can be also done through deletion series. Deletions are either designed to sequentially target the entire region of the minimal elements or to progressively trim down these sequences from both ends. The deleted regions containing regulatory function will be revealed by measuring the reporter activity of the deleted construct compared to the wild type reporter construct. These functional sequences are then subjected to be assessed for transcriptional binding sites.

Several software packages are available to identify putative transcription factor binding sites, including Jaspar (<http://jaspardev.genereg.net>), UniProbe (<http://thebrain.bwh.harvard.edu/uniprobe/index.php>), and cis-bp (<http://cisbp.cabr.utoronto.ca>). All of these databases contain a collection of transcription factor binding motifs that are either experimentally identified or generated as position weight matrices by SELEX sequencing. A search using any of

these databases will generally turn up many false positives (binding sites that are not in fact functional). A feature of cis-bp that allows for the scanning of differential binding sites between two sequences can potentially filter out false positive sites when an inactive sequence is used as negative control. Once the putative binding sites have been identified, these candidate sites are disrupted by PCR-directed point mutagenesis. Multiple binding sites can be mutated altogether in one construct or separately in different constructs. These constructs are next barcoded and pooled into one injection for the simultaneous measurement of the effects of these mutations on the ability of these regulatory elements to drive reporter expression. It is important to examine both the quantitative effects but also the spatial changes of these mutations. For example, a mutation causing ectopic expression that doubles the number of cells expressing a reporter gene would not show a significant increase in reporter activity when measured by qPCR. To assess the spatial effects, mutated reporter constructs need to be examined individually, and wild-type constructs bearing a different reporter can be used as internal controls. To test for regulatory necessity it is also important to perform site disruption in a larger construct, such as the BAC reporter itself or even in the genome.

After identifying functional binding sites, the next step is to connect these sites to their trans-acting transcriptional factors. Binding sites of closely related factors from the same family may sometimes be indistinguishable and ambiguous; this is often the case when cross-searching databases of other species. For example, a sea

urchin sequence may contain hits for Hoxa1 and Hoxa13 binding sites predicted by mouse databases. However, these sites are not necessarily associated with the binding of the sea urchin homologs Hox1 and Hox11/13 and may instead function through other Homeodomain proteins. It is thus necessary to always confirm the binding factors using perturbation analyses, wherein potential regulatory factors are knocked down using morpholinos and expression changes of the endogenous gene and minimal reporter constructs are measured. In many cases perturbation is also not sufficient to confirm the direct interaction of transcription factors and their putative binding sites. An example is a sequential activation involving two activators of same transcription factor family (e.g. forkhead family): forkhead factor A activates forkhead gene B which in turn activates gene C ($A \rightarrow B \rightarrow C$). CRA of C may identify functional forkhead protein sites and perturbation of A can decrease the expression of C, and therefore it is easy to interpret A as the direct activator of C and to fail to identify B. In cases like this, direct measurement of a physical interaction between transcription factors and their binding sites using ChIP-based experiments (ChIP-pcr and ChIP-seq) or in vitro electrophoretic mobility shift assay (EMSA) will be helpful.

THE ENDODERM GENE REGULATORY NETWORKS

The above has provided an overview of sea urchin embryogenesis, an approach for GRN model construction, and methods for cis-regulatory analysis. Below I summarize the key features of endoderm specification in the sea urchin embryo

from the perspective of the underlying gene regulatory network. The specification of sea urchin endoderm and the underlying GRN mechanisms are shown in Figure 3.

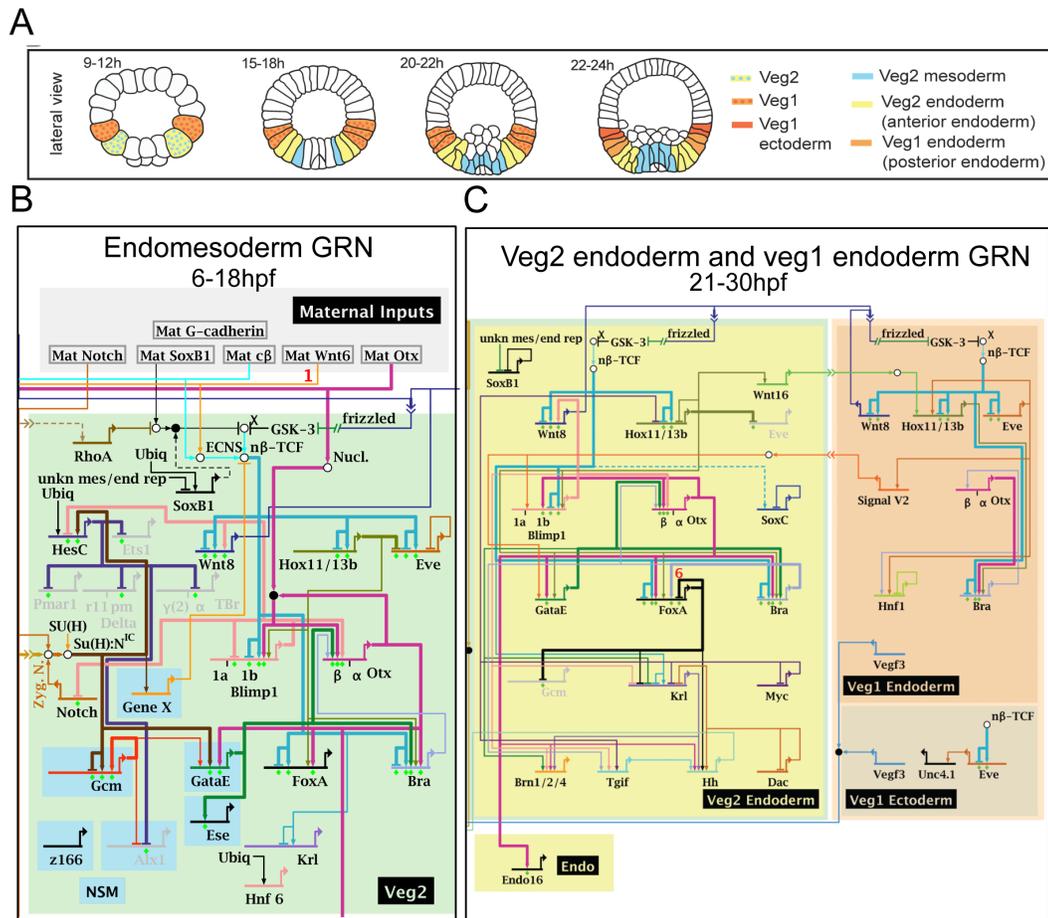


Figure 3. The specification of endoderm lineages and the underlying gene regulatory structure. (A) Schematic representation of the specification processes that lead to the formation of endodermal lineages. Two precursor cell lineages in which these processes occur are veg2 and veg1 macromeres. A common endomesoderm regulatory state is active in the veg2-derived cells before the separation of veg2 endoderm and veg2 mesoderm (also known as non-skeletogenic mesoderm), which occurs between 15-18hpf. A second set of endoderm derives from the inner veg1 cells that are in an immediate contact with the veg2 cells and become the posterior endoderm. (B) The endomesoderm GRN. Regulatory interactions between transcription factors are shown as activating arrows or repressing bars. Regulatory genes turned on exclusively in mesoderm precursor cells are shown in blue boxes. This model summarizes all interactions in the

endomesoderm up to 18hpf. For individual time points for each interactions as well the interactions between and within the neighboring lineages at the same stage, please visit <http://sugp.caltech.edu/endomes/>. (C) Anterior and posterior GRNs between 21 to 30hpf.

Endomesoderm GRN and separation between NSM and the endoderm

The sea urchin anterior endoderm and non-skeletogenic mesoderm, also known as the veg2 mesoderm, are both derived from the veg2 lineages and share a common cell lineage ancestry during the early development (Peter and Davidson 2010). These two cell fates become spatially separated by the exclusive activation of respective specification programs in a subset of cells. This complete discrimination of endoderm and mesoderm precursor cells in the veg2 lineages occurs after the mid-blastula stage at about 18 hours post-fertilization (Figure 3A).

Before mid-blastula stage, cells of veg2 progeny share the same regulatory state governed by the endomesoderm GRN (Figure 3B). The initial inputs into this GRN are largely maternal and include maternal β -catenin, maternal Sox1 and maternal Otx. Several endodermal regulatory genes have been shown to be specifically expressed in these cells, among which the earliest expressed genes are *wnt8*, *eve*, *blimp1b*, and *hox11/13b*. The expression of *wnt8* begins as early as the 4th cleavage in the micromeres, and later expands to the veg2 macromeres. At 15h post-fertilization it is specifically expressed in the veg2 cells, and by 18h, it becomes excluded from these cells and is instead only expressed in the adjacent veg1

progeny cells. The expression of *eve* exhibits a similar dynamic. Its expression starts at 5th cleavage (about 7h post-fertilization) in both macromeres and micromeres. A few hours later, at 12h post-fertilization, its expression is specific to the veg2 cells, and then become restricted to the veg1 cells after 15h post-fertilization. *Blimp1b* has maternal transcripts; its zygotic expression starts from the micromere and is later specific to the veg2 cells. *Hox11/13b* expression begins at about 9h post-fertilization, and its expression is exclusive to the veg2 cells before 18h post-fertilization. A second wave of gene activation results in the transcription of *foxa* and *brachyury (bra)*. The expression of these two genes occurs in early blastula embryos at 12h and is specific to the veg2 cells.

Tcf and the co-activator β -catenin are responsible for initiating the expression of endodermal regulatory genes in the endomesoderm GRN. Cis-regulatory analyses of *blimp1b*, *foxa*, and *hox11/13b* have shown that Tcf directly contributes to the activation of these genes in veg2 cells where nuclear β -catenin is present, whereas it represses the expression of the same genes in other cells where nuclear β -catenin is replaced by the repressor Groucho (Smith and Davidson 2008, de-Leon and Davidson 2010). The differences in the timing of the initial activation and the spatial dynamics of these Tcf target genes indicate that other regulatory inputs are required for their expression. Indeed, cis-regulatory studies have discovered that, besides Tcf/ β -catenin, the expression of *blimp1b* also requires Otx (Smith and Davidson 2008); similarly, the expression of *foxa* requires Otx and Hox11/13b (de-Leon and Davidson 2010).

The activation of the mesodermal regulatory program begins between the 7th and 9th cleavage (~9-15phf), when Delta ligands are expressed in the large micromeres and induce the specification of NSM in the adjacent veg2 tier where its receptor Notch is present. *Gcm*, a direct target gene of Delta/Notch signaling, is initially expressed in all 16 veg2 descendants at 7th cleavage and is required for the activation of the NSM specification GRN (Ransick and Davidson 2006). The expression of *gcm* activates *gatae*, whose activation in turn initiates the expression of mesodermal gene *ese*. Before the separation, endomesoderm precursor cells co-express transcription factors required for endoderm and mesoderm specification, and there seems to be little regulatory interaction between the two GRNs. After the 8th cleavage, the veg2 cells divide vertically to form two tiers of 16 cells, of which only the inner ring is in the direct contact with micromeres. The separation of mesoderm and endoderm in the veg2 progeny begins at this time, when *gcm* is only expressed in the inner veg2 cells.

By 18h post-fertilization, NSM and veg2-derived endoderm have permanently separated their fates. Endodermal regulatory genes such as *hox11/13b*, *foxa*, and *blimp1* are no longer expressed in the mesoderm precursor cells located in the inner ring of veg2 lineage and are specifically expressed in the outer veg2 cells. The clearance of endodermal genes from the NSM also depends on Delta/Notch signaling. Evidence shows that perturbation of either Delta or Notch caused *foxa* and *blimp1b* to continue to be expressed in the NSM at 24h (Croce and McClay

2010, Sethi, Wikramanayake et al. 2012). A cis-regulatory study of *foxa* showed that Tcf target sites are also required for this clearance: mutation of Tcf sites caused *foxa* transcripts to linger in the mesodermal cells (de-Leon and Davidson 2010). Thus, the same Tcf sites that are used to activate the expression of endodermal genes execute the opposite function and repress the expression of the same genes in mesoderm precursor cells. One possible explanation is that Delta/Notch signaling in the mesodermal cells interferes with the nuclear localization of β -catenin, leading to the Tcf/Groucho-mediated repression.

Veg2 endoderm (anterior endoderm) GRN

Positioned at the top of veg2 endoderm GRN is *hox11/13b* (Peter and Davidson 2011). Transcripts of *hox11/13b* are transiently expressed in the veg2 endoderm; after 21h, they become only detectable in the adjacent veg1 cells that later give rise to the hindgut (Peter and Davidson 2011). Despite its transient expression, Hox11/13b plays a prominent role in activating the veg2 endoderm GRN (Figure 3C). It provides activation linkages for many important factors including *blimp1b* and *foxa*. Another very important early coordinator of the veg2 endoderm GRN is the Wnt signaling machinery (Cui et al. 2015). Perturbation of Wnt signaling by small molecules interfering with Wnt protein secretion significantly decreases the expression level of almost all endodermal regulatory genes. Through morpholino perturbation of the individual wnt genes, we now know that this activation function is executed specifically by Wnt1 and Wnt16 (for a detailed description of the

analysis refer to Chapter 2).

Because *hox11/13b* is no longer expressed in the *veg2* endoderm after 21h, additional activators must be required to maintain the *veg2* endoderm-specific expression of *foxa* and *blimp1b*. Cis-regulatory studies have identified Otx as their direct activator. Otx is partially supplied by maternal transcripts, and its zygotic expression depends on its own proteins, GataE, and Blimp1b, which are themselves regulated by Otx (Smith, Theodoris et al. 2007). The expression of *otx* is therefore maintained by multiple positive feedback loops. These genes involved in the feedback regulation of *otx* form a sub-circuit that is referred to as the endodermal kernel. This kernel is essential for endodermal development and is conserved even in the distantly related starfish *Asterina miniata* (Hinman and Davidson 2007).

Additional players in the network include *myc*, *krl*, *brn1/2/4*, *tgif*, *Hh*, and *dac*. *Myc* and *krl* are expressed before 18h and are likely the immediate output of the endodermal kernel. *Brn1/2/4*, *tgif*, *Hh*, and *dac* are expressed a few hours before the onset of gastrulation. In this time period, the number of regulatory interactions has increased markedly. These later activated endoderm genes are at the periphery of the pre-gastrula endoderm GRN; nevertheless, they still receive regulatory inputs both from the top (endoderm kernel: *otx*, *blimp1b*, *gatae*, and *foxa*) and the middle (*krl* and *myc*) of the network hierarchy.

Before skeletogenic mesoderm cells ingress to the blastocoel, the NSM mesodermal

GRN runs independently of the *veg2* endodermal GRN, as none of the mesodermal genes were affected in embryos injected with a *hox11/13b* morpholino (Peter and Davidson 2010). However, soon after the skeletogenic mesoderm cell ingression, the expression of *delta* is activated in the NSM cells (Materna, Ransick et al. 2013). The *veg2* endoderm is then in physical contact with the *delta* expressing cells, and has the potential to activate the expression of *gcm* and initiate the pigment cells specification program. To avoid doing this, the endoderm GRN now has activated its “defense” mechanism, wherein *foxa* is employed to repress the expression of *gcm* in order to prevent pigment specification. In the embryos injected with a *foxa* morpholino, not only do the embryos fail to specify gut, but they also have an elevated number of pigment cells, which can be explained by de-repression of *gcm* (Oliveri, Walton et al. 2006).

Veg1 endoderm (posterior endoderm) GRN

The GRN specifying the *veg1* endoderm is shown in Figure 3 C. Within the *veg1* lineages only the inner cells in the proximal regions of *veg2* cells specify to endoderm, whereas the outer cells give rise to the vegetal ectoderm (Ransick and Davidson 1998, Peter and Davidson 2011). The specification of *veg1* endoderm starts at mesenchyme blastula stage (about 21h). The first gene specifically expressed in these cells is *hox11/13b*, whose activation at this time is regulated by *Eve* (Peter and Davidson 2011), and *Wnt1/Wnt16* (Cui et al. 2015). *Eve* expression

spatially defines the veg1 regulatory state beginning at 15h, but its regulatory function to *hox11/13b* can be detected only after mesenchyme blastula stage. *Eve* morpholino showed no effect on any regulatory genes other than itself at 18h. *Wnt1* and *wnt16* are required for the activation of *hox11/13b* in the veg1 endoderm, and are also target genes of Hox11/13b regulation (Cui et al. 2015). The earlier expression of *hox11/13b* in the veg2 endoderm activates the transcription of *wnt1* and *wnt16*. *Wnt1* and *wnt16* are then translated and secreted, interacting with the adjacent veg1 cells to activate the expression of *hox11/13b*, which in turn again activates *wnt1* and *wnt16* transcription in these same cells. Thus, the veg1 endoderm GRN is temporally controlled by Wnt signaling expressed under control of the veg2 endoderm GRN, and is then maintained by a positive feedforward regulatory loop formed between *wnt1/wnt16* and *hox11/13b*.

Following *hox11/13b*, *bra* is also expressed in the veg1 endoderm after 24h. Its dynamic spatial expression is mostly a function of regulatory inputs provided by Hox11/13b and Tcf in both veg2 endoderm and veg1 endoderm. Another veg1 endoderm regulator is *hnf1*, whose expression is regulated by both *Eve* and *Bra*. It is initially expressed at approximately 24h and is specific to the veg1 endoderm.

The precise activation of the veg1 endodermal GRN completely depends upon the spatial and temporal regulation of *hox11/13b*. For the network to operate exclusively in the inner cells of the veg1 lineage, expression of *hox11/13b* must be repressed in the veg2 endoderm and also not to be activated in the outer cells of the

veg1 lineage. This repression in the veg2 endoderm is mediated by auto-regulation. Perturbation of its own translation causes ectopic accumulation of *hox11/13b* transcripts in the veg2 endodermal cells and also abolishes its expression in the veg1 endoderm (Peter and Davidson 2011). An interesting question is what keeps the expression of *hox11/13b* out of the veg1 ectoderm especially considering the activator Eve is ubiquitously expressed in the veg1 lineage. The spatial expression of *hox11/13b* is also important in defining the boundary between the endoderm and the ectoderm. Perturbation of *hox11/13b* causes expansion of vegetal ectodermal genes, including *lim1*, *veg3*, and *hox7*, to the veg1 endoderm (Li, Cui et al. 2014).

Theme of thesis

As described above, the cellular fates of most lineages in sea urchin embryos are achieved through conditional specification suggesting the importance of signaling transduction in mediating embryonic partition. The availability of advanced GRN models of sea urchin development presents a unique opportunity to address the functions of signaling interactions in cell fate specification at the system-wide level. Specific questions are: 1) What are the regional specification GRNs that signaling pathways have functions into, either activating or repressing, within the whole embryos? 2) Which level of the GRN hierarchy are these signaling interactions feeding into? 3) How are the signaling molecules incorporated into their functional GRNs and can we find a general structure that explains why certain function of

signaling interaction is so conserved across different organisms while the same signaling can also be deployed for a specialized function? Only a system like sea urchins that has comprehensive morphological as well molecular understanding of cell fate specification in particular the GRN structure is equipped for addressing those questions.

To this end, we studied the function of Wnt signaling. As summarized above, it is known that Wnt signaling is crucial for the activation of the endoderm GRNs. However, previous research has often focused on the study of selected *wnt* genes in isolated developmental context, e.g. only the mesodermal cells or endodermal cells. Prior to this work we still lacked of global view of how the Wnt signaling system is utilized in the GRNs specifying embryonic lineages. In Chapter 2, I show that Wnt signaling is not required during early development, due to maternal nuclear β -catenin, but later specifically regulates endodermal as well as vegetal ectodermal GRNs. I also show that Wnt ligands function by short range signaling between and within regulatory state domains. Furthermore, I show that *wnt1* and *wnt16* incorporate into the posterior endoderm GRN by forming a positive feedback circuit with posterior endoderm specifier *hox11/13b*. In addition, I show the specific regulatory functions of Wnt ligands in embryonic patterning along the primary vegetal-animal axis.

The importance of *hox11/13b* in endoderm specification has been described at

length in the above text. It activates both the veg2 endoderm and veg1 endoderm GRNs. Its spatial expression is essential for the discrimination of cell fate between the veg2 endoderm and the veg1 endoderm, as well as between the veg1 endoderm and the veg1 ectoderm. It is thus fair to say that our understanding of the endoderm GRN is only as good as our understanding of information processing at the *hox11/13b* locus. To this end, I conducted cis-regulatory analyses of the *hox11/13b* locus. Chapter 3 summarizes this unpublished study, wherein the genomic control of *hox11/13b* expression has been systematically dissected across an extensive developmental timeframe up to 60h post-fertilization. This study has confirmed the roles of current network inputs (Eve and Hox11/13b) as indirect regulators, and also identified other new inputs. More importantly, it clarified the logic by which different factors function combinatorially to control the expression of *hox11/13b* at different times and in different domains. Furthermore, the study found an inter-modular AND regulatory gate that is required for the expression of *hox11/13b* during the later development in the hindgut. Finally, Chapter 4 concludes with a brief discussion of some implications of these studies.

REFERENCES

Arnold, C. D., D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath and A. Stark (2013). "Genome-wide quantitative enhancer activity maps identified by STARR-seq." Science **339**(6123): 1074-1077.

Barsi, J. C., E. Li and E. H. Davidson (2015). "Geometric control of ciliated band regulatory states in the sea urchin embryo." Development **142**(5): 953-961.

Barsi, J. C., Q. Tu, C. Calestani and E. H. Davidson (2015). "Genome-wide assessment of differential effector gene use in embryogenesis." Development **142**(22): 3892-3901.

Barsi, J. C., Q. Tu and E. H. Davidson (2014). "General approach for in vivo recovery of cell type-specific effector gene sets." Genome Res **24**(5): 860-868.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21 29 21-29.

Croce, J. C. and D. R. McClay (2010). "Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo." Development **137**(1): 83-91.

Cui M, S. N., Li E, Davidson EH, Peter IS. (2014). "Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo." Proc Natl Acad Sci U S A **111**(47): E5029-5038.

Davidson, E. H. (2006). "The Regulatory Genome: Gene Regulatory Networks in Development and Evolution." Academic Press/Elsevier, San Diego.

Davidson, E. H., R. A. Cameron and A. Ransick (1998). "Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms." Development **125**(17): 3269-3290.

Davidson, E. H., J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C. H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. Pan, M. J. Schilstra, P. J. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood and H. Bolouri (2002). "A genomic regulatory network for development." Science **295**(5560): 1669-1678.

De-Leon, S. B. and E. H. Davidson (2010). "Information processing at the foxa node of the sea urchin endomesoderm specification network." Proc Natl Acad Sci U S A **107**(22): 10103-10108.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-

Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenko, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.

Hinman, V. F. and E. H. Davidson (2007). "Evolutionary plasticity of developmental gene regulatory network architecture." Proc Natl Acad Sci U S A **104**(49): 19404-19409.

Howard-Ashby, M., S. C. Materna, C. T. Brown, L. Chen, R. A. Cameron and E. H. Davidson (2006). "Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*." Dev Biol **300**(1): 90-107.

Howard-Ashby, M., S. C. Materna, C. T. Brown, L. Chen, R. A. Cameron and E. H. Davidson (2006). "Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development." Dev Biol **300**(1): 74-89.

Kenny, A. P., D. Kozlowski, D. W. Oleksyn, L. M. Angerer and R. C. Angerer (1999). "SpSoxB1, a maternally encoded transcription factor asymmetrically distributed among early sea urchin blastomeres." Development **126**(23): 5473-5483.

Kumburegama, S. and A. H. Wikramanayake (2008). "Wnt signaling in the early sea urchin embryo." Methods Mol Biol **469**: 187-199.

Leonard, J. D. and C. A. Ettensohn (2007). "Analysis of dishevelled localization and function in the early sea urchin embryo." Dev Biol **306**(1): 50-65.

Li, E., M. Cui, I. S. Peter and E. H. Davidson (2014). "Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo." Proc Natl Acad Sci U S A **111**(10): E906-913.

Li, E., S. C. Materna and E. H. Davidson (2012). "Direct and indirect control of oral ectoderm regulatory gene expression by Nodal signaling in the sea urchin embryo." Dev Biol **369**(2): 377-385.

Li, E., S. C. Materna and E. H. Davidson (2013). "New regulatory circuit controlling spatial and temporal gene expression in the sea urchin embryo oral ectoderm GRN." Dev Biol **382**(1): 268-279.

Materna, S. C., M. Howard-Ashby, R. F. Gray and E. H. Davidson (2006). "The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development." Dev Biol **300**(1): 108-120.

Materna, S. C., J. Nam and E. H. Davidson (2010). "High accuracy, high-resolution prevalence measurement for the majority of locally expressed regulatory genes in early sea urchin development." Gene Expr Patterns **10**(4-5): 177-184.

Materna, S. C. and P. Oliveri (2008). "A protocol for unraveling gene regulatory networks." Nat Protoc **3**(12): 1876-1887.

Materna, S. C., A. Ransick, E. Li and E. H. Davidson (2013). "Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos." Dev Biol **375**(1): 92-104.

Matharu, N. and N. Ahituv (2015). "Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease." PLoS Genet **11**(12): e1005640.

Nam, J. and E. H. Davidson (2012). "Barcoded DNA-tag reporters for multiplex cis-regulatory analysis." PLoS One **7**(4): e35934.

Nam, J., P. Dong, R. Tarpine, S. Istrail and E. H. Davidson (2010). "Functional cis-regulatory genomics for systems biology." Proc Natl Acad Sci U S A **107**(8): 3930-3935.

Nam, J., Y. H. Su, P. Y. Lee, A. J. Robertson, J. A. Coffman and E. H. Davidson (2007). "Cis-regulatory control of the nodal gene, initiator of the sea urchin oral ectoderm gene network." Dev Biol **306**(2): 860-869.

Okazaki, K. (1975). "Spicule Formation by Isolated Micromeres of the Sea Urchin Embryo." American Zoology **15**: 567-581.

Oliveri, P., D. M. Carrick and E. H. Davidson (2002). "A regulatory gene network that directs micromere specification in the sea urchin embryo." Dev Biol **246**(1): 209-228.

Oliveri, P., Q. Tu and E. H. Davidson (2008). "Global regulatory logic for specification of an embryonic cell lineage." Proc Natl Acad Sci U S A **105**(16): 5955-5962.

Oliveri, P., K. D. Walton, E. H. Davidson and D. R. McClay (2006). "Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo." Development **133**(21): 4173-4181.

Peter, I. S. and E. H. Davidson (2010). "The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage." Dev Biol **340**(2): 188-199.

Peter, I. S. and E. H. Davidson (2011). "A gene regulatory network controlling the embryonic specification of endoderm." Nature **474**(7353): 635-639.

Peter, I. S. and E. H. Davdison, (2015). "Genomic Control Process: Development and Evolution." Academic Press, San Diego ISBN 978-0-12-404729-7.

Poustka, A. J., A. Kuhn, D. Groth, V. Weise, S. Yaguchi, R. D. Burke, R. Herwig, H. Lehrach and G. Panopoulou (2007). "A global view of gene expression in lithium and zinc treated sea urchin embryos: new components of gene regulatory networks." Genome Biol **8**(5): R85.

Range, R. and T. Lepage (2011). "Maternal Oct1/2 is required for Nodal and Vg1/Univin expression during dorsal-ventral axis specification in the sea urchin embryo." Dev Biol **357**(2): 440-449.

Range, R. C., R. C. Angerer and L. M. Angerer (2013). "Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos." PLoS Biol **11**(1): e1001467.

Ransick, A. and E. H. Davidson (1998). "Late specification of Veg1 lineages to endodermal fate in the sea urchin embryo." Dev Biol **195**(1): 38-48.

Ransick, A. and E. H. Davidson (2006). "cis-regulatory processing of Notch signaling input to the sea urchin glial cells missing gene during mesoderm specification." Dev Biol **297**(2): 587-602.

Rizzo, F., M. Fernandez-Serra, P. Squarzoni, A. Archimandritis and M. I. Arnone (2006). "Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*)." Dev Biol **300**(1): 35-48.

Sethi, A. J., R. M. Wikramanayake, R. C. Angerer, R. C. Range and L. M. Angerer (2012). "Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos." Science **335**(6068): 590-593.

Smith, J. and E. H. Davidson (2008). "Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo." Proc Natl Acad Sci U S A **105**(51): 20089-20094.

Smith, J., C. Theodoris and E. H. Davidson (2007). "A gene regulatory network subcircuit drives a dynamic pattern of gene expression." Science **318**(5851): 794-797.

Smith, R. P., L. Taher, R. P. Patwardhan, M. J. Kim, F. Inoue, J. Shendure, I. Ovcharenko and N. Ahituv (2013). "Massively parallel decoding of mammalian

regulatory sequences supports a flexible organizational model." Nat Genet **45(9)**: 1021-1028.

Sodergren, E., G. M. Weinstock, E. H. Davidson, R. A. Cameron, R. A. Gibbs, R. C. Angerer, L. M. Angerer, M. I. Arnone, D. R. Burgess, R. D. Burke, J. A. Coffman, M. Dean, M. R. Elphick, C. A. Etensohn, K. R. Foltz, A. Hamdoun, R. O. Hynes, W. H. Klein, W. Marzluff, D. R. McClay, R. L. Morris, A. Mushegian, J. P. Rast, L. C. Smith, M. C. Thorndyke, V. D. Vacquier, G. M. Wessel, G. Wray, L. Zhang, C. G. Elsik, O. Ermolaeva, W. Hlavina, G. Hofmann, P. Kitts, M. J. Landrum, A. J. Mackey, D. Maglott, G. Panopoulou, A. J. Poustka, K. Pruitt, V. Sapojnikov, X. Song, A. Souvorov, V. Solovyev, Z. Wei, C. A. Whittaker, K. Worley, K. J. Durbin, Y. Shen, O. Fedrigo, D. Garfield, R. Haygood, A. Primus, R. Satija, T. Sevrison, M. L. Gonzalez-Garay, A. R. Jackson, A. Milosavljevic, M. Tong, C. E. Killian, B. T. Livingston, F. H. Wilt, N. Adams, R. Belle, S. Carbonneau, R. Cheung, P. Cormier, B. Cosson, J. Croce, A. Fernandez-Guerra, A. M. Genevriere, M. Goel, H. Kelkar, J. Morales, O. Mulner-Lorillon, A. J. Robertson, J. V. Goldstone, B. Cole, D. Epel, B. Gold, M. E. Hahn, M. Howard-Ashby, M. Scally, J. J. Stegeman, E. L. Allgood, J. Cool, K. M. Judkins, S. S. McCafferty, A. M. Musante, R. A. Obar, A. P. Rawson, B. J. Rossetti, I. R. Gibbons, M. P. Hoffman, A. Leone, S. Istrail, S. C. Materna, M. P. Samanta, V. Stolc, W. Tongprasit, Q. Tu, K. F. Bergeron, B. P. Brandhorst, J. Whittle, K. Berney, D. J. Bottjer, C. Calestani, K. Peterson, E. Chow, Q. A. Yuan, E. Elhaik, D. Graur, J. T. Reese, I. Bosdet, S. Heesun, M. A. Marra, J. Schein, M. K. Anderson, V. Brockton,

K. M. Buckley, A. H. Cohen, S. D. Fugmann, T. Hibino, M. Loza-Coll, A. J. Majeske, C. Messier, S. V. Nair, Z. Pancer, D. P. Terwilliger, C. Agca, E. Arboleda, N. Chen, A. M. Churcher, F. Hallbook, G. W. Humphrey, M. M. Idris, T. Kiyama, S. Liang, D. Mellott, X. Mu, G. Murray, R. P. Olinski, F. Raible, M. Rowe, J. S. Taylor, K. Tessmar-Raible, D. Wang, K. H. Wilson, S. Yaguchi, T. Gaasterland, B. E. Galindo, H. J. Gunaratne, C. Juliano, M. Kinukawa, G. W. Moy, A. T. Neill, M. Nomura, M. Raisch, A. Reade, M. M. Roux, J. L. Song, Y. H. Su, I. K. Townley, E. Voronina, J. L. Wong, G. Amore, M. Branno, E. R. Brown, V. Cavalieri, V. Duboc, L. Duloquin, C. Flytzanis, C. Gache, F. Lapraz, T. Lepage, A. Locascio, P. Martinez, G. Matassi, V. Matranga, R. Range, F. Rizzo, E. Rottinger, W. Beane, C. Bradham, C. Byrum, T. Glenn, S. Hussain, G. Manning, E. Miranda, R. Thomason, K. Walton, A. Wikramanayke, S. Y. Wu, R. Xu, C. T. Brown, L. Chen, R. F. Gray, P. Y. Lee, J. Nam, P. Oliveri, J. Smith, D. Muzny, S. Bell, J. Chacko, A. Cree, S. Curry, C. Davis, H. Dinh, S. Dugan-Rocha, J. Fowler, R. Gill, C. Hamilton, J. Hernandez, S. Hines, J. Hume, L. Jackson, A. Jolivet, C. Kovar, S. Lee, L. Lewis, G. Miner, M. Morgan, L. V. Nazareth, G. Okwuonu, D. Parker, L. L. Pu, R. Thorn and R. Wright (2006). "The genome of the sea urchin *Strongylocentrotus purpuratus*." Science **314**(5801): 941-952.

Su, Y. H., E. Li, G. K. Geiss, W. J. Longabaugh, A. Kramer and E. H. Davidson (2009). "A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo." Dev Biol **329**(2): 410-421.

Tu, Q., C. T. Brown, E. H. Davidson and P. Oliveri (2006). "Sea urchin Forkhead gene family: phylogeny and embryonic expression." Dev Biol **300**(1): 49-62.

Tu, Q., R. A. Cameron and E. H. Davidson (2014). "Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*." Dev Biol **385**(2): 160-167.

Tu, Q., R. A. Cameron, K. C. Worley, R. A. Gibbs and E. H. Davidson (2012). "Gene structure in the sea urchin *Strongylocentrotus purpuratus* based on transcriptome analysis." Genome Res **22**(10): 2079-2087.

Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo

Cui M, S. N., Li E, Davidson EH, Peter IS. (2014). "Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo." Proc Natl Acad Sci U S A **111**(47): E5029-5038. doi: 10.1073/pnas.1419141111

ABSTRACT

Wnt signaling affects many specification processes throughout development. Here we take advantage of the well-studied gene regulatory networks (GRNs) that control pre-gastrular sea urchin embryogenesis, to reveal the gene regulatory functions of the entire Wnt signaling system. Five *wnt* genes, three *frizzled* genes, two *sfrp* genes, and two *dkk* genes are expressed in dynamic spatial patterns in the pre-gastrular embryo of *Strongylocentrotus purpuratus*. We present a comprehensive analysis of these genes in each embryonic domain. Total functions of the Wnt signaling system on regulatory gene expression throughout the embryo were studied by use of the porcupine inhibitor C59, which interferes with zygotic Wnt ligand secretion. Morpholino-mediated knock-down of each expressed Wnt ligand demonstrated that individual Wnt ligands are functionally distinct, despite

partially overlapping spatial expression. They target specific embryonic domains and affect particular regulatory genes. The sum of the effects of blocking expression of individual Wnt genes are shown to equal C59 effects. Remarkably, zygotic Wnt signaling inputs are required for only three general aspects of embryonic specification. These are broad activation of endodermal GRNs, regional specification of the immediately adjacent stripe of ectoderm, and the restriction of the apical neurogenic domain. All Wnt signaling in this pre-gastrular embryo is short range (and/or autocrine). Furthermore we show that the transcriptional drivers of *wnt* genes execute important specification functions in the embryonic domains targeted by the ligands, thus connecting expression and function of *wnt* genes by encoded cross-regulatory interactions within the specific regional GRNs.

INTRODUCTION:

The formation of spatial patterns of gene expression and the development of the body plan are controlled by gene regulatory networks (GRNs). Signaling interactions have a particular role in these networks, in that they provide the means of communication between cell fate specification processes operating in separate cellular domains. The timing, location, and function of each signaling interaction is determined by GRN linkages that control the expression of signaling ligands and receptors, as well as the expression of regulatory genes in response to a combination of signaling inputs and cell fate specific transcription factors. Cell fate specification GRNs active during pre-gastrular development of the sea urchin

Strongylocentrotus purpuratus are particularly well understood. During the first 30h of sea urchin embryogenesis, more than 15 gene expression domains are formed, and specifically expressed regulatory genes have been identified for each domain. In most cases, the regulatory mechanisms determining their spatial expression patterns have been resolved. Thus fairly complete GRN models have been constructed for the majority of cell fate domains in the pre-gastrula stage embryo (Oliveri, Tu et al. 2008, Peter and Davidson 2010, Peter and Davidson 2011, Li, Materna et al. 2013, Materna, Ransick et al. 2013, Li, Cui et al. 2014). The sea urchin GRN models at this point incorporate more than 60 regulatory genes and their interactions, and cover almost the entire embryo.

The principle organization of mesodermal, endodermal and ectodermal cell fate specification domains in sea urchin embryos along the animal-vegetal axis is summarized in the diagram in Fig. 1A. Cells located at the vegetal pole will become skeletogenic mesodermal cells. These cells are surrounded by the veg2 cell lineage, consisting of the veg2 mesodermal cells, located adjacent to skeletogenic cells and giving rise to all other mesodermal cell fates such as esophageal muscle cells, blastocoelar cells and pigment cells, and to veg2 endoderm cells, which will form the foregut and parts of the midgut. At a further distance from the vegetal pole, but still within the vegetal half of the embryo, is the veg1 lineage, consisting of veg1 endoderm, located adjacent to veg2 endoderm and giving rise to the other parts of the midgut and the hindgut, and of veg1 ectoderm, the future peri-anal ectoderm. Finally, the animal half of the embryo forms exclusively ectodermal cell

fates, with apical neurogenic cell fates being specified in cells at the animal pole.

The response to Wnt signaling is mediated by several alternative intracellular signaling pathways. In the canonical Wnt signaling pathway, signaling-dependent gene expression is controlled by the transcription factor Tcf/Lef, which forms a complex with the co-activator β -catenin in cells that are receiving Wnt signaling, but with the co-repressor Groucho in the absence of Wnt signaling (Range, Venuti et al. 2005). Transcriptional control by Tcf/Lef thus effects a Boolean readout of gene expression, mediating activation or repression of the same target genes in cells with or without Wnt signaling (Peter, Faure et al. 2012). *Cis*-regulatory analyses including mutation of Tcf binding sites have demonstrated direct control by Tcf in the skeletogenic GRN of *S. purpuratus* embryos, where the regulatory gene directly responsive to Tcf, *pmar1*, operates at the top of the GRN hierarchy (Oliveri, Davidson et al. 2003, Oliveri, Tu et al. 2008, Smith and Davidson 2009). Furthermore, most if not all regulatory genes expressed in early veg2 endoderm and/or veg1 endoderm cells of this embryo are direct targets of Tcf (Smith, Kraemer et al. 2008, Ben-Tabou de-Leon and Davidson 2010, Peter and Davidson 2010). Transcriptional control by Tcf is not only responsible for activation of endodermal genes in future gut cells, but is also used to exclude endodermal regulatory genes from the mesoderm (Ben-Tabou de-Leon and Davidson 2010, Croce and McClay 2010, Peter and Davidson 2011, Sethi, Wikramanayake et al. 2012). Furthermore, perturbation of particular Wnt ligand

gene expression has revealed two additional cell fate specification GRNs to be sensitive to Wnt signaling: one operating in ectodermal cells located closest to the endoderm, which responds to Wnt5 signaling (McIntyre, Seay et al. 2013), and one operating neuronal specification that is restricted to the apical domain by a mechanism dependent on Wnt signaling (Range, Angerer et al. 2013).

To understand the particular functions of signaling interactions, however, requires not only knowledge of affected target genes and cellular domains, but also identification of the cells producing the responsible signaling ligand. As in many other invertebrate embryos, the analysis of zygotic Wnt signaling functions in sea urchin embryos has been complicated by the presence of maternally localized β -catenin at the vegetal pole (Logan, Miller et al. 1999). Here we determined these zygotic Wnt signaling functions on a global scale, by assessing the temporal and spatial expression of all genomically encoded genes producing Wnt ligands, Frizzled (Fzd) receptors, or potential Wnt signaling antagonists during the pre-gastrular development of *S. purpuratus*. We have summarized these expression patterns abstractly, in order to highlight the signal-sending and signal-receiving capacity for each cell fate domain. We analyzed effects of interference with Wnt signaling on 172 specifically expressed regulatory genes, irrespective of the intracellular signaling pathways which might mediate this function. For a system-wide perturbation of Wnt signaling, we made use of the C59 inhibitor of Porcupine, which interferes with the secretion of Wnt ligands and thus with all Wnt-dependent

processes. We show that the only GRNs affected by C59 perturbation are the two endodermal GRNs, the veg1 ectoderm GRN, and the apical neurogenic GRN. For each GRN affected by C59-mediated inhibition of Wnt signaling, we identified the responsible Wnt ligands using morpholino perturbations. Furthermore, by identifying the upstream transcription factors activating *wnt* gene transcription, we established functional linkages between the GRNs regulated by Wnt signaling and the GRNs controlling Wnt ligand expression. Our intent was to achieve a system-wide understanding of the roles of Wnt signaling in this phase of development and for this embryo, and to generate a causal spatial regulatory analysis of Wnt signaling inputs into the regional embryonic GRNs.

RESULTS:

Spatial and temporal expression of *wnt*, *fzd*, *sfrp* and *dkk* genes

The Wnt signaling system encoded in the genome of *S. purpuratus* includes eleven *wnt* ligand genes and four *frizzled* (*fzd*) receptor genes (Croce, Wu et al. 2006). To identify *wnt* and *fzd* genes expressed during pre-gastrular development (12h-24h), the time courses of their expression levels were analyzed by QPCR (Fig. S1). Five *wnt* genes (*wnt1*, *wnt4*, *wnt5*, *wnt8*, and *wnt16*) are expressed in this embryo before gastrulation, only one of them transcribed maternally (*wnt16*; Fig. S1A,B). All other *wnt* genes are not expressed at all until gastrulation, and even then their transcript levels are very low (<100 transcripts per embryo). We cannot confirm the observation of maternal *wnt6* transcripts reported earlier (Croce, Range et al. 2011),

and this conclusion is substantiated in our recent transcriptome study (Tu, Cameron et al. 2014). Of the four *fzd* genes, *fzd1/2/7*, *fzd5/8*, and *fzd9/10*, were expressed at high levels (>1000 transcripts) before 30h, while *fzd4* begins to be transcribed only just before gastrulation and was not further considered in this study (Fig. S1C). Four genes encoding potential inhibitors of Wnt signaling were also found to be expressed during early sea urchin embryogenesis. These are genes encoding Dickkopf proteins, *dkk1* and *dkk3*, and two genes encoding secreted Frizzled-related protein (SFRP), *sfrp1/5* and *sfrp3/4* (Fig. S1D).

The spatial expression patterns of all expressed *wnt*, *fzd*, *dkk*, and *sfrp* genes during embryogenesis were analyzed at 3h intervals between 12h and 24h. Results for *wnt* and *fzd* genes at 12h, 18h, and 24h post fertilization are shown in Fig. 1B, and the complete data set is presented in Figs. S2 and S3. Similar results have been established in the embryos of a related species of sea urchin, *P. lividus* (Robert, Lhomond et al. 2014). Since the focus of this study is on identifying the function of Wnt signaling in the interaction between different cell fate specification processes, we have abstractly represented individual gene expression patterns according to their embryonic expression domain (Fig. 1C). In the following we summarize the expression patterns of genes encoding ligands, receptors, and potential antagonists of the Wnt signaling system by embryonic regulatory state domain (cf. Fig. 1A).

Skeletogenic mesodermal cells (SM) are the precursors of cells producing the larval skeleton, and are located at the vegetal pole before they start to ingress into the blastocoel at 21h. These cells inherit high levels of maternal nuclear β -

catenin and initially express *wnt1*, *wnt8*, and *wnt16*, but only up to 15h (Fig. 1B,C). No *wnt* gene is expressed in these cells thereafter. Thus after the degradation of maternal Fzd1/2/7 proteins, skeletogenic mesoderm cells are most likely not responsive to Wnt signaling, since no *fzd* transcripts are detectable in these cells.

Veg2 mesodermal cells give rise to all other (i.e., non-skeletogenic) mesodermal cell types including esophageal muscle cells, blastocoelar immune cells, coelomic pouch cells, and pigment cells. Veg2 mesodermal cells express *wnt4*, *wnt5*, and *wnt8* early in development at 12h and 15h, and transiently also *wnt1* and *wnt16* at 18h (Fig. 1B,C). However, after 18h and up to the onset of gastrulation, mesodermal precursor cells do not express *wnt* genes. Similarly, early expression of a *fzd* gene, *fzd9/10*, terminates after 15h, and only one *fzd* gene, *fzd5/8*, is transcribed after 21h in a subset of mesodermal cells, located in the oral portion of the veg2 mesoderm, consistent with earlier results (Croce, Duloquin et al. 2006). No *sfrp* or *dkk* genes are expressed in Veg2 mesodermal cells after 15h. Thus most mesodermal cells, with the exception of oral mesodermal cells, express neither *wnt* nor *fzd* genes after 18h and up to gastrulation, and are likely not to send nor receive Wnt signaling.

Veg2 endodermal cells are the precursors of anterior endoderm, forming the foregut and the aboral midgut (Ransick and Davidson 1998, Peter and Davidson 2011). These cells derive from the veg2 cell lineage, as do the mesodermal cells above. The common ancestor cells of the veg2 mesoderm and veg2 endoderm cells express *wnt4*, *wnt5*, and *wnt8* genes as well as *fzd9/10* at 12 and 15h. Endodermal and mesodermal cell fates become distinct in the veg2 lineage by 18h, and after that

veg2 endodermal cells transiently express *wnt1* and *wnt16* at 21h, while no *wnt* gene is expressed in this domain by 24h. Expression of *fzd9/10* continues until 18h, but no receptor gene is expressed in the anterior endoderm domain after that. Furthermore, after 18h, genes encoding potential Wnt signaling inhibitors, *sfrp3/4* and *dkk1*, start being expressed in veg2 endodermal cells, suggesting that these cells do not depend on Wnt signaling inputs after this time.

Veg1 endoderm cells are the precursors of posterior endoderm, eventually giving rise to the hindgut and the oral parts of the midgut. In an almost reversed pattern compared to anterior endoderm precursors, the posterior endoderm domain expresses no *wnt* genes before 18h, but by 24h all five *wnt* genes are expressed in these cells. These cells also transcribe *fzd9/10* at all times considered, while neither *sfrp* nor *dkk* genes are expressed after the early ubiquitous expression of *sfrp3/4*. These results indicate that the veg1 endoderm domain is capable of responding to Wnt signaling through Fzd9/10 throughout pre-gastrula stages, while they also may contribute to Wnt signaling after 18h.

Veg1 ectodermal cells derive from the veg1 lineage just as do the precursors of the posterior gut, and ultimately give rise to ectodermal cells surrounding the anus. From 18h on, before the separation of endodermal and ectodermal cell fates at 24h, veg1 cells express *wnt4*, *wnt5*, and *wnt8*, and these genes as well as *fzd9/10* continue to be expressed in veg1 ectodermal cells at 24h (Fig. 1B,C). However, unlike the veg1 endoderm, veg1 ectoderm cells do not turn on expression of *wnt1* and *wnt16*. *Sfrp* and *dkk* genes are not expressed in the veg1 lineage at pre-gastrular

stages after 15h. Thus veg1 ectodermal and veg1 endodermal cells have a similar potential to send as well as receive Wnt signaling.

Animal ectodermal cells include cells of various regulatory state domains of the animal half, which all give rise to ectodermal cell types, the stomodeal structures and neurons of the ciliated band (Yaguchi, Yaguchi et al. 2010, Angerer, Yaguchi et al. 2011). These cells express no *wnt* signaling gene and no *sfrp* or *dkk* genes after 15h. However, these cells express one *fzd* gene, *fzd1/2/7*, at first in all animal ectodermal cells, and by 21h exclusively in the oral ectoderm. Thus animal ectoderm cells are capable of responding to Wnt signaling but do not themselves emit Wnt signals.

Apical plate cells are the precursors of neurogenic cells at the animal pole. These cells transcribe no *wnt* genes, but do specifically express *fzd5/8* at all pre-gastrula stages. Furthermore, these cells express *sfrp1/5*, *dkk1*, and *dkk3* genes at all stages considered, and transiently also express *sfrp3/4*. Based on these expression patterns, cells of the apical plate likely neither respond to nor present Wnt signals.

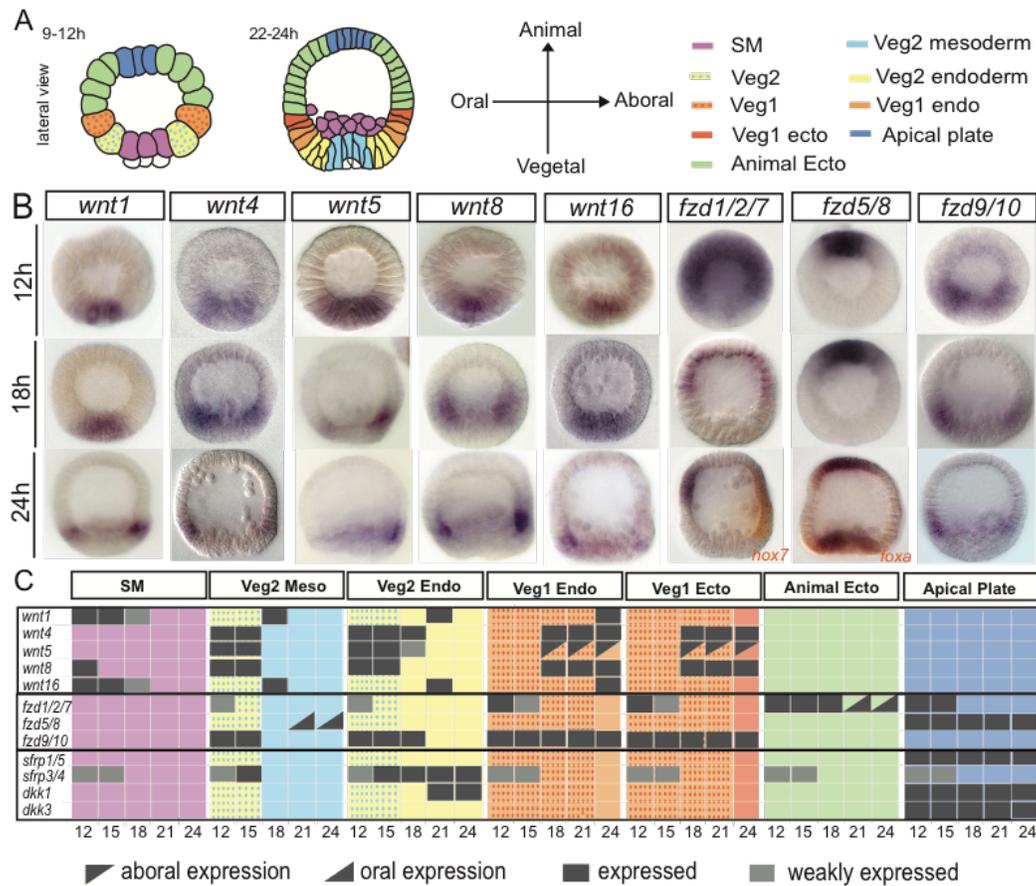


Figure 1. Spatial expression of Wnt signaling genes. (A) Schematic representation of early developmental stages of *S. purpuratus* embryos showing the spatial arrangement of regulatory state domains. SM, skeletogenic mesoderm; veg1, veg2, cell lineages descended from 6th cleavage ring of eight sister cells each giving rise to the parts of the embryo indicated in the diagrams; veg2 mesoderm, also known as non-skeletogenic mesoderm. Animal ectoderm and veg1 ectoderm denote both oral and aboral ectodermal domains. (B) Whole mount in situ hybridization (WMISH) of significantly expressed *wnt* and *frizzled* genes expressed at selected time points (12h, 18h, and 24h); additional time points for these genes and expression patterns of *dkk* and *sfrp* genes are shown in Fig. S2 and S3. (C) Expression matrix for each regulatory state domain, indicating expressed (black/grey) or not expressed (colored background) of the examined Wnt signaling genes every three hours, from 12h-24h. Regulatory domains are marked by the same color code as in (A). Developmental stages include 12h (early blastula), 15h (mid-blastula), 18h (hatching blastula), and 24h (mesenchyme blastula).

System-wide perturbation of Wnt signaling by a Porcupine inhibitor

To achieve a system-wide perturbation of Wnt signaling, we made use of a recently reported small chemical inhibitor of Porcupine, a membrane-bound O-acyltransferase required for acylation of Wnt proteins. Porcupine-mediated acylation of Wnts occurs at a conserved serine residue and is necessary for the secretion of those Wnt proteins that include this serine residue. Experimental perturbation of Porcupine interferes with secretion of all *Drosophila* Wnts except WntD (which lacks the target serine residue), and all mouse and human Wnts (Biechele, Cox et al. 2011, Herr and Basler 2012, Najdi, Proffitt et al. 2012, Biechele, Cockburn et al. 2013). Small chemical inhibitors of Porcupine have been recently proposed as efficient agents to interfere systemically with Wnt signaling in clinical applications, and here we used the Porcupine inhibitor C59 (Chen, Dodge et al. 2009, Lum and Clevers 2012, Proffitt, Madan et al. 2013). To test the efficacy of C59 in sea urchin embryos, we first assessed the phenotypes that develop in the presence of various concentrations of C59. Embryos treated with C59 showed no apparent defects in early development at pre-gastrular stages, and the ingression of skeletogenic cells occurs similarly as in control embryos. At late gastrula stage, defects in gastrulation, development of the tripartite gut, and formation of skeletogenic spicules were detected in a dose-dependent manner, while specification of mesodermal pigment cells was not affected (Fig.S4A). Embryos were exposed to C59 at different concentrations, and expression levels of the Tcf target genes *foxa*, *hox11/13b* and *eve* were strongly reduced at 0.5 μ M, while

expression of the mesodermal regulatory gene *gcm* remained unchanged (Fig. S4B). Similarly, when expression of Porcupine was blocked by morpholino injection, embryos showed defects in gut development as well as reduced expression levels of the endodermal regulatory genes *blimp1b*, *hox11/13b*, and *eve*, but not *gcm*. All these results confirm both the specificity and the efficacy of C59-mediated inhibition of Wnt signaling (Fig. S5A,B). Furthermore, all five expressed Wnt ligands contain the conserved serine residue required for Porcupine-mediated acylation, while surrounding amino acid sequences conform to a consensus sequence recently identified in Porcupine targets (Rios-Esteves, Haugen et al. 2014), indicating that all five Wnt ligands should indeed require Porcupine for their secretion (Fig. S5C).

The effect of inhibiting Wnt signaling by C59 treatment on regulatory gene expression in all embryonic domains was detected by Nanostring nCounter analysis using a probe set targeting 208 genes, including 172 genes encoding transcription factors. Sea urchin embryos were treated with C59 starting at 1h after fertilization, and gene expression levels were determined at 12h, 15h, 18h, 21h, and 24h. Results are shown for a few selected genes expressed in each of the seven embryonic domains in Fig. 2, and the complete data are listed in Table S1. A tabular summary of experimental evidence for Wnt signaling effects on all specific target genes addressed in this study can be found in Table S2. Of the 172 regulatory genes monitored in this experiment, 147 were expressed at least at one stage during the developmental time interval considered (Table S1). Treatment with C59 resulted in the down-regulation of 16 regulatory genes, of which 10 genes are components of

endoderm GRNs, and 5 are components of *veg1* ectoderm GRNs, while the expression domains of *gbx* at 24h have not been resolved yet. In addition, expression of 9 regulatory genes was up-regulated in embryos upon exposure to C59, eight of which are known to be expressed in the apical neurogenic domain, while the expression pattern of *acsc* is not known. The effects of C59 treatment on the activity of GRNs specific to the individual embryonic regulatory state domains can be summarized as follows.

Skeletogenic and veg2 mesodermal GRNs: Regulatory genes expressed in skeletogenic cells include *alx1*, *pmar1*, *dri*, *erg*, *ets1/2*, *tbr*, *tel*, and *tgif* (Oliveri, Tu et al. 2008). Expression of these genes is not changed in embryos exposed to C59, except *tgif*, which shows reduced expression levels at 24h, when it is expressed in *veg2* endoderm as well (Table S1). In *veg2* mesodermal cells, specifically expressed regulatory genes include *gcm*, *gatae*, *gatac*, *e2f3*, *erg*, *ese*, *ets1/2*, *hex*, *prox1*, *scl*, *shr2*, *six1/2*, and *z166* (Materna, Ransick et al. 2013). Of these, only *gatae* shows reduced expression levels upon C59 treatment, again at a time when it is in addition expressed in *veg2* endodermal cells. Expression of no other mesodermal regulatory genes is affected by C59 (Table S1), indicating that mesodermal cell fates do not require Wnt signaling inputs during pre-gastrular development. This is in agreement with the absence of *wnt* and *fzd* gene expression in most mesodermal cells after 18h (note that although oral *veg2* mesoderm cells later express *frz5/8* (Fig. 1B,C), expression of the canonical oral mesoderm regulatory genes *prox1*, *gatac*, *erg*, *ese*, and *scl* is impervious to C59).

Veg2 endodermal GRN: By 18h, the expression of all regulatory genes of the anterior

endoderm GRN that are exclusively transcribed in veg2 endoderm cells at this time, namely, *blimp1b*, *foxa*, *hox11/13b*, *brachyury* and *krl/z13* (Peter and Davidson 2010, Peter and Davidson 2011), is down-regulated in embryos treated with C59 (Fig. 2, Table S1). Two additional regulatory genes, *myc* and *soxc*, are expressed in veg2 endoderm cells at 18h, but their whole embryo expression levels are not affected by C59 treatment, possibly because these genes are also transcribed in cells of non-endodermal fates. Thus most, if not all regulatory genes specifically expressed in veg2 endodermal cells are down-regulated in C59 treated embryos. This result is consistent with the observed expression of *wnt* and *fzd* genes in these cells up to 18h. Furthermore, previous *cis*-regulatory evidence demonstrated that transcription of most genes of the early endoderm GRN is controlled by Tcf/ β -catenin (reviewed in ref. Peter and Davidson 2010). However, the initial expression of regulatory genes in endodermal cells prior to 15h is not affected by C59, and indeed, treating embryos with C59 only after 15h has a similar effect on regulatory gene expression at 24h as adding C59 at 1h after fertilization (Fig. S6A). For comparison, in embryos where maternal as well as zygotic accumulation of β -catenin is inhibited by injection of mRNA encoding dominant negative cadherin, expression of endodermal genes is affected at all times considered, starting at 9h (Fig. S7). These results indicate that secreted Wnt ligands are not required for regulatory gene expression in endodermal precursor cells prior to 15h, due to the presence of maternal β -catenin. A similar observation was made in mouse

embryos, where earliest developmental processes do not require Porcupine-dependent Wnt secretion (Biechele, Cockburn et al. 2013).

Veg1 endodermal GRN: By 15h, the regulatory gene *eve* is expressed throughout the *veg1* lineage, and by 21-24h, its product is responsible for activating the expression of regulatory genes specific to the future posterior endoderm, including *hox11/13b*, *brachyury*, and *hnf1* (Peter and Davidson 2011). The expression of all four genes is down-regulated in embryos treated with C59. Thus, the *veg1* endoderm GRN is being activated right at the time (24h) when all *wnt* genes and *fzd9/10*, but no *dkk* or *sfrp* genes are expressed in the same cells, and its operation depends on Wnt signaling.

Veg1 ectoderm GRNs: Regulatory gene expression in these vegetal ectodermal cells shows varied effects upon treatment with C59 (Fig. 2, Table S1). Thus, expression of several regulatory genes is affected, as indicated by lower levels of *nkl* and *unc4.1* transcripts (Li, Materna et al. 2012). Other transcripts, such as *msx* are present at decreased levels only at a stage when their expression is restricted to *veg1* ectodermal cells, while later, when these transcripts are also expressed widely in animal aboral ectoderm cells, transcript levels are comparable to control embryos (Fig. S8) (Li, Cui et al. 2014). These results suggest that *veg1* ectoderm GRNs are at least partially affected by Wnt signaling. However, the majority of regulatory genes are not exclusive to *veg1* ectoderm cells, and it is not possible to assess the extent of this regulatory input by quantitative measurements of gene expression levels. The observed effect of C59 on gene expression in *veg1* ectodermal cells is consistent with the presence of Wnt ligands and Fzd receptors in

these cells, and with previous reports of Wnt signaling affecting gene expression in *veg1* ectoderm (McIntyre, Seay et al. 2013).

Animal ectoderm GRNs: Regulatory genes expressed in animal ectoderm cells are mostly not affected by C59 perturbation, e.g., *foxg*, *not*, and *gsc* (oral animal ectoderm; ref. Li, Materna et al. 2013), *emx* (lateral animal ectoderm; ref. Li, Cui et al. 2014) and *hmx*, *hox7*, *dlx*, and *ets4* (aboral animal ectoderm; ref. Ben-Tabou de-Leon, Su et al. 2013). However, the transcription of *sp5*, which is expressed in *veg1* ectoderm and in the oral animal ectoderm domain (Fig. S8), is strongly down-regulated at 18h and up to the onset of gastrulation by C59 treatment (Fig. 2). Thus, specification of most animal ectoderm cell fates does not depend on Wnt signaling inputs, but the expression of one particular transcription factor appears to be regulated by Wnt signaling in these cells.

Apical ectoderm GRNs: Expression levels of regulatory genes transcribed in cells of the neurogenic apical ectoderm, such as *foxq2* (Yaguchi, Yaguchi et al. 2008), *foxj1* (Tu, Brown et al. 2006), *fez* (Yaguchi, Yaguchi et al. 2011), *zic* (Materna, Howard-Ashby et al. 2006), *hbn*, and *nk2.1* (Howard-Ashby, Materna et al. 2006) are up-regulated in embryos treated with C59 (Fig. 2, Table S1). Earlier observations showed that absence of Wnt signaling leads to an increase in expression levels of regulatory genes associated with neurogenic fate in adjacent animal ectoderm cells (Yaguchi, Yaguchi et al. 2008, Range, Angerer et al. 2013). In embryos exposed to C59 at 1h after fertilization, the earliest effect on regulatory gene expression, on *foxq2* and *nk2.1*, is observed at 15h. Adding C59 at 12h showed effects on apical gene expression similar to adding this drug at 1h, while

C59 treatment from 15h on showed much weaker effects, and no effects were observed when C59 was added at 18h (Fig. S6B). These results indicate that the Wnt signal critical for suppression of neurogenic fate in animal ectoderm cells is secreted after 12h and before 18h. As shown above, the neurogenic apical domain expresses no Wnt signaling ligand, but expresses *sfrp1/5*, *sfrp3/4*, *dkk1*, and *dkk3*, which encode potential Wnt signaling inhibitors. These observations thus indicate that the neurogenic fate is suppressed by Wnt signaling, and in turn, the apical neurogenic GRN ensures the expression of Wnt signaling antagonists.

Taken together, these results indicate that the anterior (*veg2*) and posterior (*veg1*) endodermal GRNs are broadly activated by Wnt signaling in pre-gastrular embryos, and that in addition, some regulatory genes of the *veg1* ectoderm GRNs receive positive Wnt signaling inputs. All three domains express only one Fzd receptor gene, *fzd9/10*, but express several genes encoding Wnt ligands, each of which or all together could be responsible for the observed effects on regulatory gene expression. In addition, Wnt signaling appears to antagonize apical neurogenic fate, but the *wnt* genes expressed closest to the animal ectoderm are *wnt4*, *wnt5*, and *wnt8* in *veg1* ectodermal cells. To determine if these different Wnt ligands execute similar or overlapping functions and can substitute for one another, or whether they operate in entirely distinct ways, requires the perturbation of individual Wnt ligand gene expression.

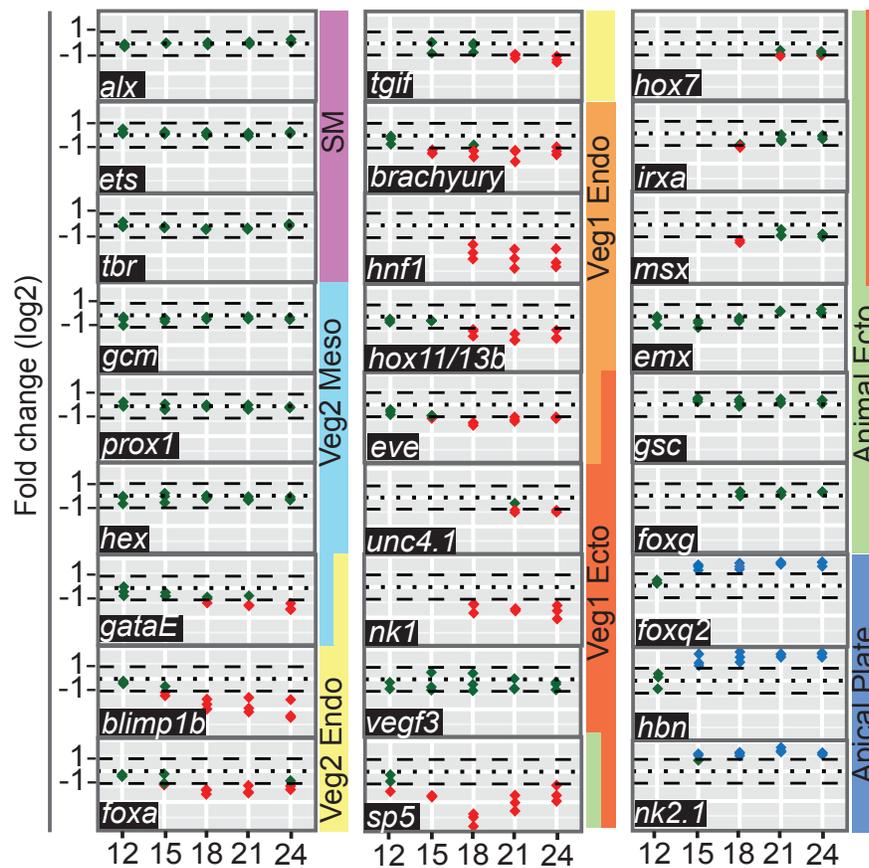


Figure 2. Effects of inhibiting Porcupine-dependent Wnt ligand secretion by C59 treatment. Embryos were treated with C59 or DMSO (control) at 1h post fertilization. Transcripts of each gene were measured by Nanostring nCounter analysis using a probeset detecting expression of 172 regulatory genes at five successive times (abscissa). Shown are fold differences of transcript abundance in embryos treated with C59 compared to control embryos. Each diamond represents one of three experimental repeats, shown in red if down-regulated >2-fold upon treatment with C59, in blue if up-regulated >2-fold, or in green if unchanged. A ratio of 1 (experimental/control) is indicated by the dotted line, and 2-fold envelope of significance is indicated by the dashed lines; ordinates, this ratio as \log_2 . Genes expressed at low levels (<100 transcripts/embryo) in treated and/or control embryos are not shown. A representative set of genes is shown for each regulatory state domain, with colored bar on the right of each column indicating the spatial expression domain(s) at 24h. Color code as in Fig. 1. Complete results are shown in Supplemental Table1.

Wnt ligands execute distinct functions

To distinguish the functional contribution of each Wnt ligand, we perturbed individually the expression of the five Wnts which could be responsible for the effects observed in embryos treated with C59. Embryos were injected with morpholino antisense oligonucleotides (MASOs) blocking the translation of Wnt1, Wnt4, Wnt5, Wnt8, or Wnt16, and gene expression levels in these embryos as compared to control embryos were analyzed at 24h for a set of regulatory genes which represent each of the four embryonic expression domains affected by C59 treatment.

The results summarized in Fig. 3 demonstrate that each Wnt ligand affects the expression of a specific set of regulatory genes. Thus, injection of Wnt1MASO broadly affected the expression of all endodermal regulatory genes tested, and also decreased the expression levels of some regulatory genes expressed in the veg1 ectoderm (*nkl*, *sp5*, and *unc4.1*), while other examined veg1 ectodermal genes and neurogenic apical genes are not affected. Wnt1 MASO also weakly affects the expression of all other *wnt* genes (Fig. 3). Embryos treated with Wnt4 MASO exhibit marginally decreased expression levels of *unc4.1*, *hox7*, and *msx*, in veg1 ectodermal cells, but have no effect on genes expressed either in the endoderm or the apical neurogenic plate. Injection of Wnt5 MASO did not reveal any change in regulatory gene expression, except for an up-regulation of *wnt5* transcripts. In Wnt8 MASO injected embryos, transcript levels of all neurogenic apical genes are increased, while genes expressed in endoderm and veg1 ectoderm are not affected. Wnt16 MASO selectively decreases the transcript levels of *blimp1b*, *eve*, and

hox11/13b in the *veg2* and *veg1* endoderm domains, but does not affect genes specific to the neurogenic apical plate and *veg1* ectodermal cells.

Taken together, these results indicate that four of the five Wnt ligands indeed execute specific functions in the regulation of transcription factor gene expression, which can not be substituted for by other Wnt ligands, despite the largely overlapping expression of all five *wnt* genes. When added up, the sum effects of perturbing the expression of individual Wnts largely correspond to the effects observed in C59 treated embryos. Thus, the expression of every gene tested here which was affected by C59 treatment was also affected by the perturbation of at least one Wnt ligand. Similarly, each embryonic regulatory state domain in which C59 treatment was observed to have altered gene expression was also affected by at least one of the Wnt ligands expressed at these developmental stages, as summarized in the following:

Veg2 endoderm: By 18h, the expression of *blimp1b*, *foxa*, *hox11/13b*, and *brachyury* in this domain is affected by C59 treatment. The responsible Wnts could be Wnt4 and Wnt5, which are expressed in these cells up to 18h; Wnt8, which is produced in these cells from 12 to 15h; or Wnt1 and Wnt16, which are expressed in the adjacent *veg2* mesodermal cells at 18h and in *veg2* endodermal cells at 21h. Neither Wnt4, Wnt5, nor Wnt8 morpholinos affect gene expression in *veg2* endoderm cells. However, Wnt1 and Wnt16 signaling are clearly required to activate gene expression in these cells by short range signaling. Injection of Wnt1 MASO results in lower expression of all tested regulatory genes expressed in *veg2* endoderm cells at 24h, *blimp1b*, *foxa*, and *gatae*, while the expression of *blimp1b* is

also affected by Wnt16 MASO, indicating that this gene requires both Wnt signaling inputs for normal expression.

Veg1 endoderm: cells of the future posterior endoderm express *fzd9/10* between 12-24h, *wnt4*, *wnt5*, and *wnt8* from 18-24h, and *wnt1* and *wnt16* at 24h. In addition, Wnt signaling may occur from the adjacent *veg2* endoderm cells, which transcribe *wnt* ligand genes between 12-21h, as discussed just above. In C59 treated embryos, expression of *eve* in *veg1* endoderm precursor cells is reduced after 15h. Similar to the *veg2* endoderm GRN, gene expression in *veg1* endoderm cells is not affected by morpholinos targeting Wnt4, Wnt5 or Wnt8 expression, but depends on Wnt1 and Wnt16 signaling. Again, injection of Wnt1MASO shows the broadest effect, reducing expression levels of *brachyury*, *hox11/13b*, *hnf1*, and *eve*, while injection of Wnt16 MASO only decreased expression levels of *hox11/13b* and *eve*. Interestingly, Wnt1 and Wnt16 not only execute overlapping functions in the two endodermal GRNs, but are also expressed in very similar patterns which are consistent with these functions. At 18h, when the expression of *foxa* and *blimp1b* is restricted to *veg2* endoderm, *wnt1* and *wnt16* expression is detected exclusively in the adjacent *veg2* mesoderm. By 21h however, transcripts of *wnt1* and *wnt16* are detected in anterior (*veg2*) endoderm cells for a brief period of time, which is when expression of their target genes *hox11/13b*, *brachyury*, and *hnf1* is induced in the adjacent *veg1* posterior endoderm cells. Thus when expressed in anterior endoderm, Wnt1 and Wnt16 again function as short range signaling ligands, now activating expression of posterior endoderm regulatory genes in adjacent *veg1* cells.

Veg1 ectoderm: Expression of regulatory genes such as *nk1* and *unc4.1* in *veg1* ectodermal cells is weakly affected by C59 treatment from 18h on. *Wnt4*, *wnt5* and *wnt8* are expressed in the *veg1* lineage from 18h on, and *fzd9/10* is expressed in these cells at all stages considered. In addition, by 24h, *wnt1* and *wnt16* are being expressed in the neighboring *veg1* endoderm domain. Injection of embryos with Wnt-specific morpholinos revealed that only Wnt1 and Wnt4 affect gene expression in *veg1* ectoderm cells, while morpholinos against Wnt5, Wnt8, and Wnt16 showed no effect by 24h (Fig. 3). Wnt1 and Wnt4 signaling in the *veg1* ectoderm domain therefore occurs by short range plus perhaps intra-domain signaling. Embryos injected with Wnt1 MASO showed reduced expression levels of *nk1*, *unc4.1*, and *sp5*, while Wnt4MASO affected expression of *hox7* and *msx* genes. Thus as in the endodermal GRNs, two Wnt ligands regulate gene expression in the *veg1* ectoderm, but in this domain, they affect separate sets of regulatory genes. These results differ from those obtained in another sea urchin species, *Lytechinus variegatus*, where Wnt5 was shown to activate regulatory genes in the *veg1* ectoderm domain, including *irxa* and *nk1* (McIntyre, Seay et al. 2013). However, in *S. purpuratus*, when Wnt5 expression was knocked down using two separate morpholinos, no change in ectodermal gene expression levels was observed except an increase in *wnt5* transcripts. This result implies that Wnt5 signaling is not required in this domain, though it might function redundantly together with other Wnts. In the *S. purpuratus* embryo, Wnt4 regulates the expression of *nk1*, while *irxa* expression was not affected by perturbation of any individual Wnt signal, even though its expression is moderately affected by C59

treatment at 18h. This inter-species difference could thus reflect a relatively recent change in the identity of the Wnt ligand used for activation of regulatory genes in the veg1 ectoderm GRNs.

Apical ectoderm: The earliest increase in expression of regulatory genes specific to the apical neurogenic fate upon C59 treatment was observed at 15h. Since no Wnt ligand is expressed in the animal half, the Wnt signal responsible for this effect must derive from the vegetal half of the embryo. The vegetal cells closest to the animal ectoderm, the veg1 lineage cells, express *wnt4*, *wnt5*, and *wnt8* from 18h on, and prior to that, at 12 to 15h, these same genes are expressed in the veg2 lineage. Injection of Wnt8 MASO at fertilization results in elevated expression levels of all apical regulatory genes tested. This result is consistent with a previous report on the function of Wnt8 in the restriction of expression of genes associated with the neurogenic apical fate (Range, Angerer et al. 2013). Our results show that blocking the expression of Wnt1, Wnt4, Wnt5, or Wnt16 had no effect on expression of apical-specific genes. *Wnt8* is the earliest and most abundantly transcribed *wnt* gene, and at 12-15h, when the embryo consists of only a little over a hundred cells, there are still only few cells separating Wnt8 expressing vegetal cells and the cells in which expression of neurogenic apical regulatory genes is to be prevented. Thus in summary these experiments revealed that the Wnt signaling function responsible for limiting the apical neurogenic domain is executed solely by Wnt8, and that this is a very early process, probably operating in late cleavage.

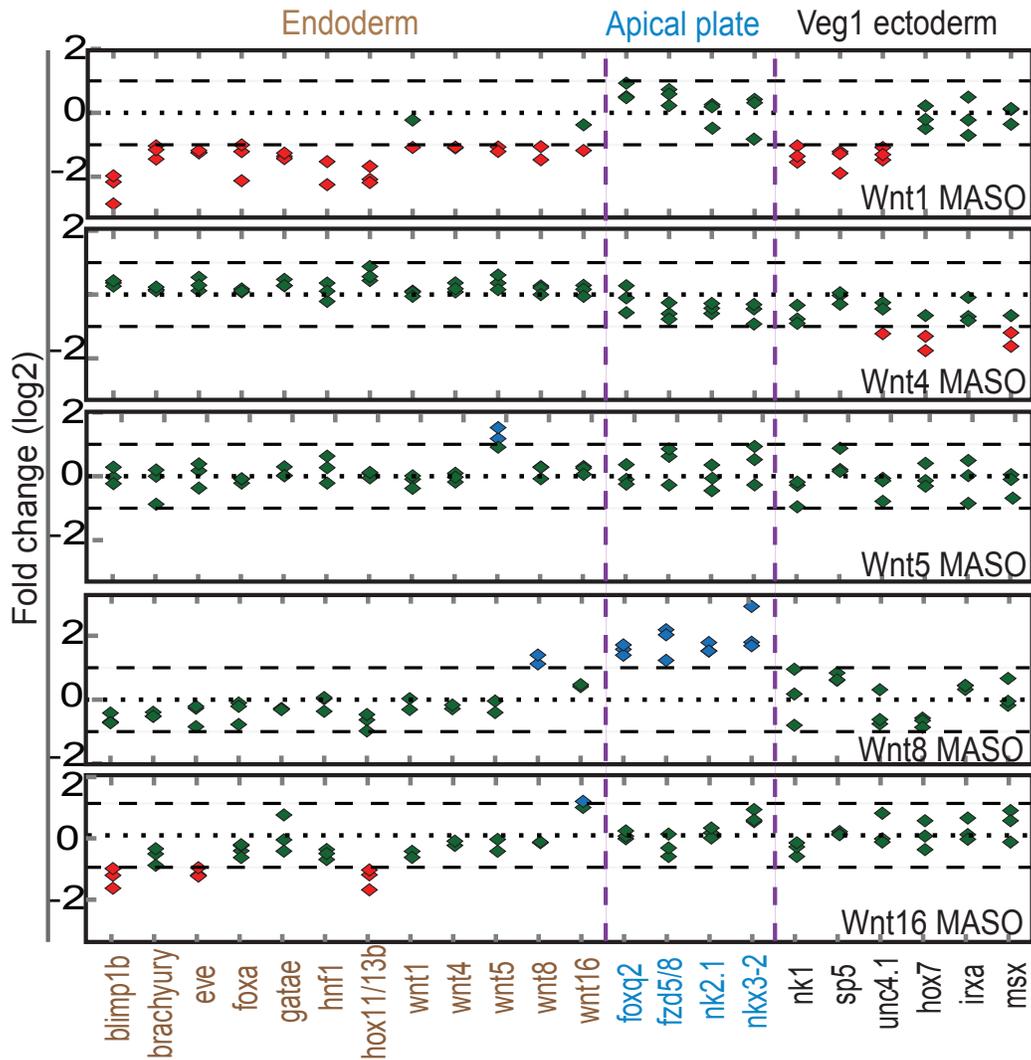


Figure 3. Morpholino perturbation of individual Wnt ligands and effects on regulatory gene expression. Embryos were injected with morpholinos targeting the expression of Wnt1, Wnt4, Wnt5, Wnt8, or Wnt16, or with randomized control morpholinos. Expression levels of regulatory genes in morpholino injected embryos were analyzed by QPCR at 24h, except for three genes analyzed at earlier developmental stages: *hox7* (asterisk, 21h), *irxa*, and *msx* (diamonds; 18h). The genes selected for this analysis are those demonstrated to respond to C59 treatment (see above). Shown are ratios (log2) of expression levels in Wnt-specific morpholino injected embryos to expression levels in control morpholino injected embryos. Symbolism as in Fig. 2.

Wnt-dependent spatial patterning functions

To assess the consequences Wnt signaling on embryonic patterning of regional regulatory states, we studied the spatial disposition of gene expression domains in embryos in which Wnt signaling is inhibited. As an initial assessment, embryos treated with C59 from 1h to 24h, were fixed at 24h and stained by WMISH with eight probes against regulatory genes specific for different expression domains throughout the embryo (Fig. S9A). Compared to control embryos, the expression domains of *alx1*, the skeletogenic mesoderm cells, and of *delta*, the *veg2* mesodermal cells, are not affected by C59. However, expression of *foxa*, marking *veg2* endoderm cells, is restricted to fewer cells in C59 treated embryos compared to control embryos. The boundary usually separating *veg2* endoderm (*foxa*) and *veg1* endoderm (*hox11/13b*) fate appears to have shifted to the vegetal pole, and expression of *hox11/13b* is at least partially overlapping with that of *foxa* in embryos treated with C59. Similarly, most gene expression boundaries are shifted towards the vegetal pole, including the boundaries of the *foxq2* expression domain at the animal pole.

To determine which Wnt signal is responsible for the vegetal shift of most expression domains except the mesodermal domains, we tested the spatial expression of these genes in embryos injected with Wnt morpholinos. The results, as summarized in Fig. S10 (for data see Fig. S9B), indicate that interfering with the expression of Wnt1 not only decreases the expression levels of *foxa*, *hox11/13b*, and *eve*, but that these genes are also expressed in fewer cells, and that these are located closer to the vegetal pole than in control embryos. The vegetal boundary of

gene expression of *lim1* (veg1 ectoderm) and particularly that of *emx* (animal ectoderm) are also shifted towards the vegetal pole in Wnt1MASO injected embryos, consistent with the previous observation that *emx* expression in veg1 cells is repressed downstream of Eve (Li, Cui et al. 2014). However, Wnt1 MASO does not affect the boundary between *emx* and *foxq2* expression domains. Effects with Wnt16MASO are similar but weaker. Embryos injected with morpholinos blocking either Wnt4 or Wnt5 expression do not appear to affect the patterning of gene expression domains assessed here. In embryos injected with Wnt8MASO, on the other hand, only the boundary between *emx* and *foxq2* expression domains, which were shown to repress each other (Li, Cui et al. 2014), is shifted vegetally, leading to an expansion of the apical neurogenic domain confirming that the up-regulation of *foxq2* expression levels in Wnt8MASO injected embryos is caused by the de-repression of apical regulatory genes within animal ectoderm cells. This leads to the prediction that Wn8 activates the expression of an early acting repressor of *foxq2* in the animal ectoderm. The observation that Wnt8 signaling is required for cell fate decisions in the animal ectoderm is consistent with the expression of *fzd1/2/7* in all cells of this domain between 12-18h, the time when Wnt signaling input is required to repress apical cell fates (Fig. S6B), and by the absence of potential Wnt signaling antagonists which could interfere with Wnt8 signaling. The boundaries between gene expression domains within the vegetal half of the embryo are not affected by Wnt8 signaling. These results indicate that gene expression patterning in the vegetal and animal half of the sea urchin embryo occurs independently.

Control of *wnt* gene expression by cell fate specification GRNs

Even though the early expression of *wnt* genes occurs in dynamic spatial patterns, by 24h, all five *wnt* genes are expressed in the veg1 endoderm, and some of them in addition also in the veg1 ectoderm (Fig. 1C). This raises the question of how the transcription of these genes is regulated by the cell fate specific GRNs operating in these cells. Moreover, we would like to know if the regulatory mechanisms that control their expression are similar, given the overlapping expression patterns of *wnt4*, *wnt5*, and *wnt8*, as well as of *wnt1* and *wnt16* throughout early sea urchin development (Fig. 1C; for a direct comparison of the expression patterns of *wnt* genes and relevant regulatory genes, see Fig. S11). The earliest specification of the veg1 cell lineage, the precursor of veg1 endoderm and veg1 ectoderm, is controlled by *eve*, a regulatory gene exclusively expressed in veg1 cells by 15h (Peter and Davidson 2010). Eve contributes to the activation of *hox11/13b* in veg1 endoderm, which is the earliest regulatory gene operating in the GRN underlying the specification of posterior endoderm fate (Peter and Davidson 2011). Perturbation of the pan-veg1 transcription factor Eve by injection of Eve MASO resulted in decreased expression of *wnt1*, *wnt4*, *wnt5*, and *wnt16*, consistent with their expression in veg1 cells, although expression of *wnt8* in these same cells was not affected (Fig. S12). Blocking the expression of the endoderm-specific Hox11/13b transcription factor with morpholinos, however, affects only the transcription of *wnt1* and *wnt16*, but of none of the other *wnt* genes, as might be expected from the exclusive expression of *wnt1* and *wnt16* in veg1 endodermal cells at 24h (Fig. S12). Blocking the expression of other endodermal regulatory

factors, FoxA, Blimp1b, Brachyury, or GataE, did not affect the expression of the five *wnt* genes (Fig. S13). Thus expression of *wnt1* and *wnt16* is regulated probably directly by both Hox11/13b and Eve, and the transcription of *wnt4* and *wnt5* is activated downstream of Eve, but independent of Hox11/13b. Late *Wnt8* expression is controlled by a separate mechanism, which was recently shown to involve the pan-ectodermal regulator SoxB1 (Li, Cui et al. 2014).

Thus in all three cases, that is [*wnt1* + *wnt 16*], [*wnt4* + *wnt5*], and [*wnt8*], the *wnt* genes are wired into the GRNs they control. The GRN circuitry, which is summarized in Fig. 4, explains the spatial and temporal expression pattern of *wnt* genes, and indicates a remarkable relationship between the transcriptional control of *wnt* genes and their downstream functions. For instance, *eve* expression is specific to all cells of the veg1 lineage from 15h on, and expression of *wnt4* and *wnt5* in these same cells, under the control of Eve, is first observed shortly after that, at 18h, and continues up to 24h. In turn, Wnt4 activates gene expression in veg1 ectodermal cells, where Eve continues to be expressed. That is, Eve plays a dual role in the specification of veg1 ectoderm: it directly represses genes of the animal ectoderm in these cells and it activates the expression of *wnt4* in veg1 endoderm and veg1 ectoderm, which leads to the activation of genes of the veg1 ectodermal GRN.

In veg1 endoderm cells, expression of *hox11/13b* initiates at 21-24h, under the control of Eve, and transcription of *wnt1* and *wnt16*, activated by Hox11/13b, starts at 24h. Their expression is not observed in veg1 ectoderm, where Hox11/13b is not expressed, though *eve* is. Since signaling by Wnt1 and Wnt16 also affects the

expression of *hox11/13b* and *eve*, these results indicate that expression of Hox11/13b, a crucial upstream transcription factor in the posterior endoderm GRN, is controlled by a positive feedback circuit between *wnt1*, *wnt16*, and *hox11/13b*, thus ensuring specification of the posterior endoderm cell fate (Fig. 4). Curiously, this circuit might also be responsible for the expression of *wnt1*, *wnt16* and *hox11/13b* at earlier time points. By the time when *hox11/13b* expression becomes sensitive to Wnt signaling according to the C59 experiments, at 18h, *hox11/13b* is being transcribed specifically in veg2 endoderm, while *wnt1* and *wnt16* are transcribed in adjacent veg2 mesoderm. By 21h, *wnt1* and *wnt16* are expressed in veg2 endoderm, controlled by Hox11/13b, and transcription of *hox11/13b* is activated in adjacent veg1 endoderm cells. By 24h, expression of all three genes has terminated in veg2 endoderm, by a mechanism involving auto-repression of Hox11/13b (Peter and Davidson 2011), and the entire positive feedback circuit is operative exclusively in veg1 endoderm cells. Thus the sequence begins as an inductive relay, where first veg2 mesoderm Wnt1 and Wnt16 activate the *hox11/13b* gene in veg2 endoderm, and then veg2 endoderm Hox11/13b activates the *wnt1* and *wnt16* genes in veg2 endoderm, whereupon Wnt1 and Wnt16 activate the *hox11/13b* gene in Veg1 endoderm. In veg1 endoderm their mutual positive feedback locks in the circuit. The existence of a signaling interaction between the veg2 endoderm GRN and the veg1 endoderm GRN, which is responsible for activating *hox11/13b* transcription in veg1 endoderm cells after 21h, was predicted earlier (Peter and Davidson 2011). Here we identify this signal to be both *wnt1* and *wnt16*, which are expressed in veg2 endoderm cells at 21h, under the control of

Hox11/13b, and are required for *hox11/13b* transcription in the adjacent veg1 endoderm domain, thus fulfilling all criteria for the predicted signal.

Finally, the expression of *wnt8* in the veg1 endoderm and veg1 ectoderm domains is regulated differently from that of *wnt4* and *wnt5*, even though these show the same expression pattern. By 24h, *wnt8* expression is activated by SoxB1, a transcription factor present throughout the animal half of the embryo as well as in the veg1 lineage, which functions as an activator of many ectodermal regulatory genes (Kenny, Oleksyn et al. 2003, Li, Cui et al. 2014). SoxB1 thus activates the specification GRNs in animal ectoderm cells, and controls the expression of the signaling ligand which is required to exclude an alternative fate in these cells.

Thus in summary, even though all five *wnt* genes are expressed in veg1 endoderm, only those, *wnt1* and *wnt16*, affecting the activity of the endoderm GRN are also controlled by the endoderm GRN regulator Hox11/13b (see network in Fig. 4). Conversely, those *wnt* genes, *wnt4* and perhaps *wnt5*, which affect the activity of the veg1 ectoderm GRN are independent of Hox11/13b expression and are controlled instead by the veg1 regulator Eve. And *wnt8*, which causes expression of a repressor of apical neurogenic genes in the animal ectoderm, is activated by the animal ectoderm factor SoxB1.

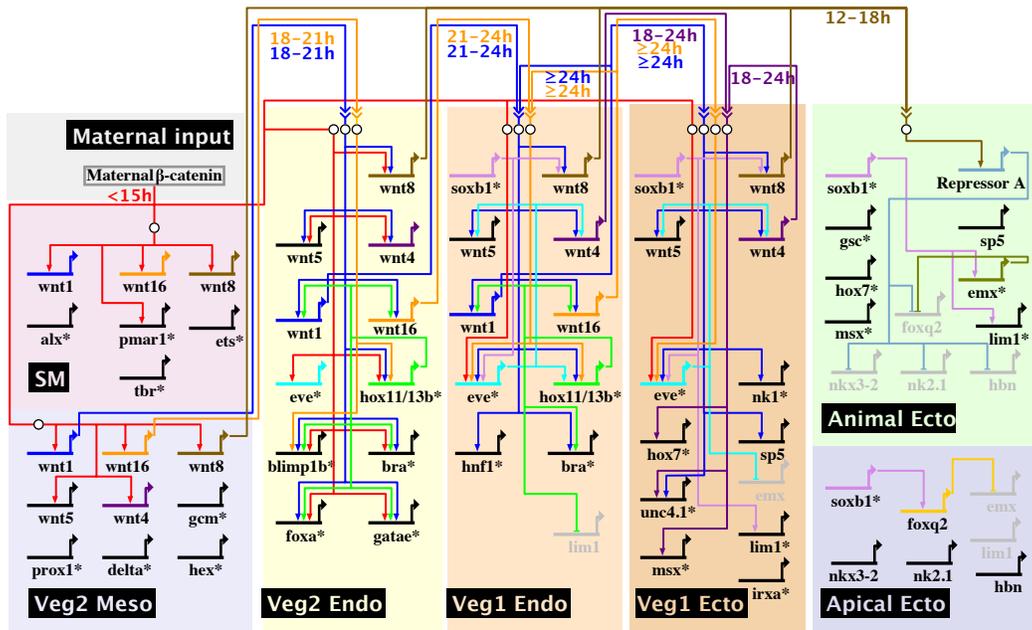


Figure 4. Model of regulatory interactions between *wnt* genes and domain-specific GRNs. This diagram summarizes the regulatory interactions between *wnt* genes and the regulatory genes composing the cell fate specification GRNs operating in embryonic regulatory state domains, as well as the regulatory inputs from maternal β -catenin. The GRNs shown omit known input and output linkages at most regulatory genes, as symbolized by asterisks at the gene name. For the complete GRNs see <http://sugp.caltech.edu/endomes/>. The time windows during which Wnt signaling inputs are active are indicated below regulatory linkages.

DICUSSION

We show here the results of a system-wide analysis of the functions of Wnt signaling in the regulation of the cell fate specification GRNs operating across the entire sea urchin embryo during pre-gastrular development. Our approach was to analyze in every regulatory state domain of this embryo the spatial expression of all

genes involved in intercellular Wnt signaling, encoding Wnt ligands, Frizzled receptors, and possible antagonists such as Sfrp and Dkk proteins. We addressed the overall function of Wnt signaling in developmental GRNs during the first 24h of embryogenesis, by use of the C59 Porcupine inhibitor, which interferes globally with the secretion of Wnt signaling ligands. This perturbation affects the expression of specific regulatory genes that in the GRNs lie downstream of Wnt signal inputs. The contribution of each Wnt signaling ligand to the regulation of expression of the genes affected by C59 treatment was further analyzed by individually blocking the expression of each Wnt ligand using injection of antisense morpholinos. Our results lead to overall conclusions regarding the functions of the Wnt signaling system in an entire embryo and how these functions are mediated by the linkages between the Wnt signaling system and the GRNs controlling the developmental process. This study traverses a large scale developmental process during which diverse regulatory state domains are established throughout the embryo.

Specific functions of Wnt signaling in the embryonic GRNs

Only five of the 11 Wnt ligands encoded in the *S. purpuratus* genome are expressed during pre-gastrular development. Their spatial expression, even though dynamically changing, is at all times confined to at most two adjacent gene expression domains. The remainder of the embryo expresses no *wnt* genes. The spatial expression of Frizzled receptors occurs in broad contiguous domains. After 18h, all cells express no more than one Frizzled receptor, but most cells of mesodermal fate express none. The system-level analysis of Wnt ligand and

receptor expression patterns yields revealing insights into the functions of Wnt signaling in the whole process of early sea urchin development. We see for example that the Wnt signaling system can have nothing to do with specification of most mesodermal fates, because except for the oral mesoderm after 21h, no mesodermal cells express Wnt signaling receptors. This prediction is confirmed in the C59 experiment, where no effects on mesodermal gene expression were recorded in the face of global knock down of zygotic Wnt signaling. Whereas Wnt signaling plays important roles in endoderm specification, as shown by the C59 and Wnt morpholino experiments, the expression patterns indicate that after about 21h, this role must be restricted to the posterior endoderm, since transcripts of wnt ligands and receptors cease to be detectable in the anterior endoderm by this time, while expression of the potential antagonists *sfrp3/4* and *dkk1* is activated. On the other hand, these expression patterns imply the importance of Wnt signaling in the posterior endoderm, where all five *wnt* ligands are expressed at 24h, and *fzd9/10* continues to be expressed at all times considered. The same arguments apply also to the *veg1* ectoderm, where expression of Wnt ligands and receptors but of no antagonists are consistent with the observed Wnt signaling function. In the animal ectoderm, presence of a receptor marks the potential to receive Wnt signals, though the absence of Wnt ligand expression precludes these domains as sources of Wnt signaling. Indeed, these cells receive Wnt8 signaling inputs from more vegetally located cells, which lead to exclusion of apical neurogenic fate, but Wnt signaling is not required for activation of ectodermal GRNs. Finally, the apical neurogenic domain does not express Wnt ligands, but does express a Wnt receptor and also

several antagonists which protect it from Wnt8 signaling. The expression patterns when viewed in detail at a system level thus confer a remarkably accurate set of predictions that illuminate which domains of the embryo engage in Wnt signaling.

In the sea urchin embryo the key important function of Wnt signaling in this developmental phase consists of their impact on the endodermal GRNs, since the expression of a majority of regulatory genes in these networks is activated by Wnt signaling (Peter and Davidson 2010, Peter and Davidson 2011). However, contrary to previous assumption, the earliest Tcf-dependent expression of endodermal regulatory genes and also the initiation of the skeletogenic GRN, do not depend on Wnt ligand signaling, but are mediated by the presence of maternal β -catenin which *ab initio* is specifically localized in cells of the skeletogenic and *veg2* lineages (Logan, Miller et al. 1999). Thus despite the early expression of *wnt* genes starting at 7h in this embryo, and the presence of maternal *Fzd1/2/7*, Wnt signaling is not required for early gene expression, since the earliest effects upon treatment with the Porcupine inhibitor C59 are observed only at 15-18h. This is a strong conclusion which depends on our and other's evidence that C59 is highly efficient in blocking an essential step in Wnt signaling. In this work a convincing argument is that the sum total of individual Wnt morpholino effects equals the effect of C59 treatment, as we show for every domain affected by this perturbation. Outside the endodermal GRNs, only a few regulatory genes expressed in *veg1* ectoderm cells require activation by Wnt signaling. In animal ectoderm cells, Wnt signaling is additionally required to repress activation of the apical neurogenic GRN, but this function is

probably also confined to activation of a yet unknown early repressor of apical gene transcription. Specification of all mesodermal cell fates as well as of all apical cell fates occurs independently of Wnt signaling in this embryo during pre-gastrular development. Thus a comprehensive view of Wnt signaling system function reveals only a small set of target genes and these belong to precisely confined GRNs of the embryo. Where the Wnt signaling system does impact the embryonic GRNs, it does so by use of characteristic circuitry that affects many nodes of the GRN, in which cross-regulatory interactions ensure continued expression of the active Wnt ligands.

Our results indicate that in this embryo and at the stages we consider here, Wnt signaling functions by short-range intercellular interactions. Thus each Wnt ligand affects gene expression either in cells of the same domain where it is also expressed, or in cells of an immediately adjacent domain. Examples of the latter, which constitute short range inductive signaling effects, are, as shown in Fig. 4, the response of veg2 endoderm to Wnt signals emitted by the veg2 mesoderm (Wnt1 and Wnt16); the response of veg1 endoderm to Wnt signals emitted by veg2 endoderm (Wnt1 and Wnt16); and the response of veg1 ectoderm to signals emitted by veg1 endoderm (Wnt1). In every case where Wnt signaling ligands affect an adjacent domain, this domain is located to the animal side of the domain of Wnt ligand expression. The mechanism underlying this vegetal to animal polarity is on the face of it enigmatic. It is clearly not due to a layered or oriented expression of diverse Frizzled receptors, as this is excluded by their spatially simple patterns of expression. A remarkable aspect of Wnt signaling revealed in this study is that it repeatedly functions in an intra-domainic manner, such that cells within a given

regulatory state domain both receive and emit the Wnt signal. This feature corresponds to the output of community effect circuitry, which in the cases examined relies on regulation of the signaling ligand gene by the signal transduction system which it activates in responding cells (Bolouri and Davidson 2010). The import of this circuitry is to ensure the homogeneity of regulatory states among cells of a given domain, by establishing a positive intercellular feedback throughout the domain. In this work we find intra-domainic community effect signaling specifically in the *veg1* endodermal domain, where the feedback consists of activation of *wnt1* and *wnt16* by *Hox11/13b*, while in turn, *Wnt1* and *Wnt16* activate *hox11/13b* (via *Tcf/β-catenin* input; Fig. 4). Similarly, *wnt4* is expressed within the *veg1* ectoderm and also activates gene expression within this domain.

Remarkably, the control of expression of all five relevant *wnt* genes is tightly correlated with their respective functions. Thus the *wnt* genes that operate in the endodermal GRNs are also transcribed under the control of endodermal regulators. Similarly, *Wnt8*, which operates on the animal ectoderm GRN, is transcriptionally controlled by an animal ectoderm regulator. The *wnt* genes which operate in the *veg1* GRNs are controlled by *Eve*, which regulates gene expression in *Veg1* ectoderm. Partly, this would follow de facto from the participation of these ligands in intra domainic signaling, but this cannot be the complete explanation, as some of the relevant signaling is inductive. What emerges is that particular regulatory cassettes link transcriptional control of Wnt signaling to the cell fate specification GRNs they regulate. The implication of this circuitry is that the

temporal co-expression of the Wnt signaling ligands and their target genes, as well as the spatial proximity of signal-sending and signal-receiving domains, is ensured by the common control of *wnt* genes and their target regulatory genes by the same cell fate-specific transcriptional regulators.

The specificity conundrum in the Wnt signaling system

An important question is how specificity of Wnt signaling is mediated, and this is a general issue given the multiplicity of *wnt* and *fzd* genes encoded in all animal genomes. While our study was not particularly designed to address this question, our results may hint at the basic design principle of this signaling system. Most important, in the context of GRN analysis, is the question of how Wnt signaling in different embryonic domains or different developmental phases leads to the regulation of specific, different sets of target genes. At the *cis*-regulatory level, the causal explanation for which genes are expressed in response to Wnt signaling must rely on combinatoriality in the regulation of gene transcription. Tcf, the transcription factor activated by Wnt signaling, always operates together with other transcription factors, some of which are exclusively expressed in cells of a given fate at a particular developmental time. We see here, as specific examples, that most regulatory genes activated in endodermal precursor cells by Wnt1 signaling are also activated by Hox11/13b, while Wnt1 signaling in *veg1* ectoderm leads to expression of *nk1*, which is also a target gene of Not and Lim1 transcription factors expressed in these cells. Thus the impact of Wnt signaling on gene expression in cells of different fates causally depends on the regulatory state in

the signal receiving cells. Note here that all gene expression domains regulated by Wnt1 signaling express the same Frizzled receptor, Fzd9/10, and we can therefore exclude the idea for the early sea urchin embryo that difference in target gene sets depends on utilization of diverse Fzd receptors.

An additional level of specificity must be invoked to account for the selective interaction of given Wnt ligands with given domains. For example, all five Wnt ligands are expressed in veg1 endoderm cells, but only Wnt1 and Wnt16 are required for GRN function in this domain. Similarly, Wnt4 affects veg1 ectoderm cells while Wnt16 and Wnt8 do not. In each of these domains, specific Wnt ligands are required for signaling response, while that domain is blind to the presence of the other Wnt ligands. Moreover, veg1 endoderm and veg1 ectoderm respond to different Wnt ligands, despite expression of the same receptor Fzd9/10. This points directly at a Wnt ligand recognition function in addition to the Fzd9/10 receptor present on these cells. An explanation could be the differential expression in the diverse embryonic domains of Wnt co-receptors such as the low-density lipoprotein receptor-related protein (LRP) family in these cells (for review see ref. van Amerongen and Nusse 2009).

METHODS

Gene cloning and Whole mount in situ hybridization (WMISH). cDNAs of *wnt*, *fzd*, *dkk*, and *sfrp* genes were amplified by PCR using primers listed in Table S2, and cloned into pGEM-T constructs. cDNA prepared from various developmental stages were used as templates for PCR reactions. Sequences used to prepare WMISH probes of regulatory genes, i.e. *alx*, *delta*, *foxa*, *hox11/13b*, *eve*, *lim1*, *emx*, and *foxq2*, were previously described (Peter and Davidson 2010, Peter and Davidson 2011, Li, Materna et al. 2012, Li, Materna et al. 2013, Li, Cui et al. 2014). Cloned constructs were linearized for transcription of antisense RNA probes labeled with digoxigenin (DIG) or fluorescein (FL). WMISHs were performed according to the standard methods (Ransick A and Davidson 2012). In brief, embryos were fixed in 2.5% glutaraldehyde, 32.5% sea water, 32.5 mM MOPS (pH 7) and 162.5 mM NaCl at 4°C overnight. 25 ng/μl of proteinase K in TBST was used to treat embryos at room temperature for 5-10 minutes, followed by 30-minute fixation in 4% paraformaldehyde at room temperature. Hybridizations were performed overnight at 58-60°C using a probe concentration of 0.5-1ng/μl. Probes were detected using anti-DIG or anti-FL Fab fragments conjugated to alkaline phosphatase (Roche). Color was developed using NBT/BCIP or INT/BCIP reagents (Roche).

C59 treatment. The Porcupine inhibitor C59 was obtained from Cellagen Technology (C7641-2s). C59 was dissolved in DMSO at 10mM, and experiments

were performed in the presence of C59 diluted in sea water to 0.5 μ M (see text), unless indicated otherwise. Except for the timing experiment in Fig. S6, embryos were treated with C59 at 1h post fertilization and were exposed until embryos were collected, at developmental times indicated in figure captions.

MASO injection and RNA isolation. MASOs were provided by GeneTools. Sequences of gene-specific MASOs are shown in Table S3. MASOs were diluted in injection solution including 0.12M KCl at 300 μ M and injected into fertilized eggs in a volume of 2-4 μ l. Randomized control MASOs (N₂₅) were injected at the same concentration. Experiments were performed on 2-5 independent embryonic batches. Embryos were cultured at 15°C, and approximately 300 MASO-injected embryos were collected at different time points. Total RNA was extracted using the RNeasy Micro Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

QPCR analysis. cDNA was synthesized using iScript cDNA synthesis kit (BioRad). QPCR was performed using Power SYBR Green Master Mix (Life Technology) with the primers listed in Table S2. Gene expression levels were normalized to those of poly-ubiquitin. Fold change was calculated by comparing normalized expression levels in the embryos injected with gene specific MASOs to expression levels in embryos injected with control MASOs.

NanoString nCounter. Approximately 300 embryos were collected at different time points and lysed in 15 μ l of RLT buffer (Qiagen, Hilden, Germany). Hybridization reactions were performed according to manufacturer's instructions in

total 30 μ l solution with 10 μ l of detection probes, 10 μ l of hybridization buffer, 5 μ l of embryo lysate, and 5 μ l of capture probes. All hybridization reactions were incubated at 65 °C for a minimum of 18 h. Hybridized probes were recovered with the NanoString Prep Station and immediately evaluated with the NanoString nCounter. The resulting count for each gene was subtracted by counts of negative spikes for background correction and then normalized using the sum of all corrected counts for all genes in the codeset. Fold changes of normalized counts in C59 treated versus untreated control embryos is shown for genes with counts >200, corresponding to approximately 100 transcripts. Experiments were performed on 3 independent embryonic batches. Probe sequences and accession numbers for the genes included in the codeset are as same as previously reported (Li, Materna et al. 2012)

ACKNOWLEDGMENTS. We are very grateful to Prof. Randall Moon (University of Washington) and Prof. Robert Burke (University of Victoria) for their careful and expert reviews of this manuscript. We would like to thank Prof. Mike Collins (UCLA) for helpful discussions and for bringing to our attention the small molecule antagonist C59. It is our pleasure to acknowledge support of this research by National Institutes of Health Grant HD-037105 and the Lucille P. Markey Charitable Trust.

SUPPLEMENTAL FIGURES AND TABLES

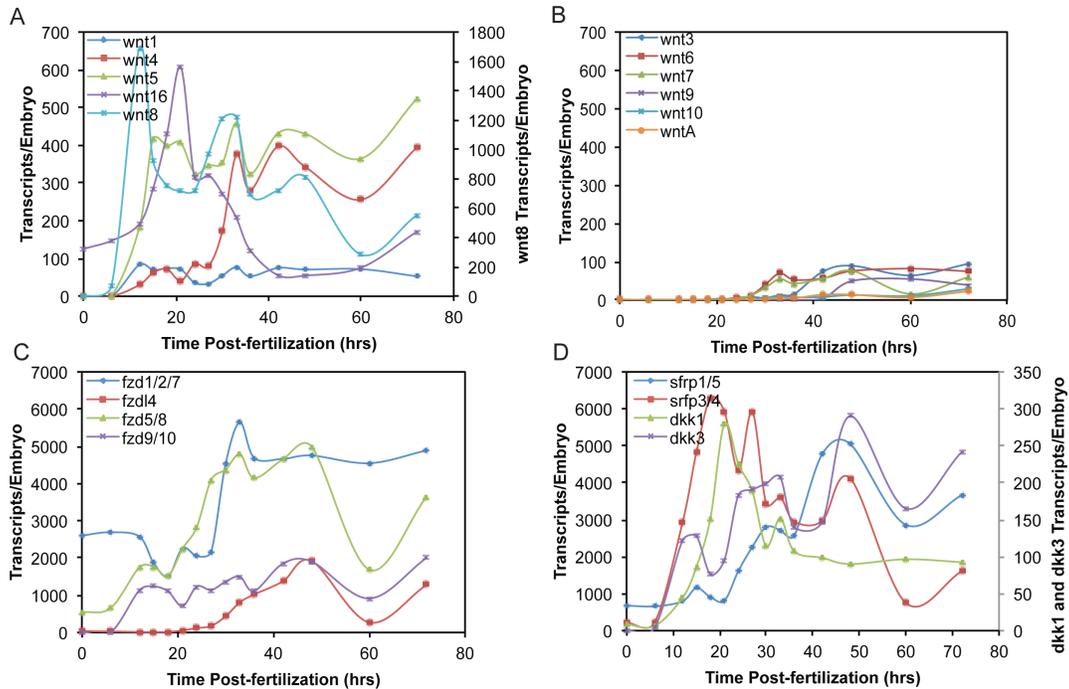


Figure S1. Temporal expression profiles of Wnt signaling components from 0 to 72h. (A) Early expressed *wnt* genes: *wnt1*, *wnt4*, *wnt5*, *wnt8*, and *wnt16*; transcript levels for *wnt8* are plotted on a different scale shown at right. (B) Late expressed *wnt* genes: *wnt3*, *wnt6*, *wnt7*, *wnt9*, *wnt10*, and *wntA*. (C) *frizzled* genes: *fzd1/2/7*, *fzd4*, *fzd5/8*, and *fzd9/10*. (D) Genes encoding soluble frizzled related proteins, and Dickkopf proteins: *sfrp1/5*, *sfrp3/4*, *dkk1*, and *dkk3*.

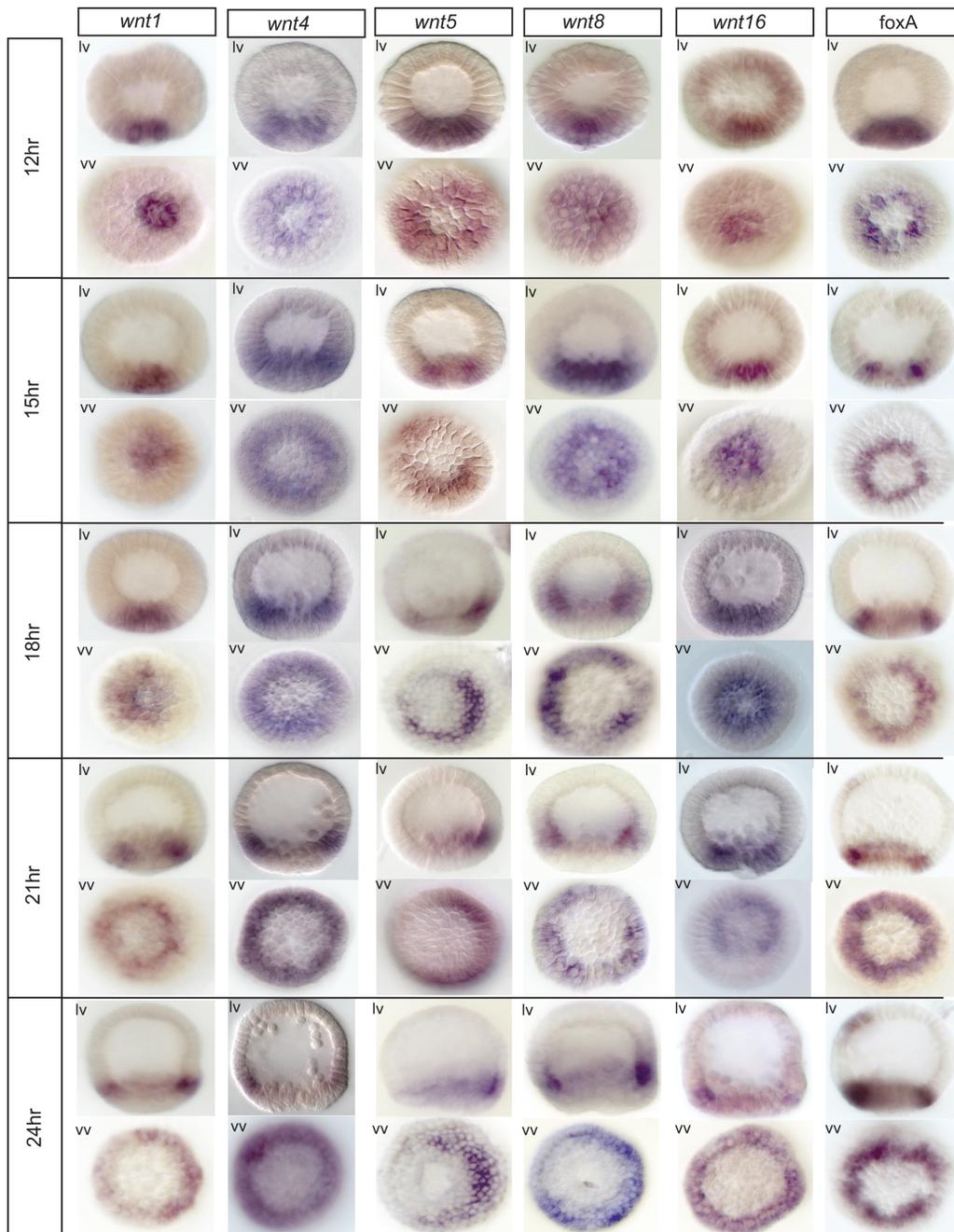


Figure S2. Spatial expression of *wnt* genes, 12-24h. Representative WMISH images for all early expressed *wnt* genes at 12h, 15h, 18h, 21h, and 24h, using probes against all early expressed *wnt* genes. Expression of *foxA* in *veg2* endoderm at the same stages is shown to provide a spatial reference in mapping *wnt* expression domains.

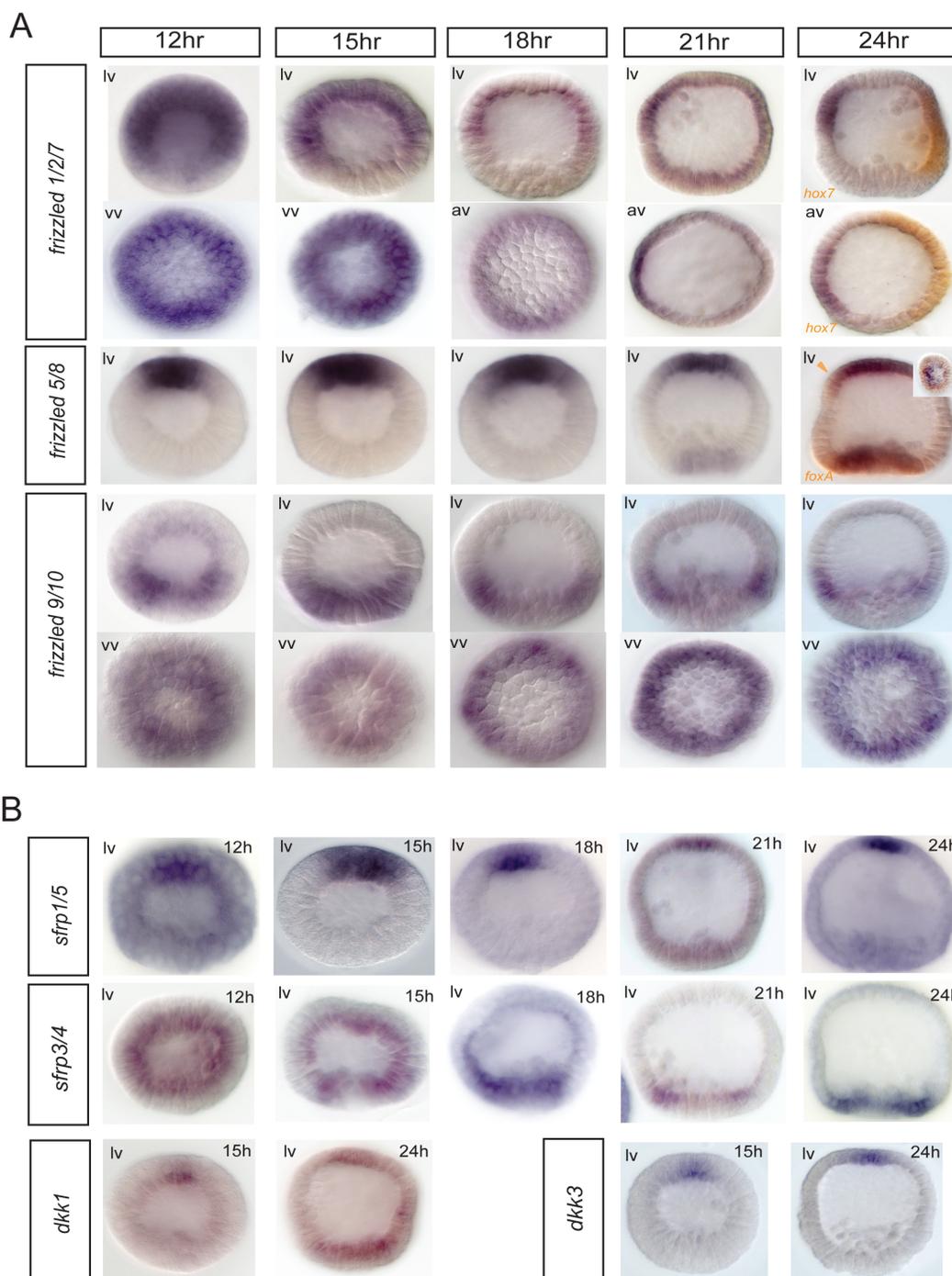


Figure S3. Spatial expression of *fzd*, *sfrp*, and *dkk* genes, 12-24h. (A) Representative WMISH images for *fzd* genes. (B) Representative WMISH images for *sfrp* and *dkk* genes (*dkk* genes shown only at 15 and 24h).

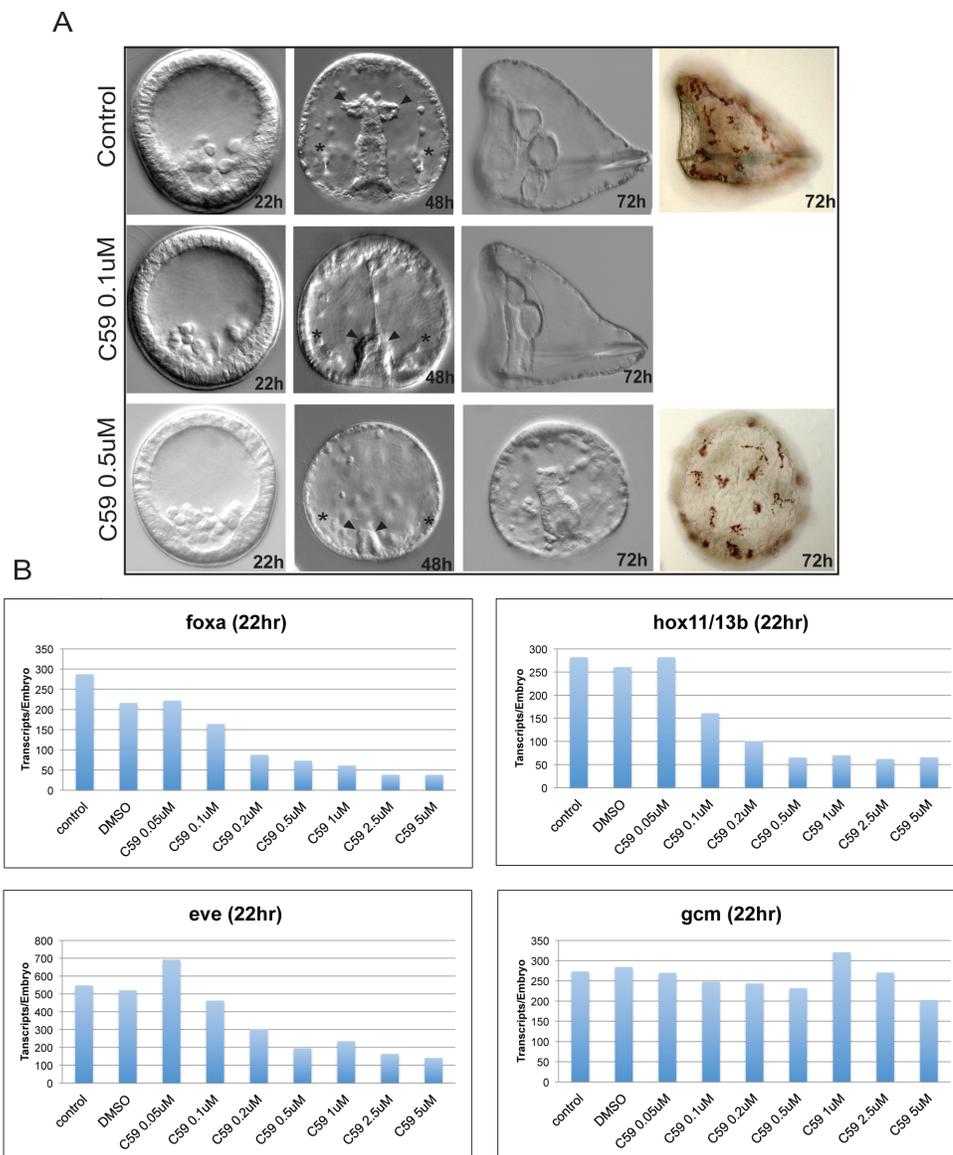


Figure S4. Effects of C59 porcupine inhibitor on morphology and gene expression. (A) Embryonic morphology at 22h, 48h, and 72h in the presence of 0.1% DMSO (control), 0.1µM C59 or 0.5µM C59. Inhibition of Wnt signaling affects the formation of the gut, the development of coelomic pouches (arrow head), and the arrangement and production of spicules by skeletogenic cells (asterisk) in a dose-dependent manner. Specification and patterning of pigment cells are not affected, as shown in light photographic images at right. (B) C59 effects on endomesodermal regulatory genes at various concentrations. QPCR measurements are shown for the veg2 endodermal regulator *foxa*, the veg1 posterior endoderm regulators *eve* and *hox11/13b* and aboral mesodermal regulator *gcm*, in control embryos and in embryos exposed to C59 at the concentrations between 0.1µM and 5µM.

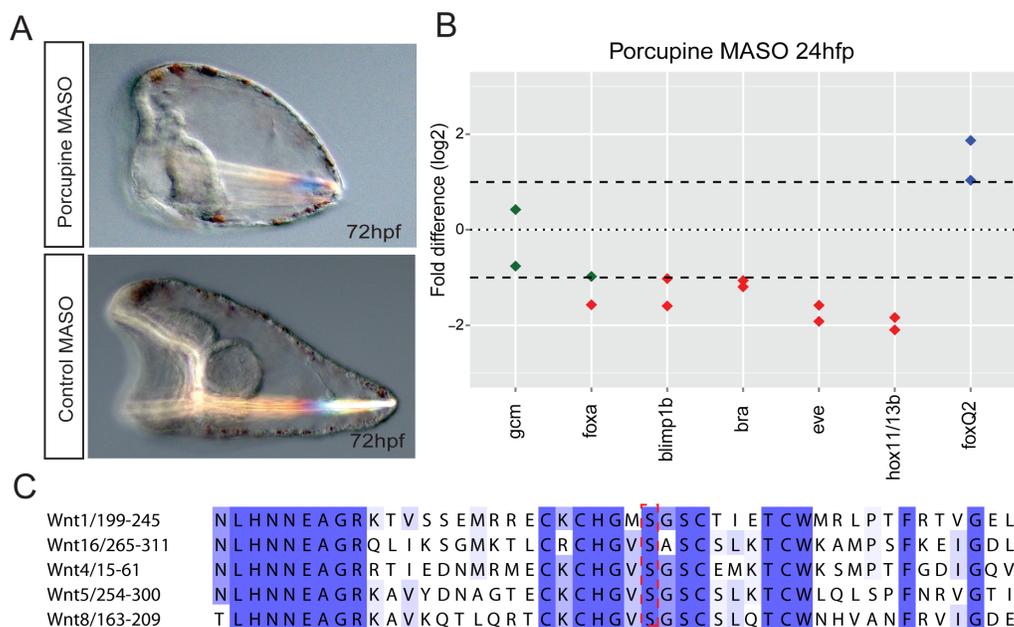


Figure S5. Porcupine in sea urchin embryos. (A) Morphological effects of Porcupine knock-down by morpholino, seen at 72h. (B) Endomesodermal genes expression of which is significantly decreased by Porcupine MASO are the same as those responding to inhibition of Wnt signaling by C59; QPCR analysis, symbolism as in Fig.2 of text. (C) Porcupine target sequences shared among all early expressed the *wnt* genes.

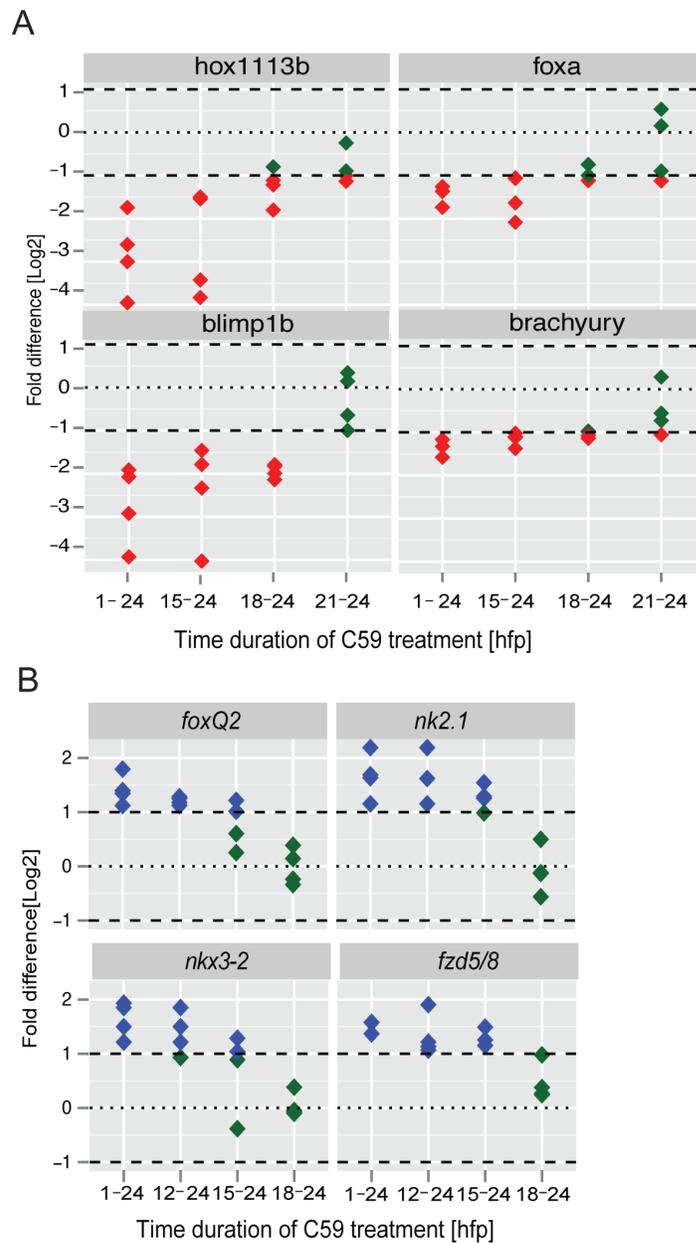


Figure S6. Efficacy windows for C59 inhibition of regulatory genes expression. Embryos were incubated with 0.5 μ M of C59 for different developmental intervals, as indicated on the abscissae. Transcripts of genes were measured by QPCR at 24h. (A) Endodermal genes, *hox11/13b*, *foxa*, *blimp1b*, *brachyury*; as above red symbols indicate significant decreases. Maximal effects were obtained in C59 treatment between 15 to 24h. (B) Apical genes, *foxq2*, *nk2.1*, *nkx3-2*, *fzd5/8*; here as discussed in text, C59 causes up-regulation as due to expansion of expression domain, and maximal effects were obtained at the earliest exposure interval.

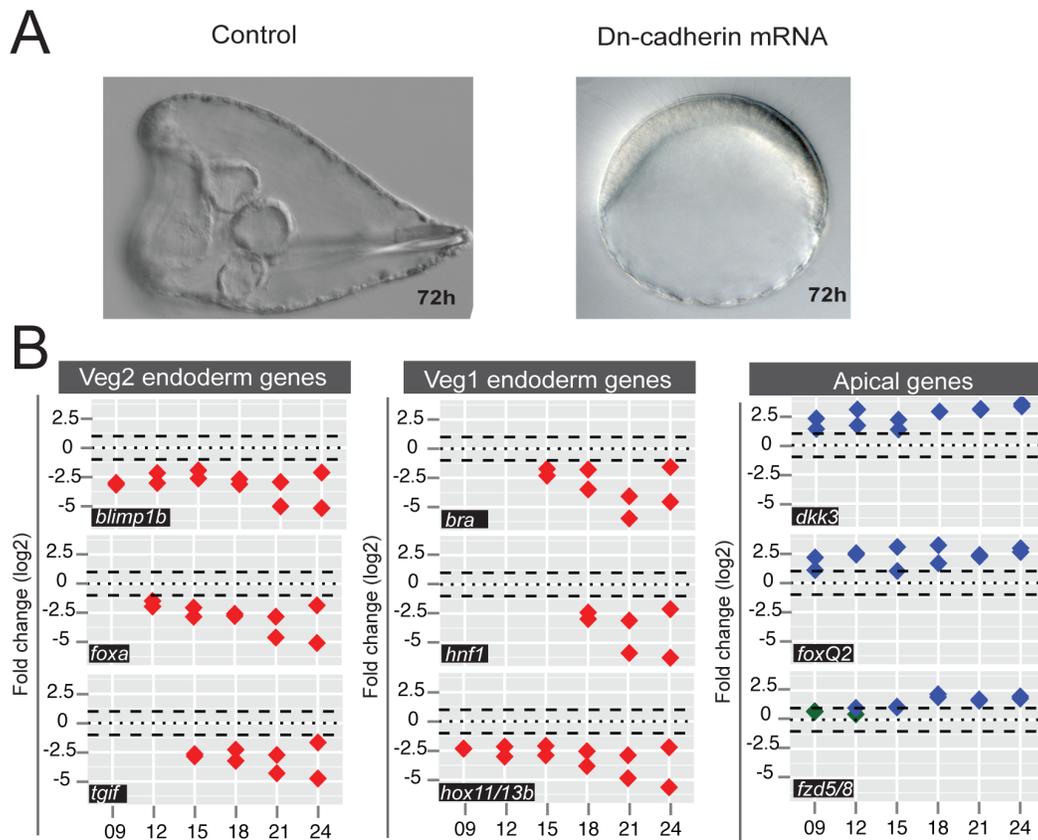


Figure S7. Effects of maternal β -catenin depletion by Dn-cadherin mRNA injection. A) Complete animalization phenotype of embryos shown at 72hpf. B) Timing of effects on veg1 and veg2 endodermal gene, *blimp1b*, *foxa*, *hox11/13b*; veg1 endodermal genes, *eve*, *bra*, *hnf1*; apical genes, *dkk3*, *foxq2*, and *fzd5/8*. Significant depression of veg1 and veg2 endodermal genes is observed from the earliest time these genes are expressed. Similarly up-regulation of apical genes is observed at the earliest time of their expression; symbolism as in Fig.2 of the text.

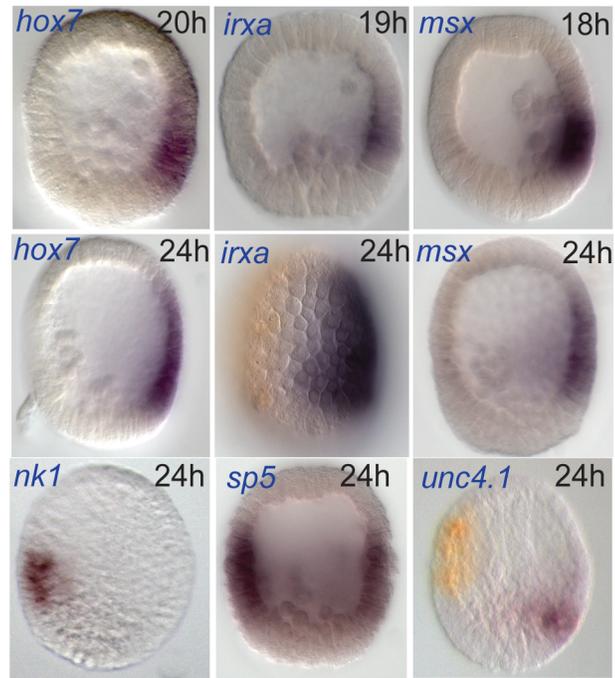
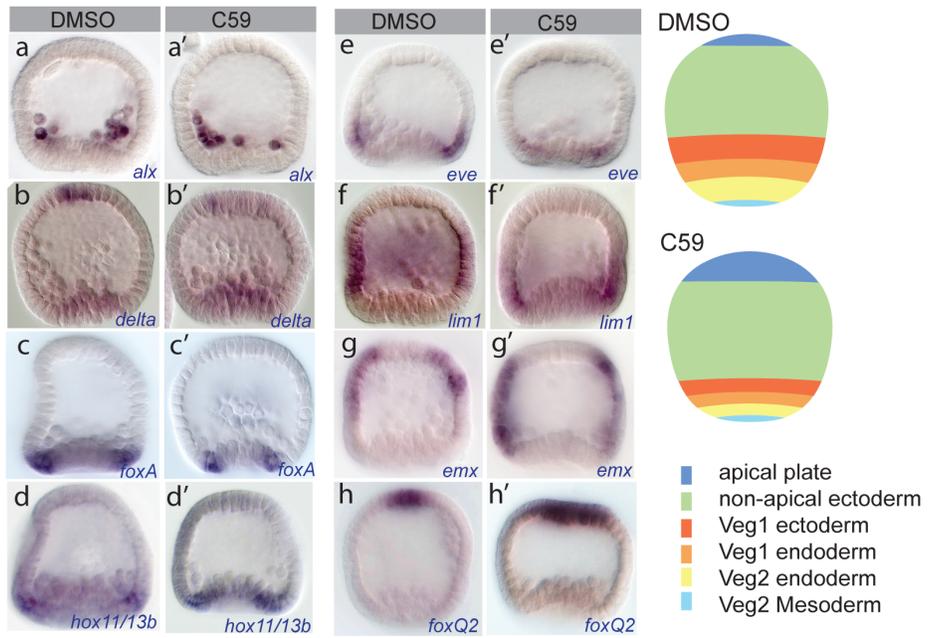


Figure S8. Normal spatial expression of Veg1 genes. Changing expression patterns of *hox7*, *irxa*, and *msx*, 4-6 hours before 24h. All three of these genes are initially activated in the veg1 cells whereupon the transcription expands to include the aboral animal ectoderm. In the lower tier are shown the expression of *nk1* and *sp5* in the oral veg1 ectoderm, and of *unc4.1* in the aboral veg1 ectoderm.

A



B

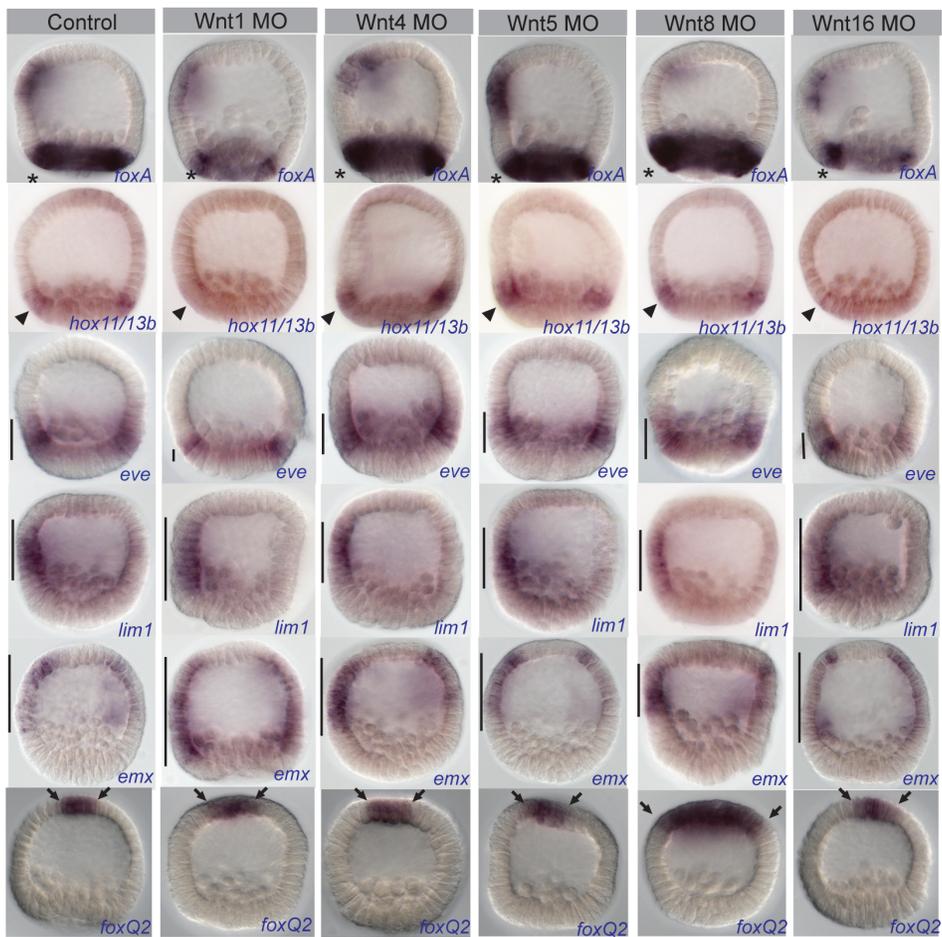


Figure S9. Wnt signaling functions in the spatial patterning of regulatory territories along the animal-vegetal axis. (A) WMISH displays of regulatory genes expression in C59 treated embryos, and in controls exposed only to DMSO. Genes studied are normally expressed in the following domains: *alx*, skeletogenic mesoderm; *delta*, *veg2* mesoderm; *foxa*, anterior endoderm; *hox11/13b*, posterior endoderm; *eve*, posterior endoderm plus vegetal ectoderm; *lim1* and *emx*, animal ectoderm; and *foxq2*, apical neurogenic domain. Schematic, summarizing overall alterations in spatial disposition of regulatory state domains indicated in the key below. In general, it can be seen that the endomesodermal domains contain fewer cells and the apical domain expands in C59 treated embryos. (B) Effects on expression patterns of same genes resulting from morpholino knockdown of each of the early expressed wnt genes. Symbols: asterisks, arrowheads, vertical lines, and arrows mark the expression domains in normal and perturbed embryos.

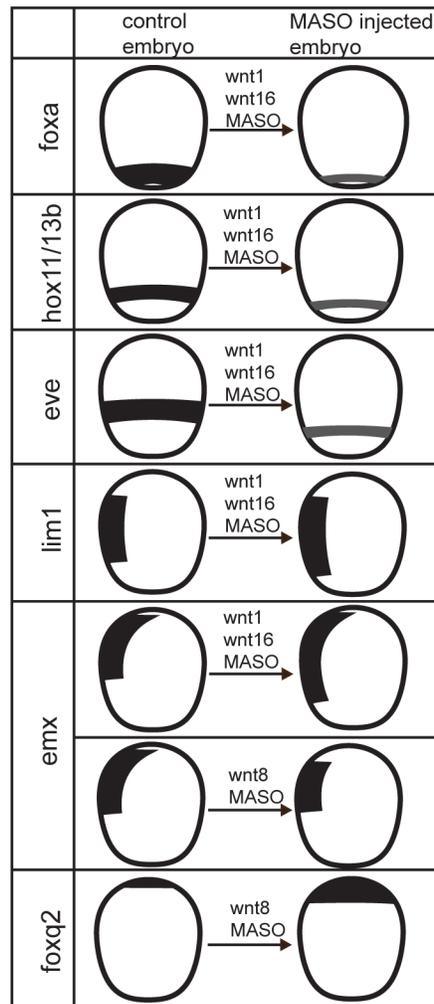


Figure S10. Schematic summary of changes in spatial gene expression upon Wnt morpholino injection. Embryos were injected with morpholinos targeting

Wnt1, Wnt4, Wnt5, Wnt8 or Wnt16 and analyzed by WMISH at 24h. Spatial expression of regulatory genes representing regulatory state domains throughout the embryo are shown in Fig. S9. Here the expression domains of regulatory genes are schematically represented for control embryos (left schemes) and Wnt morpholino injected embryos (right schemes). Wnt morpholinos responsible for observed effects are indicated. Expression domains represented by regulatory genes at 24h are as follows: *foxa* (veg2 endoderm), *hox11/13b* (veg1 endoderm), *eve* (veg1 endoderm and veg1 ectoderm), *lim1* (veg1 ectoderm, oral animal ectoderm), *emx* (oral animal ectoderm), and *foxq2* (apical neurogenic ectoderm)

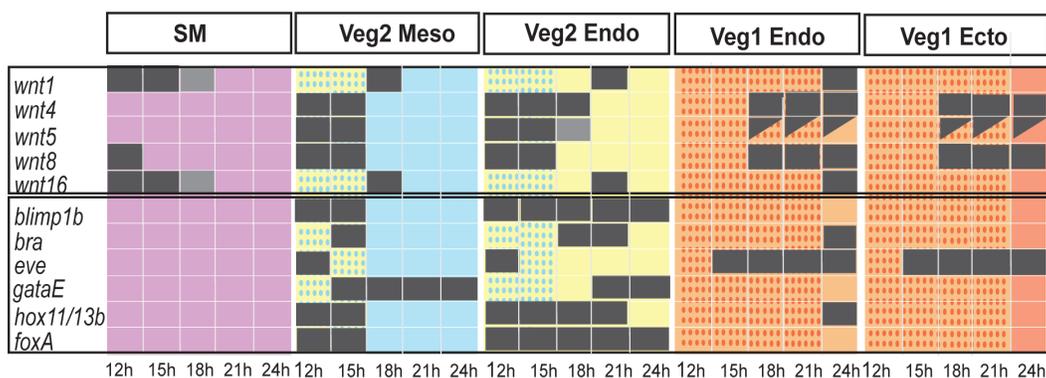


Figure S11. Expression matrix for candidate *wnt* activators in the vegetal lineages. Expression matrix for the five early expressed *wnt* genes is shown as in Fig.1, and below for comparison, similar expression matrixes are shown for candidate endomesodermal regulatory genes which were tested as candidate drivers of the *wnt* genes.

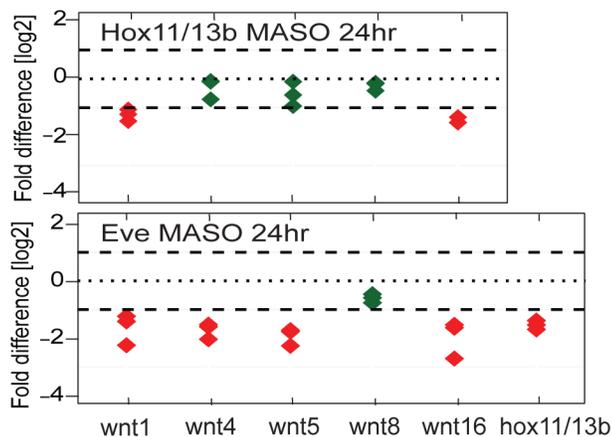


Figure S12. Roles of Hox11/13b and Eve in the control of *wnt* gene expression. Expression levels of *wnt* genes were analyzed by qPCR at 24h in embryos injected with morpholinos blocking expression of Hox11/13b or Eve, or with randomized control morpholinos. Results are shown as ratios (log₂) of expression levels of *wnt* genes in embryos injected with gene-specific morpholinos versus control morpholinos. Results of three independent experiments are shown, with symbolism as in Fig. 2

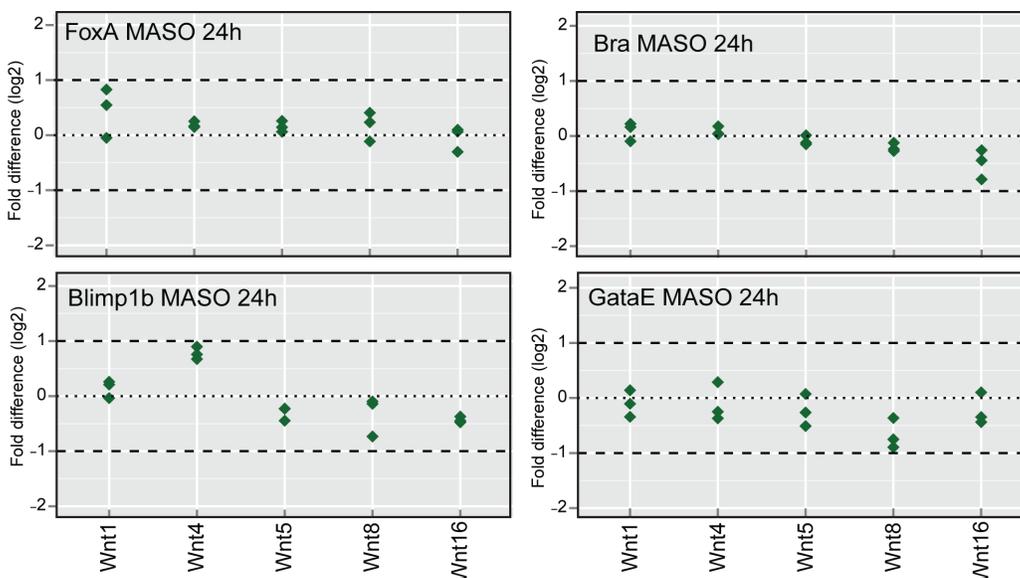


Figure S13. Lack of effect of *foxa*, *brachyury*, *blimp1b*, and *gatae* morpholinos on transcript levels of the five early expressed *wnt* genes. QPCR data are shown, symbolism as in Fig.2 of text.

Genes	12h	15h	18h	21h	24h	Genes	12h	15h	18h	21h	24h
acsc	N/E	N/E	N/E	2.72	1.84	ngn	N/E	N/E	N/E	N/E	N/E
acti nb	N/E	N/E	N/E	N/E	N/E	nk1	N/E	N/E	0.33	0.28	0.26
alx1	0.91	1.02	1.01	1.03	1.15	nk2.1	N/E	2.06	2.77	3.46	2.81
arnt	N/E	N/E	N/E	1.06	1.08	nk2.2	0.80	0.76	0.65	0.66	0.66
atbf1	1.07	0.95	1.18	1.23	1.11	nkx3-2	N/E	2.43	3.89	4.65	4.27
atf2	1.06	1.12	1.18	1.03	1.09	nlk	0.94	1.01	1.02	0.98	0.92
blimp 1a	N/E	N/E	N/E	N/E	N/E	nodal	1.06	0.91	0.73	0.77	0.85
blimp 1b	0.81	0.50	0.24	0.23	0.18	not	1.00	1.06	0.80	0.81	1.04
bmp2 /4	1.12	0.98	0.93	0.99	1.12	notch	0.99	1.09	1.02	0.95	0.97
bmp5 /8	0.88	0.94	1.07	1.10	1.11	nr1h6b	1.10	N/E	N/E	N/E	1.06
bra	0.81	0.38	0.41	0.36	0.52	otp	1.06	N/E	N/E	N/E	N/E
brn1/ 2/4	1.07	0.98	1.21	1.02	0.99	otx alpha	1.15	1.25	0.80	0.92	1.21
cdx	N/E	N/E	N/E	N/E	N/E	otx beta1/ 2	1.21	0.87	0.79	0.90	1.94
cebpa	1.00	1.19	1.43	1.31	1.47	pax2/5 /8	N/E	N/E	N/E	0.80	0.88
chd	N/E	0.95	0.82	0.78	0.87	pax4	0.84	0.91	1.03	0.99	1.10
coe	N/E	N/E	N/E	N/E	N/E	pax6	N/E	N/E	N/E	N/E	N/E
cyclo philin	0.89	0.93	1.07	0.98	1.01	paxb	0.92	1.00	1.00	1.01	0.97
dach	0.93	1.01	1.07	1.08	1.05	paxc	N/E	N/E	N/E	N/E	N/E
delta	0.90	0.87	1.15	1.43	1.49	pea	0.85	0.89	1.14	1.02	0.95
dlx	N/E	N/E	0.72	0.70	0.77	pitx2	N/E	N/E	N/E	N/E	N/E
dmrt	N/E	N/E	N/E	N/E	N/E	pks	N/E	0.70	0.91	0.87	0.88
dpc4	1.05	1.05	1.05	1.04	0.96	pmar	0.99	1.10	1.11	0.96	1.02
dri	0.77	0.88	1.11	1.16	0.99	prox1	1.09	0.94	1.03	0.92	0.94
e2a	0.95	0.90	1.11	1.13	1.10	ptc1	1.11	0.98	1.17	1.30	1.31
e2f3	0.93	0.96	0.99	1.09	1.03	ptf1a	N/E	N/E	N/E	N/E	0.55
e2f4	1.05	0.97	1.11	1.03	1.06	rel	N/E	N/E	N/E	1.12	1.09

ecr	1.28	0.95	1.06	1.04	0.84	reverb	N/E	N/E	N/E	0.99	1.05
egf2	0.86	0.90	0.97	0.98	0.98	rfp	N/E	N/E	N/E	N/E	N/E
elfb	N/E	0.96	1.03	1.02	0.97	runt1	1.03	1.25	1.35	1.26	1.30
elk	0.90	1.09	1.03	0.92	1.00	rx	N/E	N/E	N/E	3.00	2.39
emx	0.82	0.66	0.83	1.36	1.32	scl	0.91	1.07	0.99	0.86	0.95
endo16	N/E	0.56	0.64	0.59	0.56	scratchx	N/E	N/E	N/E	N/E	N/E
erf	1.09	0.97	0.93	0.81	1.02	shr2/tf2.4	0.99	1.03	1.13	1.02	1.03
erg	0.87	0.82	0.93	0.96	1.01	sim	N/E	N/E	N/E	N/E	N/E
ese	1.03	1.19	1.05	1.01	1.02	sip/sm adip	0.97	1.34	1.33	1.10	1.19
ets1/2	1.19	1.12	1.15	1.01	1.11	six1/2	N/E	N/E	N/E	1.03	0.78
ets4	1.17	0.95	0.91	1.23	1.14	six3	1.12	1.27	1.57	1.54	1.31
eve	0.66	0.51	0.34	0.43	0.54	sm50	0.82	1.22	1.07	0.99	1.01
fgf	N/E	N/E	N/E	1.05	1.07	smad2/3	0.94	0.98	1.05	0.99	1.03
fgfr3	0.83	1.11	1.37	1.28	1.09	smad4	N/E	N/E	N/E	N/E	N/E
fic	0.81	0.88	1.04	1.18	1.30	smad5	0.94	1.00	1.00	0.93	0.95
follistatin	N/E	N/E	N/E	N/E	N/E	smo	1.01	1.11	1.01	0.95	1.11
foxa	0.66	0.61	0.32	0.35	0.46	snail	N/E	N/E	N/E	N/E	0.68
foxb	N/E	N/E	N/E	0.84	1.26	soxb1	0.95	1.04	1.03	1.31	1.37
foxc	N/E	N/E	N/E	N/E	N/E	soxb2	0.87	0.73	1.07	1.14	1.01
foxd	N/E	N/E	N/E	N/E	N/E	soxc	0.91	0.82	0.98	1.04	1.05
foxf	N/E	N/E	N/E	N/E	N/E	soxd1	1.21	1.25	1.56	0.86	1.11
foxg	N/E	1.10	1.15	1.17	1.24	soxe	N/E	N/E	N/E	N/E	N/E
foxi	N/E	N/E	N/E	N/E	N/E	soxf	1.15	N/E	N/E	N/E	1.31
foxj1	0.92	1.16	2.12	2.12	1.87	spec1	0.85	0.87	0.94	0.94	0.94
foxj2	1.02	1.05	1.12	1.19	1.15	spz12	0.92	1.02	1.01	1.09	1.02
foxk	0.92	0.99	0.99	1.06	0.99	srf	0.99	1.05	1.02	0.91	0.95
foxm	0.97	0.97	1.05	1.04	1.01	su(h)	0.95	1.11	1.22	0.97	1.05
foxn2/3	0.84	0.79	0.71	0.55	0.67	tbr	1.05	0.90	0.81	0.83	1.04
foxo	N/E	N/E	N/E	1.19	1.06	tbx2/3	N/E	N/E	0.98	0.92	0.79
foxp	0.91	0.99	1.11	1.05	0.96	tbx6	1.13	0.96	1.04	1.24	1.08
foxq2	1.24	2.95	3.12	3.67	3.63	tcf	1.09	0.98	1.40	1.04	1.09
foxy	1.12	1.20	1.20	1.35	1.12	tead4	0.94	0.96	1.19	1.10	1.07
fzd4	1.22	1.10	0.96	1.04	0.78	tel	0.92	0.92	0.97	0.97	1.05
fxr	N/E	N/E	N/E	N/E	N/E	tgif	N/E	0.90	0.84	0.51	0.41
gatac	N/E	N/E	N/E	0.85	1.04	thr	1.06	1.15	1.24	1.12	1.01
gatae	N/E	0.56	0.63	0.45	0.33	tll1	N/E	N/E	N/E	0.55	0.86
gbx	N/E	N/E	N/E	N/E	0.44	tulp4l	1.05	1.01	1.08	1.04	1.02
gcm	0.76	0.78	0.86	0.85	0.82	ubq	1.08	1.05	1.00	1.04	1.00

gfp	N/E	N/E	N/E	N/E	N/E	unc4.1	N/E	N/E	N/E	0.62	0.44
gsc	N/E	1.31	1.13	1.19	1.21	univin	0.89	0.79	0.76	0.94	1.09
hbn	0.96	2.42	2.87	2.72	2.30	usf	0.88	0.86	1.06	0.93	0.96
hes	1.17	1.50	1.42	1.19	1.25	vegfb3	0.90	0.76	0.82	0.78	0.78
hesc	0.84	0.78	0.71	0.81	0.88	vegfr	N/E	N/E	1.01	1.03	1.06
hex	0.84	0.93	0.94	0.84	0.85	vitellog enin2	N/E	1.52	2.22	2.51	2.50
hey4	0.89	0.63	0.56	0.50	0.51	wnt1	0.80	0.67	0.60	0.57	0.58
hh	N/E	N/E	N/E	N/E	N/E	wnt16	0.92	0.86	0.88	0.68	0.43
hlf	0.86	0.93	1.18	1.19	1.17	wnt3	N/E	N/E	N/E	N/E	N/E
hlx	N/E	N/E	N/E	N/E	N/E	wnt4	0.88	0.88	0.67	0.62	0.68
hmg1	0.93	0.95	1.05	1.05	1.04	wnt5	0.76	0.83	0.58	0.45	0.46
hmg2	0.86	0.83	1.10	1.04	1.02	wnt6	N/E	N/E	N/E	N/E	N/E
hmx1	0.73	0.51	0.65	1.01	0.93	wnt8	0.89	0.65	0.38	0.47	0.64
hnf1	N/E	N/E	0.23	0.17	0.16	wnta	N/E	N/E	N/E	N/E	N/E
hnf4	N/E	N/E	N/E	N/E	N/E	z108	0.91	1.08	1.00	1.00	1.01
hnf6	1.07	1.05	1.15	1.23	1.24	z115	0.93	1.00	1.08	0.98	0.94
hox1 1/13 b	0.79	0.79	0.47	0.28	0.33	z121	N/E	N/E	N/E	N/E	N/E
hox7	N/E	N/E	N/E	0.55	0.56	z13/kr l	0.80	0.60	0.37	0.45	0.89
id	0.90	0.76	1.20	0.83	0.85	z133/f ez	N/E	N/E	N/E	N/E	N/E
irf4	N/E	N/E	N/E	N/E	N/E	z141	0.97	1.02	1.01	1.04	0.95
irxa	N/E	N/E	0.49	0.76	0.82	z157/o vo	0.96	1.03	1.06	1.08	1.31
jun	1.11	1.10	1.15	1.21	1.18	z166	N/E	0.74	0.87	0.89	0.97
lefty	1.09	0.96	0.75	0.74	0.89	z188/k lf13	0.99	1.15	1.63	1.41	1.30
lhx2. 9	N/E	N/E	N/E	3.24	2.69	z199/s p5	0.58	0.25	0.07	0.18	0.30
lim1	0.73	0.78	0.58	0.61	0.70	z204	N/E	N/E	N/E	N/E	1.07
lmo4	N/E	N/E	N/E	N/E	0.82	z214	0.88	0.96	1.09	1.03	0.96
lmx1	N/E	N/E	N/E	N/E	N/E	z22/gli 1	1.07	0.93	0.98	0.91	0.97
lox	N/E	N/E	N/E	N/E	N/E	z244/z ic	1.04	2.15	2.97	3.26	2.30
mad	0.94	1.02	1.11	1.02	0.97	z30	0.97	1.06	1.20	1.28	1.05
max	0.98	0.97	1.04	0.98	0.97	z400	0.86	0.97	1.10	1.02	0.96
mbx1	N/E	N/E	N/E	N/E	N/E	z48	1.02	0.84	1.19	0.84	0.83
mef2	0.91	0.91	1.03	1.06	1.13	z487	0.98	1.07	1.12	1.06	1.09
mitf/t fe3	1.02	0.95	1.15	0.92	0.99	z54/sp alt	0.92	0.98	1.05	1.02	0.86
mlx	1.05	1.05	1.04	1.00	0.99	z55	0.96	0.96	1.11	1.05	1.01
Msx	N/E	N/E	0.43	0.61	0.85	z57	0.85	0.91	1.01	1.11	1.04
myb	0.97	0.99	1.04	1.03	1.03	z60	0.92	0.92	0.87	1.07	1.25

myc	0.85	0.70	0.83	0.77	0.79	z62	0.89	0.96	1.06	1.04	1.02
myor 2	N/E	N/E	N/E	N/E	N/E	z85/klf 2/4	0.91	0.90	1.09	0.99	0.87
nfe2	0.93	0.97	0.96	0.95	0.99	z86	0.99	0.95	0.99	1.05	0.91
nfkb	0.95	0.97	1.07	1.08	1.07	z92	0.92	0.95	1.00	1.01	0.97

N/E: counts \leq 200

Table S1. Complete list of genes examined by nCounter and their fold-change upon inhibition of Wnt signaling by C59 treatment. Gene counts between samples are normalized by total reads and are corrected for background using negative spikes. Only gene counts that are great than 200 after normalization and correction are thus calculated for fold changes. Numbers shown are the average of three independent experiment.

Regulatory genes affected by Wnt signaling	Evidence ¹	Effect of C59 on gene expression	Tcf target ²	Domain of expression at 24h	Source
blimp1b	C59, Wnt1 MASO, Wnt16 MASO, CRA	Decreased	Y	veg2 Endo	This study, 1, 9
bra	C59, Wnt1 MASO, CRA	Decreased	Y	veg1 Endo	This study, 1
eve	C59, Wnt1 MASO, Wnt16 MASO, CRA	Decreased	Y	veg1 Endo, veg1 Ecto	This study, 9
foxa	C59, Wnt1 MASO, CRA	Decreased	Y	veg2 Endo	This study, 1, 7
gatae	C59, Wnt1 MASO, CRA	Decreased	NA	veg2 Meso and veg2 Endo	This study
gbx	C59	Decreased	NA	NA	This study
hey4	C59, Wnt1 MASO	Decreased	NA	NA	This study
hnf1	C59, Wnt1 MASO	Decreased	NA	veg1 Endo	This study
hox7	Wnt4 MASO	Decreased	NA	Aboral veg1, Aboral animal Ecto	This study
hox11/13b	C59, Wnt1 MASO, Wnt16 MASO, CRA	Decreased	NA	veg1 Endo	This study, 1
irxa	C59	Decreased	NA	Aboral veg1, Aboral animal Ecto	This study, 10
lim1	Wnt5 MASO	Decreased	NA	Lateral and Oral veg1, Lateral and Oral animal Ecto	10
msx	C59, Wnt4 MASO	Decreased	NA	Aboral veg1, Aboral animal Ecto	This study
nk1	C59, Wnt1 MASO	Decreased	NA	Oral veg1	This study, 10
pax2/5/8	Wnt5 MASO	Decreased	NA	Lateral veg1 ectoderm	10
tgif	C59, Wnt1 MASO	Decreased	NA	veg2 Endo	This study
unc4.1	C59, Wnt1 MASO	Decreased	NA	Aboral veg1	This study
wnt1	C59, Wnt1 MASO	Decreased	NA	veg1 Endo	This study
wnt4	C59, Wnt1 MASO	Decreased	NA	veg1 Endo, veg1 Ecto	This study
wnt5	C59, Wnt1 MASO, Wnt5 MASO	Decreased	NA	veg1 Endo, veg1 Ecto	This study, 10
wnt8	C59, Wnt1 MASO, CRA	Decreased	Y	veg1 Endo, veg1 Ecto	This study, 8
wnt16	C59, Wnt1 MASO	Decreased	NA	veg1 Endo	This study
z13/kr1	C59, Wnt1 MASO	Decreased	NA	veg2 Endo	This study
z199/sp5	C59, Wnt1 MASO	Decreased	NA	Oral veg1, Oral animal Ecto	This study
acsc	C59, Wnt8 MASO	Increased	N	NA	This study
foxj1	C59, Wnt8 MASO	Increased	N	Apical plate	This study
foxq2	C59, Wnt8 MASO	Increased	N	Apical plate	This study, 2

hbn	C59, Wnt8 MASO	Increased	N	Apical plate	This study
ihx2.9	C59, Wnt8 MASO	Increased	N	Apical plate	This study
nk2.1	C59, Wnt8 MASO	Increased	N	Apical plate	This study
nkx3-2	C59, Wnt8 MASO	Increased	N	Apical plate	This study
rx	C59, Wnt8 MASO	Increased	N	Apical plate	This study
z244	C59, Wnt8 MASO	Increased	N	Apical plate	This study

¹ CRA: Cis-Regulatory Analysis

² Regulatory genes directly regulated by TCF as shown by cis regulatory studies; Y, direct target genes; N, genes affected by C59 but are not direct TCF targets; NA, potential TCF target genes, but not analyzed at the cis-regulatory level.

Table S2. Summary of experimental evidence for regulatory genes affected by Wnt signaling.

Gene	ID	Q-PCR Primer fwd	Q-PCR Primer rev	WMISH primer fwd	WMISH primer rev
wnt1	WHL22.59678 2	TGCGATCTTATGTGCT GCTC	GAAACGACGTGCAC TCTTCA	TTAATCCGCAGC AAATAGCC	ACAACCTCCTCAG CGTTTCC
wnt3	WHL22.87121	GAGAGCGAGGACAAAG ATGG	CCACCAGCAGGTCT TGAGTT		
wnt4	WHL22.58760 6	GCGGACTTTAAGCCTC ACAC	GTCGTCCATGAGTT CCCAGT	CCTTTGAGCCAA TTTGTTTG	TCTCGGACATTAC GGAAACA
wnt5	WHL22.52081	TGCTGTGGAAGAGGCT ACAA	TTCTGCACTTCCGA CACTTG	AGGAAGGACTGT GCTCGAA	AGGAAAACACTCC CGAAGAC
wnt6	WHL22.59678 4	CGGGCTCCTGTACTCT CAAG	GGTGGCTGTTTGA CAGTTT		
wnt7	WHL22.47502 7	AGGCATGTATGCAGGG AAAC	GAAAAGCCGTGAAA ACCTGA		
wnt8	WHL22.8923	TGTCGTTTCAATCAAGC CATC	TATCACTCGCCATT CGTTCA	GTCACCAGCAAG CAACGTTT	AACACCAAACGAA GTTGCAG
wnt9	WHL22.59677 5	TGACCTTGGAATAAGG ACCG	TGACCTGAACACTT CGTTTCG		
wnt10	WHL22.23530	TGTACACTACGCCGA AAAG	CATGAGACGGTTTC CAACCT		
wnt16	WHL22.73523 2	CGATCCCGAGACTCTC TGTC	CGATTTCCCGGTTA GTACGG	TCTCTCATTTTC TGTGTCAG	GTCCATGGTTTAA GCAGACC
wntA	WHL22.54094 2	ATGGGTCACTCGTGGGA ACTC	CAGTTCCATCGTTC GTTTCT		
frizzled 1/2/7	WHL22.45238	TTGCCACCACTACAGC TTTG	AACTGGTCCACGA TCTCAC	AGAGGGAAGTTA CGGCAACA	CCCATGGAATAG CACACCT
frizzled 4	SPU_008022	AGGAGGGGTGGAGAA CACT	TGATGACGGTTTGG ACGAAG		
frizzled 5/8	WHL22.42488	TCCTATCTGTTTGGCG GACT	CACTCGTTTCTGCA TTCGTA	GTGGAACAATCC ATCAGTTG	CGTGGTTGCCTAC GTAACAG
frizzled 9/10	WHL22.60681	ACGATCCTGACGTTCT CACC	GTGGCAGGCACTGT GTAGAG	TCCTTCGTGTTT TCTTGATG	GTTTCACTGATAA CAC
sfrp 1/5	WHL22.14603 1	CATGTGCGAGAACTTG GAGA	CTTTCCCGTCTTGT GTTGGT	TCACCTTGCTCG ATCACTCA	GTTTCTCCCGTT TGTCAGA
sfrp 3/4	WHL22.62451	TTGGATCTAGGGGCTT TCCT	TCTTGCCGACTTCT GATCCT	GCAGTTTGCTCC TCTCATCC	AGCTCCGAACAGG GTAAGAC
dkk1	WHL22.34222 6	TGCTATGTGAGGCAGA CAGG	GTTTCCCTGGCAAC ACATCT	CGGAGCAACTGG GTATTTGT	GCGACAGACAGGG AGAGTTC
dkk3	WHL22.68546 3	ATGGTTCGGATTATGG ACACC	CTGGGATGTTCTCT TTCCAG	CGAAACCAGCAT AGGCTCTC	TAGTTGCTTCGGC TTGGTCT

Table S3. Sequences of primer sets used for temporal and spatial characterization of Wnt signaling components.

gene	MASO sequence	MASO interfering with
wnt1	ACGCTACAAACCACTCAAGTTTCAT	Translation ¹
wnt4	GATATAAATTCCTTACTCTTCTCC	Splicing
wnt4	GAGCAGTCCATCTCCTGTTTTGGA	Translation
wnt5	TCATGGGACAGAATGATCTTCGTCA	Translation
wnt5	GAGTTCGTACACGTTTCCATTTTGC	Translation
wnt8	AGACATCCATGATGTACTCCAAT	Translation
wnt8	GTAAAGTGTTTTCTTACCTTGGAT	Translation ²
wnt16	TCTCAACAACCTCGATAGTTCAACC	Translation
porcupine	CGCACCTGCATAAACAAAGAGAGTA	Splicing
blimp1b	CTCCCTTTCGCTTGAAAAACCCGC	Translation ³
brachyury	CGCTCATTCGAGGCATAGTGGCG	Translation ⁴
eve	CAGAAACCACTCGATCAATGTTTGC	Translation ⁴
foxa	TGGGTTCCCTTTGAAATCCACGAT	Translation ⁵
gatae	GACTTACACCGACCTGATGTGGCAT	Translation ⁴
hox11/13b	AAGCCTGTTCCATGCCGATCTGCAT	Translation ⁶

Table S4. Morpholino antisense oligonucleotide (MASO) sequences.

Supplemental references:

1. Sethi AJ, Wikramanayake RM, Angerer RC, Range RC, & Angerer LM (2012) Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos. *Science* 335(6068):590-593.
2. Range, R. C., R. C. Angerer and L. M. Angerer (2013). "Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos." *PLoS Biol* 11(1): e1001467.
3. Livi, C. B. & Davidson, E. H. Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. *Dev. Biol.* 293, 513-525 (2006).

4. Peter, I. S. and E. H. Davidson (2011). "A gene regulatory network controlling the embryonic specification of endoderm." *Nature* 474(7353): 635-639.
5. Oliveri, P., Walton, K. D., Davidson, E. H. & McClay, D. R. Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo. *Development* 133, 4173-4181 (2006).
6. Arenas-Mena, C., Cameron, R. A. & Davidson, E. H. Hindgut specification and cell- adhesion functions of Sphox11/13b in the endoderm of the sea urchin embryo. *Dev. Growth Differ.* 48, 463-472 (2006).
7. Ben-Tabou de-Leon SB & Davidson EH (2010) Information processing at the foxa node of the sea urchin endomesoderm specification network. *Proc. Natl Acad. Sci. USA* 107(22):10103-10108.
8. Minokawa T, Wikramanayake AH, Davidson EH (2005) cis-Regulatory inputs of the wnt8 gene in the sea urchin endomesoderm network. *Dev Biol* 288:545–558.
9. Smith J, Kraemer E, Liu H, Theodoris C, & Davidson E (2008) A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo. *Dev. Biol.* 313(2):863-875.
10. McIntyre DC, Seay NW, Croce JC, & McClay DR (2013) Short-range Wnt5 signaling initiates specification of sea urchin posterior ectoderm. *Development (Cambridge, England)* 140(24):4881-4889.

REFERENCES:

Angerer, L. M., S. Yaguchi, R. C. Angerer and R. D. Burke (2011). "The evolution of nervous system patterning: insights from sea urchin development." Development **138**(17): 3613-3623.

Ben-Tabou de-Leon, S., Y. H. Su, K. T. Lin, E. Li and E. H. Davidson (2013). "Gene regulatory control in the sea urchin aboral ectoderm: spatial initiation, signaling inputs, and cell fate lockdown." Dev Biol **374**(1): 245-254.

Ben-Tabou de-Leon, S. B. and E. H. Davidson (2010). "Information processing at the foxa node of the sea urchin endomesoderm specification network." Proc. Natl Acad. Sci. USA **107**(22): 10103-10108.

Biechele, S., K. Cockburn, F. Lanner, B. J. Cox and J. Rossant (2013). "Porcn-dependent Wnt signaling is not required prior to mouse gastrulation." Development **140**(14): 2961-2971.

Biechele, S., B. J. Cox and J. Rossant (2011). "Porcupine homolog is required for canonical Wnt signaling and gastrulation in mouse embryos." Dev Biol **355**(2): 275-285.

Bolouri, H. and E. H. Davidson (2010). "The gene regulatory network basis of the "community effect," and analysis of a sea urchin embryo example." Dev. Biol. **340**(2): 170-178.

Chen, B., M. E. Dodge, W. Tang, J. Lu, Z. Ma, C. W. Fan, S. Wei, W. Hao, J. Kilgore, N. S. Williams, M. G. Roth, J. F. Amatruda, C. Chen and L. Lum (2009). "Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer." Nat Chem Biol **5**(2): 100-107.

Croce, J., L. Duloquin, G. Lhomond, D. R. McClay and C. Gache (2006). "Frizzled5/8 is required in secondary mesenchyme cells to initiate archenteron invagination during sea urchin development." Development **133**(3): 547-557.

Croce, J., R. Range, S. Y. Wu, E. Miranda, G. Lhomond, J. C. Peng, T. Lepage and D. R. McClay (2011). "Wnt6 activates endoderm in the sea urchin gene regulatory network." Development **138**(15): 3297-3306.

Croce, J. C. and D. R. McClay (2010). "Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo." Development **137**(1): 83-91.

Croce, J. C., S. Y. Wu, C. Byrum, R. Xu, L. Duloquin, A. H. Wikramanayake, C. Gache and D. R. McClay (2006). "A genome-wide survey of the evolutionarily

conserved Wnt pathways in the sea urchin *Strongylocentrotus purpuratus*." Dev. Biol. **300**(1): 121-131.

Herr, P. and K. Basler (2012). "Porcupine-mediated lipidation is required for Wnt recognition by Wls." Dev Biol **361**(2): 392-402.

Howard-Ashby, M., S. C. Materna, C. T. Brown, L. Chen, R. A. Cameron and E. H. Davidson (2006). "Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development." Dev. Biol. **300**(1): 74-89.

Kenny, A. P., D. W. Oleksyn, L. A. Newman, R. C. Angerer and L. M. Angerer (2003). "Tight regulation of SpSoxB factors is required for patterning and morphogenesis in sea urchin embryos." Dev. Biol. **261**(2): 412-425.

Li, E., M. Cui, I. S. Peter and E. H. Davidson (2014). "Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo." Proc Natl Acad Sci U S A.

Li, E., S. C. Materna and E. H. Davidson (2012). "Direct and indirect control of oral ectoderm regulatory gene expression by Nodal signaling in the sea urchin embryo." Dev Biol **369**(2): 377-385.

Li, E., S. C. Materna and E. H. Davidson (2013). "New regulatory circuit controlling spatial and temporal gene expression in the sea urchin embryo oral ectoderm GRN." Dev Biol **382**(1): 268-279.

Logan, C. Y., J. R. Miller, M. J. Ferkowicz and D. R. McClay (1999). "Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo." Development **126**(2): 345-357.

Lum, L. and H. Clevers (2012). "Cell biology. The unusual case of Porcupine." Science **337**(6097): 922-923.

Materna, S. C., M. Howard-Ashby, R. F. Gray and E. H. Davidson (2006). "The C2H2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development." Dev. Biol. **300**(1): 108-120.

Materna, S. C., A. Ransick, E. Li and E. H. Davidson (2013). "Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos." Dev Biol **375**(1): 92-104.

McIntyre, D. C., N. W. Seay, J. C. Croce and D. R. McClay (2013). "Short-range Wnt5 signaling initiates specification of sea urchin posterior ectoderm." Development **140**(24): 4881-4889.

Najdi, R., K. Proffitt, S. Sprowl, S. Kaur, J. Yu, T. M. Covey, D. M. Virshup and M. L. Waterman (2012). "A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities." Differentiation **84**(2): 203-213.

Oliveri, P., E. H. Davidson and D. R. McClay (2003). "Activation of pmar1 controls specification of micromeres in the sea urchin embryo." Dev. Biol. **258**(1): 32-43.

Oliveri, P., Q. Tu and E. H. Davidson (2008). "Global regulatory logic for specification of an embryonic cell lineage." Proc. Natl. Acad. Sci. USA **105**(16): 5955-5962.

Peter, I. S. and E. H. Davidson (2010). "The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage." Dev. Biol. **340**(2): 188-199.

Peter, I. S. and E. H. Davidson (2011). "A gene regulatory network controlling the embryonic specification of endoderm." Nature **474**(7353): 635-639.

Peter, I. S., E. Faure and E. H. Davidson (2012). "Feature Article: Predictive computation of genomic logic processing functions in embryonic development." Proc Natl Acad Sci U S A **109**(41): 16434-16442.

Proffitt, K. D., B. Madan, Z. Ke, V. Pendharkar, L. Ding, M. A. Lee, R. N. Hannoush and D. M. Virshup (2013). "Pharmacological inhibition of the Wnt acyltransferase PORCN prevents growth of WNT-driven mammary cancer." Cancer Res **73**(2): 502-507.

Range, R. C., R. C. Angerer and L. M. Angerer (2013). "Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos." PLoS Biol **11**(1): e1001467.

Range, R. C., J. M. Venuti and D. R. McClay (2005). "LvGroucho and nuclear beta-catenin functionally compete for Tcf binding to influence activation of the endomesoderm gene regulatory network in the sea urchin embryo." Dev. Biol. **279**(1): 252-267.

Ransick, A. and E. H. Davidson (1998). "Late specification of Veg1 lineages to endodermal fate in the sea urchin embryo." Dev. Biol. **195**(1): 38-48.

Rios-Esteves, J., B. Haugen and M. D. Resh (2014). "Identification of key residues and regions important for porcupine-mediated Wnt acylation." J Biol Chem **289**(24): 17009-17019.

Robert, N., G. Lhomond, M. Schubert and J. C. Croce (2014). "A comprehensive survey of wnt and frizzled expression in the sea urchin *Paracentrotus lividus*." Genesis **52**(3): 235-250.

Sethi, A. J., R. M. Wikramanayake, R. C. Angerer, R. C. Range and L. M. Angerer (2012). "Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos." Science **335**(6068): 590-593.

Smith, J. and E. H. Davidson (2009). "Regulative recovery in the sea urchin embryo and the stabilizing role of fail-safe gene network wiring." Proc. Natl Acad. Sci. USA **106**(43): 18291-18296.

Smith, J., E. Kraemer, H. Liu, C. Theodoris and E. Davidson (2008). "A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo." Dev. Biol. **313**(2): 863-875.

Tu, Q., C. T. Brown, E. H. Davidson and P. Oliveri (2006). "Sea urchin Forkhead gene family: phylogeny and embryonic expression." Dev. Biol. **300**(1): 49-62.

Tu, Q., R. A. Cameron and E. H. Davidson (2014). "Quantitative developmental transcriptomes of the sea urchin *Strongylocentrotus purpuratus*." Dev Biol **385**(2): 160-167.

van Amerongen, R. and R. Nusse (2009). "Towards an integrated view of Wnt signaling in development." Development **136**(19): 3205-3214.

Yaguchi, S., J. Yaguchi, R. C. Angerer and L. M. Angerer (2008). "A Wnt-FoxQ2-nodal pathway links primary and secondary axis specification in sea urchin embryos." Dev Cell **14**(1): 97-107.

Yaguchi, S., J. Yaguchi, R. C. Angerer, L. M. Angerer and R. D. Burke (2010). "TGF β signaling positions the ciliary band and patterns neurons in the sea urchin embryo." Dev Biol **347**(1): 71-81.

Yaguchi, S., J. Yaguchi, Z. Wei, Y. Jin, L. M. Angerer and K. Inaba (2011). "Fez function is required to maintain the size of the animal plate in the sea urchin embryo." Development **138**(19): 4233-4243.

*Chapter 3***Endoderm-specific HOX gene expression mediated by temporally and spatially distinct interacting cis-regulatory modules**

(Manuscript in preparation)

ABSTRACT

Precise expression of Hox genes is required for cells to maintain their relative positions and acquire their appropriate cellular identities within a developing embryo. Sea urchin *hox11/13b* is specifically and dynamically expressed in endodermal progeny cells, and is crucial for the specification of the posterior endoderm that gives rise to the hindgut region. In this study we experimentally dissect the cis-regulatory system that controls this complex pattern of *hox11/13b* expression during first two days of embryogenesis. We have discovered that the expression of *hox11/13b* is operated by two AND logic gates at different developmental stages. During the early phase of expression (<30h), an AND logic gate exists within the early cis-regulatory module. This regulatory module employs broadly expressed transcription factors (TFs) to promote prompt and robust gene expression, and T cell factor (TCF) switches to mediate strict spatial restriction. During the late phase of *hox11/13b* expression (42h-60h), an AND logic gate exists that requires two distinct regulatory modules to function, with neither module being

sufficient to drive gene expression on its own. We investigated chromatin accessibility and chromosomal architecture at both of these modules in order to clarify the mechanism underlying this rarely observed enhancer dependent/cooperative phenomenon. This study provides insight into how tightly-regulated gene expression can be achieved during development. More importantly, this work emphasizes the importance of studying cooperation among enhancers as a regulatory mechanism underlying gene expression.

INTRODUCTION

The expression of genes involved in development is precisely regulated in terms of both the timing and location. Cis-regulatory modules (CRMs) are the primary determinants of temporal/spatial patterns of gene expression. They function as integrated TF binding platforms, and are recognized by both major lineage specifiers and DNA binding effectors of signaling pathways (Davidson 2006). Understanding how gene expression is regulated by identifying cis-regulatory modules and annotating TF binding sites provides crucial insights into how specific patterns of gene expression are achieved. This information allows for a more mechanistic understanding of how nucleic acid sequence changes, such as mutations or insertions/deletions, can lead to changes in gene expression that may drive the evolution of unique body-plans or cause genetic diseases in humans (Erwin DH 2009, Laurell T 2014, VanderMeer JE 2014).

Hox genes are arguably the most important developmental genes, as they are critical for the proper establishment of regional identities along the body axes of bilaterian animals (Krumlauf 1994, Ferrier and Holland 2001). In many species Hox genes are grouped into genomic clusters, and their expression exhibits temporal and spatial collinearity in which tight correspondence exists between the order of Hox genes and the succession of their expression territories along body axes as well as the temporal sequence in their transcriptional activation (Ferrier and Holland 2001, Lemons and McGinnis 2006, Duboule 2007). Hox proteins expressed in an axial arrangement can activate downstream regulatory programs, thereby leading to differential specification along the axes of different body parts (Peter and Davidson 2015). Given these facts, it is clear that precise control of both the localization and timing of Hox gene transcription is critical for proper embryonic development.

Basal deuterostome sea urchins express 11 Hox genes which are encoded in a single disorganized cluster of their genome (Figure 2A, (Cameron, Rowen et al. 2006)). Although the spatial collinearity of some sea urchin Hox genes can be observed in larval somatocoel, such sequential expression is absent during embryonic development (Arenas-Mena, Martinez et al. 1998, Arenas-Mena, Cameron et al. 2000). In fact, only two Hox genes, *hox7* and *hox11/13b*, are expressed during embryogenesis. *Hox11/13b* is expressed in the progeny of cells of the endodermal lineage, including those of the anterior and posterior endoderm, and

it is later restricted to the posterior endoderm that gives rise to the hindgut (Arenas-Mena, Cameron et al. 2006, Peter and Davidson 2010, Peter and Davidson 2011). Hox11/13b shares homology with the vertebrate proteins Hox11 and Hox13 (Arenas-Mena, Cameron et al. 2006, Cameron, Rowen et al. 2006). The expression of *hoxd13* and *hoxa13* has been also reported in the posterior gut of vertebrates including mice, zebrafish, and chickens, suggesting a conserved or converged mechanism of spatial regulation of the expression of these genes between vertebrates and echinoderms (Warot, Fromental-Ramain et al. 1997, de Santa Barbara and Roberts 2002, Zacchetti, Duboule et al. 2007, Scotti, Kherdjemil et al. 2015). To date, however, studies of this conserved and intriguing regulatory mechanism have not yet been published.

The function of sea urchin Hox11/13b is crucial in endodermal gene regulatory networks (GRNs), particularly in the GRN of the posterior endoderm where it is the first gene to be specifically expressed and where it activates many downstream regulators of the GRN (Peter and Davidson 2010, Peter and Davidson 2011). Perturbation of Hox11/13b in sea urchin embryos causes gut malformation, with the hindgut not being properly specified and midgut genes being ectopically expressed (Arenas-Mena, Cameron et al. 2006). In our previous study, we showed that the activation of *hox11/13b* requires Wnt1 and Wnt16 signaling (Cui et al. 2015). Other prior work suggests that EVE may function as an activator during certain developmental time, and that Hox11/13b itself represses its own transcription in the anterior endoderm (Peter and Davidson 2011). The dynamic and lineage-specific

expression of *hox11/13b* cannot be adequately explained using these potential regulatory inputs. Indeed, we still lack an understanding of how the genomic control system integrates multiple inputs in order to produce the dynamic and endoderm-specific expression pattern of *hox11/13b*.

To clarify these regulatory mechanisms, in this study we performed a comprehensive cis-regulatory analysis (CRA) of *hox11/13b*. We utilized a high-throughput tag system to scan an approximately 150kb of genomic sequence of *hox11/13b* for regulatory modules that function during an extensive developmental period, from 15-60 hours post fertilization. We discovered one intronic module (*E*) and one distal module (*D*), both of which contribute to the early (15h-30h) but not late (42h-60h) expression of *hox11/13b* during embryogenesis. We determined that this early expression is regulated by the sequential activation of module *E* by two TFs, ETS, and EVE, and is spatially restricted by a TCF toggle-switch. We additionally found a second intronic module (*L*) that is required to drive *hox11/13b* expression at later developmental stages, when it is specifically expressed in the hindgut. Interestingly, we show that both module *E* and *L* are required for the late gene expression, but that neither module is sufficient on its own to mediate this expression. We investigate the nature of this “AND” logic gate of shared regulatory activity between two cis-regulatory modules by assessing the sequence basis, chromatin accessibility, and chromosomal structure. Our results explain the dynamic and lineage-specific early expression of *hox11/13b*, and more importantly

they uncover mechanisms of regulatory cooperation that are employed to control late expression in the hindgut. We suggest that this cooperative enhancer phenomenon may broadly exist as a conserved mechanism used for generating spatially and temporally regulated gene expression.

RESULTS

Expression dynamics of *hox11/13b* and identification of active regulatory modules

Transcriptional expression of the *hox11/13b* gene starts at 9 hours post-fertilization (hpf). During the mid-blastula stage around 15hpf, *hox11/13b* is expressed in all the descendants of veg2 cells, and its transcripts reach to the highest levels at this time (Figure 1). The veg2 cells later specify to two distinct lineages, with the inner veg2 descendants giving rise to veg2 mesodermal cells (esophageal muscle cells, blastocoelar immune cells, coelomic pouch cells, and pigment cells), and the outer veg2 descendants forming the anterior endoderm (veg2 endoderm) (Davidson, Cameron et al. 1998) (Figure 1). The expression of *hox11/13b* continues in both the veg2 mesoderm and the veg2 endoderm until 18hpf. Between 18hpf and 20hpf, *hox11/13b* transcripts are cleared in the veg2 mesodermal progenitors and are thereafter only expressed in veg2 endoderm (Figure 1). By 22hpf *hox11/13b* is expressed in adjacent veg1 cells, but only in the inner ring of veg1 cells that will give rise to the posterior endoderm (veg1 endoderm), and not in the outer ectodermal cells (veg1 ectoderm) (Figure 1). At the same time point, the expression of *hox11/13b* is decreased significantly in the veg2 endoderm such that it can be only detected in the veg1 endoderm. The activation of *hox11/13b* in the veg1 endoderm marks the physical separation of the distinct regulatory states within the veg1 descendant cells and, more importantly, the formation of the boundary

between embryonic endoderm and ectoderm (Li, Cui et al. 2014). Later during embryogenesis, the expression of *hox11/13b* continues in the posterior endoderm after the onset of gastrulation. At 54hpf, the invaginated endodermal archenteron partitions into morphologically recognizable foregut, midgut, and hindgut structures, and at this time *hox11/13b* is specifically expressed in the hindgut (Figure 1).

To provide a genomic context for CRA, a *hox11/13b* bacterial artificial chromosome (Gaj, Gersbach et al.) was constructed with *gfp* coding sequences knocked into its start codon. This recombinant BAC (sp4005C17) covers a sequence 60kb upstream and 90kb downstream from the *hox11/13b* TSS, as well as a partial sequence of the neighboring gene *hox11/13c* (Figure 2A). When injected into sea urchin eggs, this *hox11/13b:gfp* BAC is stably integrated into the genome of clonal founder cells during early cleavage, resulting in a random mosaic pattern of incorporation. The spatial expression pattern of our *hox11/13b:gfp* BAC was characterized by assessing its co-localization with the endogenous *veg2* marker *foxa* using double chromogenic whole mount RNA in situ hybridization (dWMISH) from a pool of embryos injected with *hox11/13b:gfp* BAC. For every *gfp* mRNA positive embryo, *gfp* expression was scored to its expressing cell lineage(s) to generate an overall distribution of its expression pattern. Transcripts of *gfp* were detected in *foxa* positive cells in over 80% of examined embryos at 15hpf (Figure 2C and Figure S1A). By 20hpf, still about 80% of embryos had undetectable *gfp* mRNA levels in *veg2* mesodermal cells and expressed *gfp* exclusively in *veg2*

endodermal cells. At 24hpf, *gfp* transcripts were detected in veg1 endodermal cells adjacent to *foxa* expressing cells in almost 85% of examined embryos, and in veg2 endodermal cells in about 60% of embryos. The prolonged expression of *gfp* in veg2 endoderm can be explained by its increased mRNA stability as compared to that of *hox11/13b*. We thus observed that the number of embryos expressing *gfp* in veg2 endoderm decreased from 80% at 20hpf to 60% at 24hpf, and to even lower levels about 40% at 30hpf indicating that *gfp* transcripts were undergoing clearance in the veg2 endoderm. At 48hpf *gfp* transcripts are only present in the posterior end of archenteron (Figure 2C), in the same cells where endogenous *hox11/13b* transcripts are present. These observations demonstrate that our *hox11/13b:gfp* BAC contains all the regulatory information necessary to mediate veg2 activation, veg2 mesodermal repression, at least part of the veg2 endodermal clearance, and veg1 endodermal activation of *gfp* expression, all of which are necessary to produce the correct spatial expression of *hox11/13b* (Figure S1A). A temporal profile measuring the transcripts level of the *hox11/13:gfp* BAC from 8hpf to 24hpf demonstrated similar expression kinetics to those of *hox11/13b* (Figure 2B). Given these results and the implications that the sequence in our *hox11/13:gfp* BAC contains all of the regulatory information necessary for normal spatial and temporal regulation of *hox11/13b*, we decided to proceed with the use of this BAC for our CRA studies.

This CRA screen was carried out using the high-throughput “barcode” system that has been previously described (Nam, Dong et al. 2010, Nam and Davidson 2012).

Briefly, the entirety of the BAC was amplified by PCR into 51 fragments, each of which was 3-5kb long and contained ~500bp of overlapping sequences on both ends (Figure 2A). These 51 fragments were each individually fused into a set of barcode-tagged vectors containing the *hox11/13b* basal promoter and a *gfp* reporter. These cis-regulatory expression constructs were then pooled and injected into fertilized eggs along with carrier DNA where they concatenate together and incorporate into the genome of early cleavage nuclei (Livant, Hough-Evans et al. 1991). The regulatory activity of each fragment was monitored by Q-PCR at specific developmental stages including 15hpf, 24hpf, 42hpf, and 60hpf using primers that specifically recognized the individual barcoded tags. Using this approach we identified two active *hox11/13b* cis-regulatory modules: *E* and *D*. Module *E* is located in an intronic region and demonstrated early regulatory activity at 15hpf and 24hpf; Module *D* is located downstream of *hox11/13b* and was active at 15hpf. Interestingly, neither active module drives expression in later stages of embryonic development, suggesting distinct mechanisms of regulatory control for early and late *hox11/13b* expression.

In order to determine whether module *E* and *D* are capable of recapitulating the spatial and temporal expression of *hox11/13b*, we injected a GFP reporter construct driven by either fragment into sea urchin eggs and measured GFP fluorescence. Because of the mosaic nature of exogenous DNA incorporation, we examined a pool of injected embryos (>50). We found that construct *D:gfp* exhibits ubiquitous GFP expression in blastula stage embryos, with fluorescence evident in almost

every cell lineage (Figure 3B, Figure S1D). Construct *E:gfp* drives specific GFP expression that largely recapitulates the early expression pattern of *hox11/13b* and resembles the regulatory behavior of *hox11/13:gfp* BAC (Figure 3B, Figure S1A and B). We thus chose to first study module *E* in detail to decipher the information that it uses to regulate the early spatially-restricted expression of *hox11/13b*.

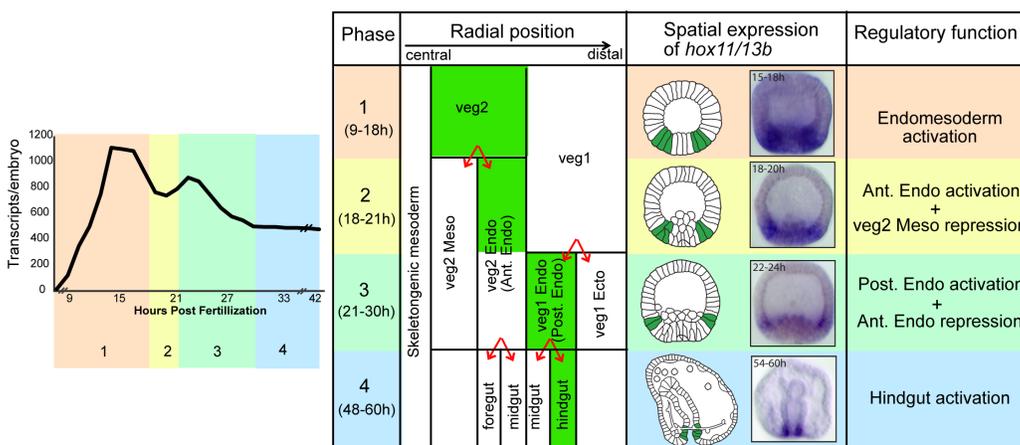


Figure 1: Dynamic and multi-phasic expression of *hox11/13b* in sea urchin embryos. The temporal expression of *hox11/13b* from 9h to 42h is shown on the left. The transcripts reach to the peak levels at 15h and 21h. The spatial expression and the expected regulatory operation at the genomic locus at each of the four expressional phases are shown on the right. The cells expressing *hox11/13b* are highlighted in green in both the process diagram, where the specification of the vegetal lineages is shown in respect to time (vertical axis) and radial position (horizontal axis), and the embryonic schemes, based on the *in situ* hybridization results (shown on the further right). All embryos are shown in lateral views. The red arrows indicate separation of regulatory states. Veg2 Mesoderm, also known as non-skeletogenic mesoderm, and the veg2 endoderm (anterior endoderm) are descendants of veg2 macromeres; Veg1 macromeres specify into veg1 endoderm (posterior endoderm) and the veg1 ectoderm, also known as the vegetal ectoderm. Ant. Endo, anterior endoderm; Post. Endo, posterior endoderm; veg2 Meso, veg2 mesoderm; veg1 Ecto, veg1 ectoderm.

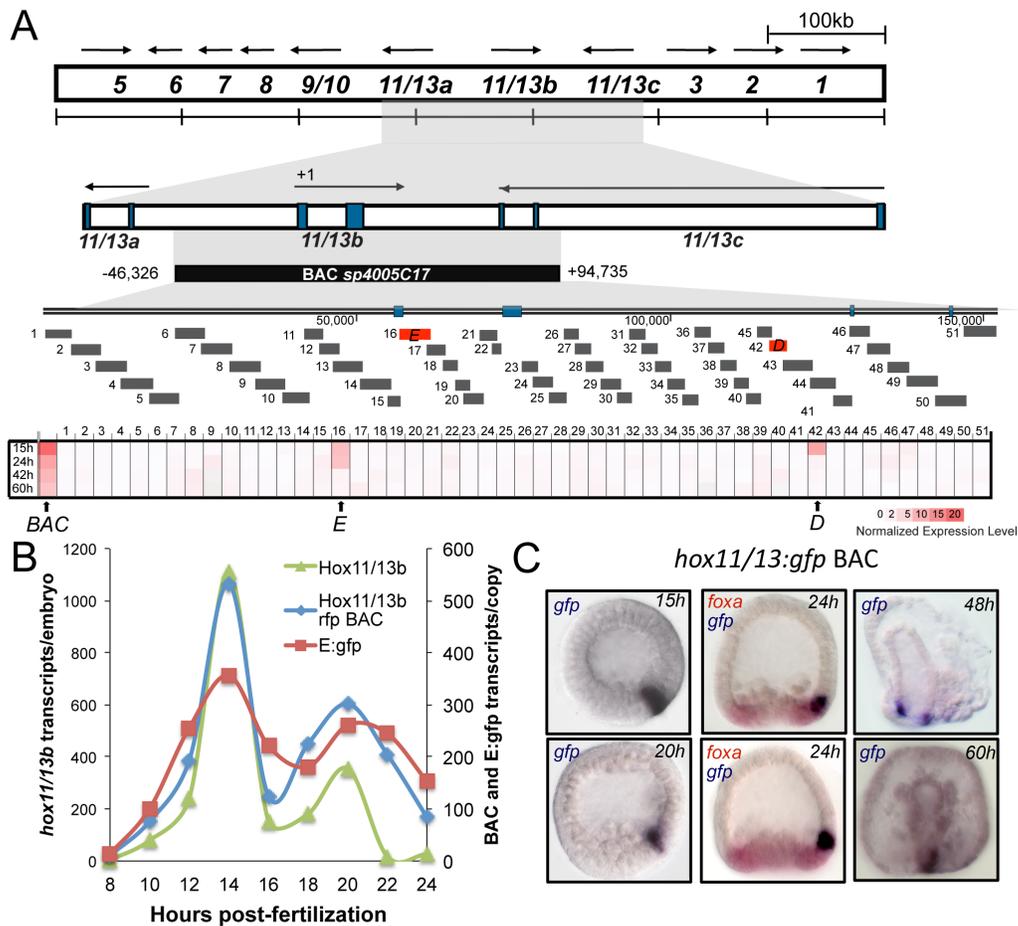


Figure 2: multiplexing reporter assay discovered early active cis-regulatory modules (E and D) of *hox11/13b* in a region of temporally and spatially correct BAC reporter. (A) A map showing the coverage of *hox11/13b* BAC sp4005C17 in the *hox* gene cluster and the 51 fragments contained in the BAC that were used to screening for active regulatory modules. The regulatory activity of the 51 fragments was measured using a “barcoded” reporter system at 15h, 24h, 42h and 60h in the embryos co-injected with the BAC reporter. Normalized activity is shown in heat map and was obtained by comparing the reporter transcript level of each fragment-reporter construct to that of an empty reporter construct. An intronic module E (active at 15h and 24h) and a downstream module D (active at 15h) were discovered; none of the fragment was able to drive the reporter expression in the later stages (42h and 60h). (B) BAC sp4005C17 and the module E exhibit similar temporal expressional dynamics compared to the endoderm *hox11/13b* gene in pre-gastrula stages of embryos (before 27h). (C) GFP *in situ* hybridization in the embryos injected with sp4005C17:GFP BAC showing the correct dynamics of spatial expression in all four phases. GFP signal was only detected in a clone of cells due to the mosaic incorporation of the BAC. The endogenous *foxa* was used to mark the anterior endoderm at 24h. A statistical

distribution of the cell lineages expressing sp4005C17:GFP BAC in a pool of injected embryos is shown in Figure S1A.

Early regulatory module *E* contains two functionally distinct sub-modules

The regulatory activity of the early regulatory module *E* is contained within a 597bp minimal sequence for which we call element *ME* (Figure 3D). The boundaries of this *ME* element were established via progressive trimming and retesting in vivo. Briefly, the entire 5kb region of module *E* was first dissected into smaller fragments that were in turn sub-sectioned in search of the minimal sequence with the regulatory capacity necessary and sufficient to recapitulate that of the module *E* as a whole (Figure S2A, B). The regulatory activity of all of these truncated fragments was measured simultaneously with the abovementioned high-throughput tag reporter system. The identified *ME* element was found to effectively recapitulate quantitatively and spatially the expression pattern of the complete sequence of module *E*, which suggests we did not lose regulatory information by minimizing the size of the module (Figure 3A, B). Deletion of *ME* element from the BAC reporter abolished the reporter activity proving its necessity for driving the early expression (Figure 3C).

Further functional dissection revealed sub-modularization within the sequence of *ME*. Deleting 5' end sequences gradually reduced the regulatory activity of *ME* at 15hpf, but not at 22hpf. The opposite effect was observed when base pairs were removed from the 3' end of *ME* (Figure S2C). These data suggest that the

regulatory information needed for *hox11/13b* activation at 15hpf and 22hpf is encoded separately at the two ends of this *ME* sequence. To further analyze the nature of this sub-modularization, we characterized the behavior of *gfp* reporter constructs driven by either the 5' end sequence (*ME5*) or the 3' end region (*ME3*). The temporal kinetics of the regulatory activity of *ME3* and *ME5* were established by measuring *gfp* transcript levels over time. The profile of *ME*-driven *gfp* transcripts levels was found to be biphasic. In the first phase between 9hpf and 18hpf, transcript levels climb to an initial peak at 15hpf followed by a moderate 30% decline at 18hpf (Figure 3E). In the second phase from 19hpf to 30hpf, the transcript levels rise again to a second peak at 22hpf and decline sharply by about 80% at 30hpf (Figure 3E). We found that *gfp* expression promoted by *ME5* is significantly elevated in the first phase but is almost undetectable in the second phase, whereas *ME3* induces much lower *gfp* expression in the first phase, but with similar expression levels and kinetics as those of *ME* in the second phase (Figure 3E). To examine how these regulatory modules control spatial gene expression, we performed dMWISH using probes specific for *gfp* and *foxa* transcripts or directly examined GFP fluorescence expression in embryos injected with either *ME3:gfp* or *ME5:gfp*. We found that *ME3* is sufficient to drive dynamic and endoderm-specific *gfp* expression at 24h and 30h, consistent with the expression pattern of endogenous *hox11/13b* (Figure 3F, Figure S1E). Unexpectedly, we found that *ME5* regulates ubiquitous spatial expression as it promotes detectable GFP expression in almost every embryonic cell lineage (Figure 3F, Figure S1F).

These observations suggest that there is a functional sub-division within the early module. The sequence on the 5' end of *ME* works as a “booster” module that ensures sufficient *hox11/13b* levels in the first phase of its expression; it does not contribute to *hox11/13b* expression in the second phase, nor does it regulate spatial gene expression. The sequence on the 3' end of *ME* regulates *hox11/13b* activation in the second phase, and more importantly it controls this spatially confined pattern of *hox11/13b* expression. In order to orchestrate the tightly regulated expression of *hox11/13b*, these two sub-modules of *ME* need to cooperate in a spatial and temporal manner.

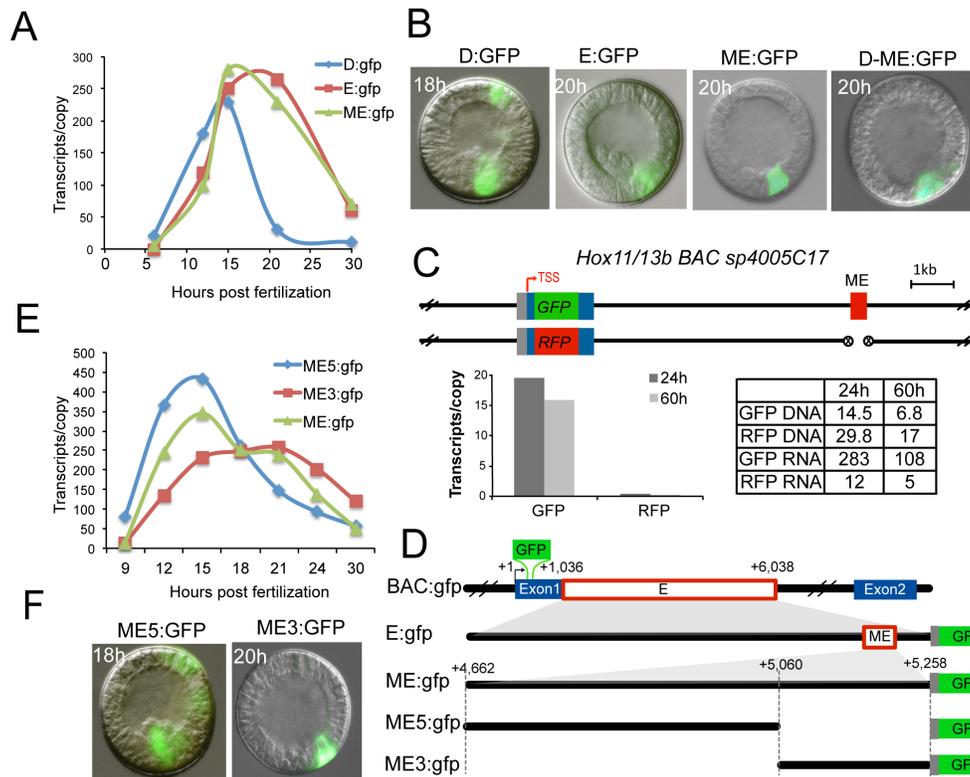


Figure 3: characterization of spatial and temporal regulatory behavior of the early active modules. (A) Temporal expression of reporter constructs driven by module D, E and ME showing that module D does not activate after 20h and module E and ME have comparable regulatory activity. (B) Spatial expression of reporter constructs driven by module D, E, ME and D-ME. Ectopic expression was observed for module D, and the anterior endoderm-specific expression was shown for module E, ME as well as module D-ME. (C) The necessity of module ME for the early expression proven by the lost expression of ME mutant BAC. A significant decrease of the RNA level, but not the incorporated DNA level, was observed for the ME-mutant RFP BAC in comparing to the wild-type GFP BAC. (D) A map showing position of module E, module ME, module ME5 and module ME3. (E) Time course expression of module ME, ME5 and ME3 showing different temporal expressional dynamics. (F) Ectopic expression was observed for ME5, whereas the endoderm-specific expression was shown for ME3. All embryos are shown in lateral views.

Dual functionality of TCF is responsible for the specific spatial pattern of *hox11/13b* expression

To clarify the DNA sequences responsible for the spatial control of *hox11/13b* expression, pieces of *ME3* were amplified and fused to *ME5* in attempt to reverse its ubiquitous expression at 15hpf. We found that addition of 40bp to *ME5* was sufficient to constrain its expression to *veg2*-specific cells (Figure 4A). Deletion of this 40bp from *ME* consequently reverted gene expression to an ectopic pattern (Figure 4A, Figure S3B), confirming that this is the only sequence necessary to coordinate this spatial regulation. Further experiments indicated that this spatial regulatory information is encoded by two TCF (T cell factor) sites within this 40bp sequence of *ME3*. Mutation of both TCF sites caused the mutated *ME:gfp* construct to be expressed ubiquitously in every cell of the embryo (Figure 4C). Individually mutating either TCF site did not change the spatial expression pattern of *gfp* (Figure S3B), suggesting that these two TCF sites have redundant regulatory activities. This *ME:gfp* construct with double TCF site mutations also drives ectopic *gfp* expression at 24hpf, but with a more confined overall pattern. Careful analysis using dMWISH to target *foxa* showed that this ectopic expression encompasses the entire *veg1* lineage including the *veg1* endoderm and the *veg1* ectoderm, whereas at this time the endogenous *hox11/13b* is only expressed in the *veg1* endoderm (Figure 4C). Thus, two TCF sites located at the 3' end of *ME* are required to promote the correct spatial expression of *hox11/13b*, and mutating these modules causes ectopic gene expression at 15hpf and 24hpf.

It is known that TCF can act as a toggle switch to control gene expression. They can function as activators in the presence of nuclear β -catenin, or as repressors in combination with Groucho (Kim, Oda et al. 2000, Range, Venuti et al. 2005). In the cells containing nuclear β -catenin, TCF acts to promote the expression of target genes such as *foxa*, *blimp1b*, *eve*, and *hox11/13b*, while it represses these same genes in other cells (Smith and Davidson 2008, Smith, Kraemer et al. 2008, de-Leon and Davidson 2010). Thus, the activation function of TCF strictly relies on the availability of nuclear β -catenin. β -catenin can be stabilized and undergo nuclear localization through two pathways, one utilizing a maternal cytoplasmic system that involves phosphorylated Disheveled protein and one involving the canonical Wnt signaling pathway (Logan, Miller et al. 1999, Cui et al. 2015). In the mid-blastula stage of development for sea urchin embryos (15hpf), nuclear β -catenin is detected in the *veg2* progeny including *veg2* mesoderm and *veg2* endoderm, likely through this maternal pathway, and at this time *hox11/13b* expression is accordingly restricted to these cells (Logan, Miller et al. 1999). During the late blastula stage (18hpf) nuclear β -catenin clears from the *veg2* mesoderm, potentially through indirect interaction with Delta-Notch signaling (Peter and Davidson 2011, Sethi AJ 2012), and remains only in the *veg2* endoderm. The clearance of nuclear β -catenin in the *veg2* mesoderm leads to the silencing of endodermal genes including *foxa* due to TCF-Groucho repression (de-Leon and Davidson 2010). Expression of *hox11/13b* is also cleared from the *veg2* mesoderm at this time point, likely through the same mechanism. At the mesenchyme blastula

stage (>22hpf), detectable β -catenin is absent in *veg2* descendants and is localized to the nuclei of a subset of *veg1* cells that are predicted to become endoderm, corresponding to the expression pattern of *hox11/13b*. Our previous work has shown that *Wnt1* and *Wnt16* are required for the activation of *hox11/13b* in the *veg1* endoderm (Cui et al. 2015). The discovery of functional TCF sites in *hox11/13b* regulatory module suggests that this activation is likely through the canonical Wnt pathway, which utilizes TCF/ β -catenin as a nuclear mediator of gene expression.

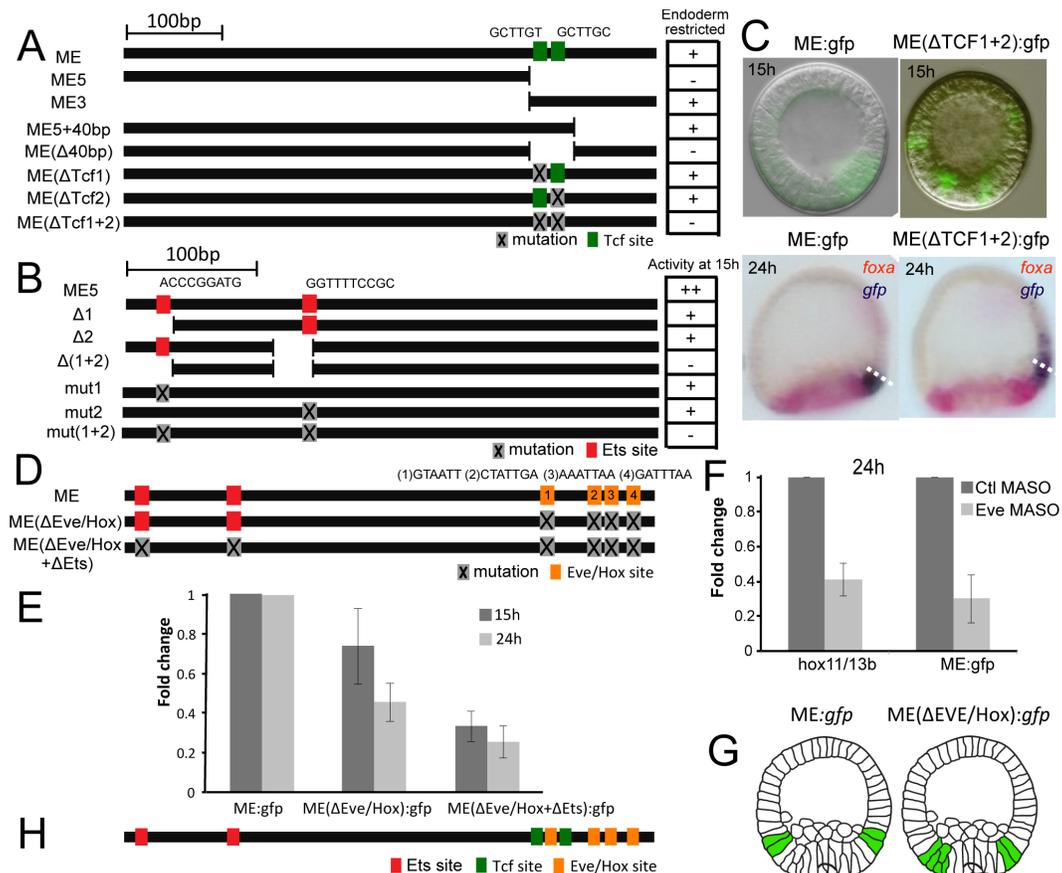


Figure 4: Discovery of functional binding sites within the early module ME.

(A) Deletions and mutations of the ME3 region on module ME discovered two functionally Tcf sites (green bars) required for restricting the expression to

endoderm. The spatial expression of the GFP reporter of various constructs was examined at 24h by fluorescent microscope. Constructs with endoderm-specific expression is marked “+”; constructs with ectopic expression is marked “-”. A complete deletion series to discover the sequences responsible for the spatial restriction is shown in Figure S3A. (B) Deletions and mutations of module ME5 identified two functional Ets binding sites shown in red vertical bars. The regulatory activity of the deleted and mutant constructs was measured at 15h and compared to that of the wild-type construct, whose level is quantified as “+++”; 1/3 reduction from the wild-type activity is marked as “+”, and 2/3 reduction is shown in “-”. The complete measurement can be found in Figure S4B. (C) The ectopic GFP expression driven by Tcf-sites-mutant ME construct is ubiquitous at 15h and more narrowed but expanded anteriorly into the vegetal ectoderm at 24h. The white dashed lines indicate the boundary between endoderm and ectoderm. *Foxa* staining is used to mark the anterior endoderm. (D, E) Mutation of Eve/Hox sites (orange bars) decreased the expression of ME reporter construct at 24h. Further reduction at 24h as well as 15h was observed when both Ets and Eve/Hox sites were mutated. The histogram shows quantitative PCR measurement of reporter activity of mutant constructs normalized to that of the wild-type ME construct. (F) Expression of *hox11/13b* and ME reporter construct in posterior endoderm at 24h depends on *eve* expression. (G) Schematic illustration shows that mutation of Eve/Hox sites interfered with the activation of *gfp* expression in the posterior endoderm and also its clearance in the anterior endoderm. Detailed spatial analysis of this mutant construct is shown in Figure S7A, B. (H) Map of functional binding sites discovered in module ME. Sequences of binding sites are shown in Figure S5A.

ETS and EVE work sequentially to activate early *hox11/13b* expression

Because of the dual regulatory capacities of these TCF sites, deleting them not only ablated their activation capacity through β -catenin but also eliminated their repression capabilities normally mediated by Groucho. In fact, the ectopic expression pattern observed when these TCF sites were mutated in our *ME:gfp* construct suggests that additional factors exist to activate the expression of *hox11/13b*, and that those additional factors are likely to be broadly expressed.

To identify potential factors that drive ubiquitous expression of *ME5* in the first phase, we generated systematic deletion constructs as illustrated in Figure S4A. The regulatory activity of these constructs was measured at 15hpf using a GFP reporter system. We found that the deletion of two specific sequences decreased reporter activity by 2-fold, and simultaneous deletion of both of these sequences significantly reduced reporter activity by more than 3-fold (Figure 4B, Figure S4B). Both deleted DNA sequences are ~25bp in length and contain a predicted ETS family binding motif according to the JASPER database (Figure S5A). Mutation of the core consensus binding sequence of these ETS motifs reduced reporter activity to a magnitude similar to that present in samples in which the entire 25bp sequence was deleted (Figure S4B). These results suggest that an ETS family transcription factor is likely to be a direct activator of *ME5*, and that there is unlikely to be any other regulatory information encoded in the remainder of this sequence. During sea urchin embryogenesis, there are five ETS family genes expressed ubiquitously in blastula stages, and they are *Ets1/2*, *Tel*, *Erf*, *Ets4*, and *Elk* (Rizzo, Fernandez-Serra et al. 2006). With the exception of *Elk*, all of these factors exhibit abundant maternal expression, which made it difficult for us to conduct loss-of-function studies via morpholino perturbation. The transcription of *elk* precedes the activation of *hox11/13b* by 1-2 hours, and its temporal profile exhibits similar kinetics to those of *hox11/13b* in the first phase of its expression (Figure S6). Perturbation of *Elk* using a morpholino to block translation increased *elk* transcript levels by 2-fold suggesting a negative feedback loop as well as the efficacy of the morpholino;

however, it had no significant effect on the transcripts level of *hox11/13b* (Figure S6). It is possible that other maternal ETS factor(s) may compensate for the loss of Elk; however, it appears that Elk is either not required for the activation of *hox11/13b* or it shares functional redundancy with other ETS factors.

An activator controlling the second phase of *hox11/13b* expression in the veg1 endoderm should exhibit a pan-veg1 expression pattern as inferred from the ectopic expression pattern of our TCF site mutants at 24hpf (Figure 4C). None of the genes encoding ETS factors are expressed in veg1 progeny cells at this time point, suggesting a distinct factor may dictate this activation of gene expression. Prior studies have shown that morpholino perturbation of the Homeobox protein EVE decreases the expression level of *hox11/13b* at 24hpf, but with no change of expression observed at 15hpf or 18hpf (Peter and Davidson 2011). Transcripts of *eve* are present in the entire veg1 lineage at 24hpf, mirroring the ectopic expression pattern of our TCF-mutated *ME:gfp* construct. This spatial expression, coupled with these previous perturbation studies, suggests that EVE is a plausible candidate regulator of the second phase of *hox11/13b* expression. To establish whether the *ME* is under the regulatory control of EVE, we measured the expression change of *ME:gfp* in response to EVE perturbation. We observed reduced *gfp* expression at 24hpf in the embryos co-injected with EVE morpholino as compared to those co-injected with random control morpholino (Figure 4E), suggesting that EVE mediates the regulatory activity of *ME* at this time. Indeed, we identified functional EVE binding sites in the early module that are responsible for the second phase

activation of *hox11/13b*. There are four predicted EVE binding motifs located at the 3' end of *ME* (Figure S5A). Because EVE is a Homeodomain protein and three of the four potential EVE binding sites are also predicted as target sites for other Homeodomain proteins, we thus refer these sites to Eve/Hox sites. Mutation of all four Eve/Hox sites decreased the activity of our *ME:gfp* construct by ~50% at 24hpf (Figure 4D, E). The effect of these mutations at 15hpf was negligible, suggesting that these Eve/Hox sites are required for the activation only in the second phase of *hox11/13b* expression. The remaining transcript levels in our Eve/Hox site mutant reporter embryos were likely due to the residual activity of the ETS-mediated initial activation. Indeed, when we mutated both ETS and Eve/Hox binding sites, we observed a nearly complete ablation of reporter activity at 24hpf (Figure 4D, E). We have thus identified EVE as a direct activator that contributes to the second phase of *hox11/13b* expression in the veg1 endoderm.

We next examined the effect on spatial gene expression when these Eve/Hox sites were mutated in our *ME:gfp* construct. As expected, embryos injected with the mutated construct failed to activate *gfp* expression in the veg1 endoderm at 24hpf; the majority (90%) of them instead expressed *gfp* in the veg2 endoderm (Figure S7A). The *gfp* expression in the veg2 endoderm was also detected in the embryos injected with WT *ME:gfp* construct; however, this expression gradually decreases as development proceeds, as evidenced by the reduced percentage of embryos expressing *gfp* in veg2 endoderm at 30hpf as compared to 24hpf (Figure S7B) and 20hpf (data not shown). We noticed that this clearance was impaired in the embryos

injected with our Eve/Hox-mutated *ME:gfp* construct, as *gfp* transcripts persisted in the *veg2* endoderm in ~80% of embryos examined at 30hpf (Figure S7A). This suggests that mutation of Eve/Hox sites hampers the repression of *gfp* transcripts in the *veg2* endoderm. Prior studies have shown that perturbation of the Hox11/13b gene itself interferes with the clearance of its transcripts from the *veg2* endoderm (Peter and Davidson 2011). If by mutating Eve/Hox sites we also disrupted Hox11/13b binding sites, this could explain the failure of *gfp* mRNA clearance in *veg2* endoderm. Indeed, as we mentioned earlier that there are also four Hox binding sites in *ME*, and three of them partially overlap with EVE sites (Figure S5A). Due to the difficulty of separating Hox sites from EVE sites, we tested whether *ME* would spatially respond to the perturbation of Hox11/13b in order to confirm the regulatory information of Hox11/13b is actually encoded in *ME*. In the embryos co-injected with *ME:gfp* and a control morpholino, we observed clearance of *gfp* transcripts in the *veg2* endoderm of 80% of embryos at 30hpf (Figure S7C). This clearance failed to occur in >60% of embryos that were co-injected with a Hox11/13b morpholino (Figure S7C). This result therefore suggests that Hox11/13b binding to the *ME* is required for the *veg2* endoderm repression of *hox11/13b* transcripts as a negative feedback mechanism.

Our results thus far indicate that combinatorial control within the early module is required for the proper spatial and temporal early expression of *hox11/13b*. During the first phase of its expression, *hox11/13b* is expressed in the *veg2* progeny and is later restricted to the *veg2* endoderm. Its robust activation during this phase is

mediated by ubiquitously expressed ETS factor(s), and its spatial restriction is regulated by the dual functionality of TCF. During the second phase when *hox11/13b* is expressed in the veg1 endoderm, its activation relies on the pan-veg1 transcription factor EVE, and the same TCF binding sites used in the first phase now serve to restrict its expression to the veg1 endoderm. Meanwhile, Hox11/13b negatively regulates its own transcription in the veg2 endoderm, further ensuring its specific expression in the veg1 endoderm. Thus, in two phases, *hox11/13b* utilizes conserved TCF binding sites in concert with different regulators to ensure its correct spatial expression and activation. This suggests that the early expression of *hox11/13b* is operated by AND logic gate: spatial regulation + activation. This AND logic gate ensures regulatory specificity as well as flexibility by coupling with different activators to achieve dynamic expressional pattern.

A distal regulatory module *D* contributes to robust *hox11/13b* activation

Module *D* was another fragment that showed regulatory activity in our initial screen. Temporal and spatial characterization of a reporter construct (*D*:GFP) showed that its sequence possesses similar regulatory functionality as the *ME5* region of module *E*, as the *gfp* expression of *D:gfp* construct showed a similar temporal and spatial pattern to that of *ME5:gfp* construct (Figure 3A, B).

Module *D* is located about 62kb downstream from the TSS of *hox11/13b*. In fact, its sequence is only 5kb downstream of the last exon of the neighboring *hox11/13c* gene. However, *hox11/13c* is not expressed consistently during embryogenesis (Arenas-Mena, Martinez et al. 1998, Arenas-Mena, Cameron et al. 2000). Moreover, no other genes within 300kb of module *D* are transcribed during this time frame post-fertilization. Therefore, the regulatory function of module *D* is most likely associated with the activation of *hox11/13b*, serving as a distal regulatory module. When we previously characterized the regulatory behavior of module *E*, we noticed a reduction of its regulatory activity at 15hpf in comparison with our *hox11/13b* BAC (Figure 2B). We reasoned that this reduction could be due to lack of regulatory information encoded in this distal module. To demonstrate that module *D* directly contributes to the activation of *hox11/13b* expression, we made a *D-ME:gfp* reporter construct containing the fused sequences of *D* and *ME* upstream of a *gfp* reporter (Figure S8A). In the embryos injected with this fusion construct, we observed a ~40% increase of *gfp* expression at 15hpf comparing to the embryos injected with *ME:gfp* (Figure S8A). Importantly, the addition of module *D* to the fusion construct did not alter the specific expression pattern of *D-ME:gfp*. The fusion construct maintained proper spatial restriction, with GFP specifically expressed in the same expression pattern as we observed for *ME:gfp* (Figure 3B, Figure S8B). These observations suggest that, although module *D* doesn't contribute to the spatial regulation, it is required to achieve robust activation of *hox11/13b* during the first phase of its expression.

Late expression in the hindgut requires two separate cis-regulatory modules

To our surprise, in our initial screen none of the 51 dissected fragments were found to be active at later stages of embryogenesis (42hpf and 60hpf), even though the *hox11/13b:gfp* BAC is capable of driving reporter expression during this time frame. In order to clarify the cis-regulatory sequence responsible for this late gene expression, we took a different approach in which we examined the late expression of multiple versions of BAC reporter constructs. Three different *hox11/13b:gfp* BACs were constructed and injected into sea urchin embryos individually to study patterns of reporter activity. These three *hox11/13b* BACs span different regions across the *hox11/13 a-c* cluster and are largely non-overlapping with each other, but all three of them contain the first exon and part of the intron of *hox11/13b* (Figure 5A). We found that all three BACs were able to drive both early and late expression of the GFP reporter, suggesting that the late regulatory module is located within their shared sequence. This shared sequence is ~30kb in length, covering 12kb upstream and 18kb downstream from the TSS of *hox11/13b*. We next amplified fragments of this 30kb region and examined their regulatory activity, which allowed us to identify the late regulatory module within the intronic region. As shown in Figure 5A, an 11kb fragment encompassing ~400bp upstream, the exon1-GFP coding region, and an 8kb intron produced accurate spatial reporter expression at 60hpf (Figure S9). Progressive trimming from the 3' end of this 11kb fragment further confined the late module (*L*) to a ~700bp DNA sequence located 700bp

downstream of the region *ME*. The initial screen using high-throughput reporter constructs failed to identify this late module *L*, which suggests that it alone is not sufficient to drive the late gene expression. Indeed, a reporter driven by module *L* alone did not show any activity (data not shown). Furthermore, deletion of *ME* from the *hox11/13b:gfp* BAC not only abolishes the early but also the late expression of GFP reporter (Figure 3C), suggesting that *ME* is also required for this late activity. Thus, the expression of *hox11/13b* in the hindgut at later stages of development requires both the early and late cis-regulatory modules: both modules are necessary, but neither is sufficient.

Regulatory module E and L are both in the intronic region and are only 700bp apart. To exclude the possibility that these two modules are a single functional unit, we performed ATAC-seq analysis, which measures chromatin accessibility (Buenrostro, Giresi et al. 2013, Buenrostro, Wu et al. 2015), in the embryos at different developmental stages —18hpf, 24hpf, and 60hpf. We determined that the sequences of these early and late modules exhibit distinct and dynamic chromatin accessibility that mirrors their regulatory activity. Module E is highly accessible at 18hpf and 24hpf, but becomes less accessible at 60hpf, whereas the opposite pattern is present for module L which contains opened chromatin structures at 60hpf but not 18hpf or 24hpf (Figure 5A). Importantly, the intermediate sequence between these two modules has no mapped ATAC-seq signal at any time examined, suggesting that this is a closed chromatin region that separates the early and late regulatory modules. Further evidence supporting that these two modules are indeed

separate regulatory modules comes from a sequence conservation analysis, in which the sequences of *hox11/13b* homologs in *S. purpuratus* and *L. variegatus*, two related sea urchin species that are 30 to 50 million years evolutionarily divergent (Smith, Pisani et al. 2006), were aligned for conservation comparison. As shown in Figure 5A, module *E* and *L* are conserved between species, whereas the intermediate region is not. This suggests that the intermediate region has undergone unconstrained evolutionary selection and is unlikely to be a part of a functionally conserved regulatory module. Taken together, the distinct chromatin accessibility dynamics of the early and late modules together with their sequence conservation suggests that they are two separate regulatory entities.

The 400bp TSS-proximal region (upstream of the promoter) contained within the 11kb fragment is also required for both the early and late stages of *hox11/13b* activation. Deletion of this proximal sequence reduced the *gfp* transcripts level by 2-fold at 24hpf and 3-fold at 48hpf, to a level similar to that of the negative control construct (Figure 5A, Figure S10). This 400bp region was encompassed within a larger fragment included in our initial reporter screen; however, that fragment did not show any regulatory activity at the time points examined. This is seemingly in contrast with the fact that it is clearly needed for *hox11/13b* expression. It is known that CRMs, such as insulators or barriers, can contribute to gene expression by regulating chromosomal interactions (Geyer 1997, Lin, Li et al. 2011). They may, for example, mediate enhancer-promoter looping events or the separation of distinct chromosomal territories (Matharu and Ahituv 2015). We thus chose to explore the

possibility that this proximal region may contribute to *hox11/13b* expression by serving as an anchor that facilitates looping interactions between both regulatory modules and the basal promoter region of this gene. Among many mechanisms of DNA looping, Zinc-finger factor CTCF is best characterized to bind together strands of DNA through a mechanism involving homodimerization, thus forming chromatin loops (Splinter, Heath et al. 2006, Guo, Xu et al. 2015). Indeed, we found one well-scored CTCF binding site within this proximal region that is oriented in a forward direction towards the transcriptional start site of this gene (Figure 5A). We also identified an additional CTCF site 250bp upstream from this first site that also possesses a forward orientation (Figure 5A). The discovery of this second CTCF site outside of this region may explain why the *hox11/13:gfp* BAC with the 400bp region deletion did not reduce *gfp* expression (data not shown). The sequence containing both of these CTCF sites is conserved between two sea urchin species (Figure 5A), suggesting that its regulatory function is evolutionarily selected.

We next performed 4C-seq assay in order to test the hypothesis in which the gene proximal region contributes to the expression of *hox11/13b* by facilitating the interaction between cis-regulatory modules and the promoter through a looping mechanism. We designed our bait for 4C-seq analysis to target this proximal region and observed signals mapped to both module ME and module L at 15h and 48h (Figure 5B). This result clearly shows that there is a chromosomal interaction between two cis-regulatory modules and the gene proximal region in the genome.

More interestingly, the same chromosomal interactions at the gene proximal region persist across developmental stages between the early and late expression of *hox11/13b*, so that the late module is also interacting with the proximal region at 15h when it is clearly not functional. This lack of dynamic changes of chromosomal 3D structure at the *hox11/13* locus is consistent with previous observation in cell lines using Hi-C. Together, we show the late expression of *hox11/13b* in the hindgut requires two separate cis-regulatory modules: ME and L. These two modules possess dynamic chromatin accessibility at different developmental stages, but nevertheless both form looping interactions with the gene proximal region during the early and late expression of *hox11/13b*.

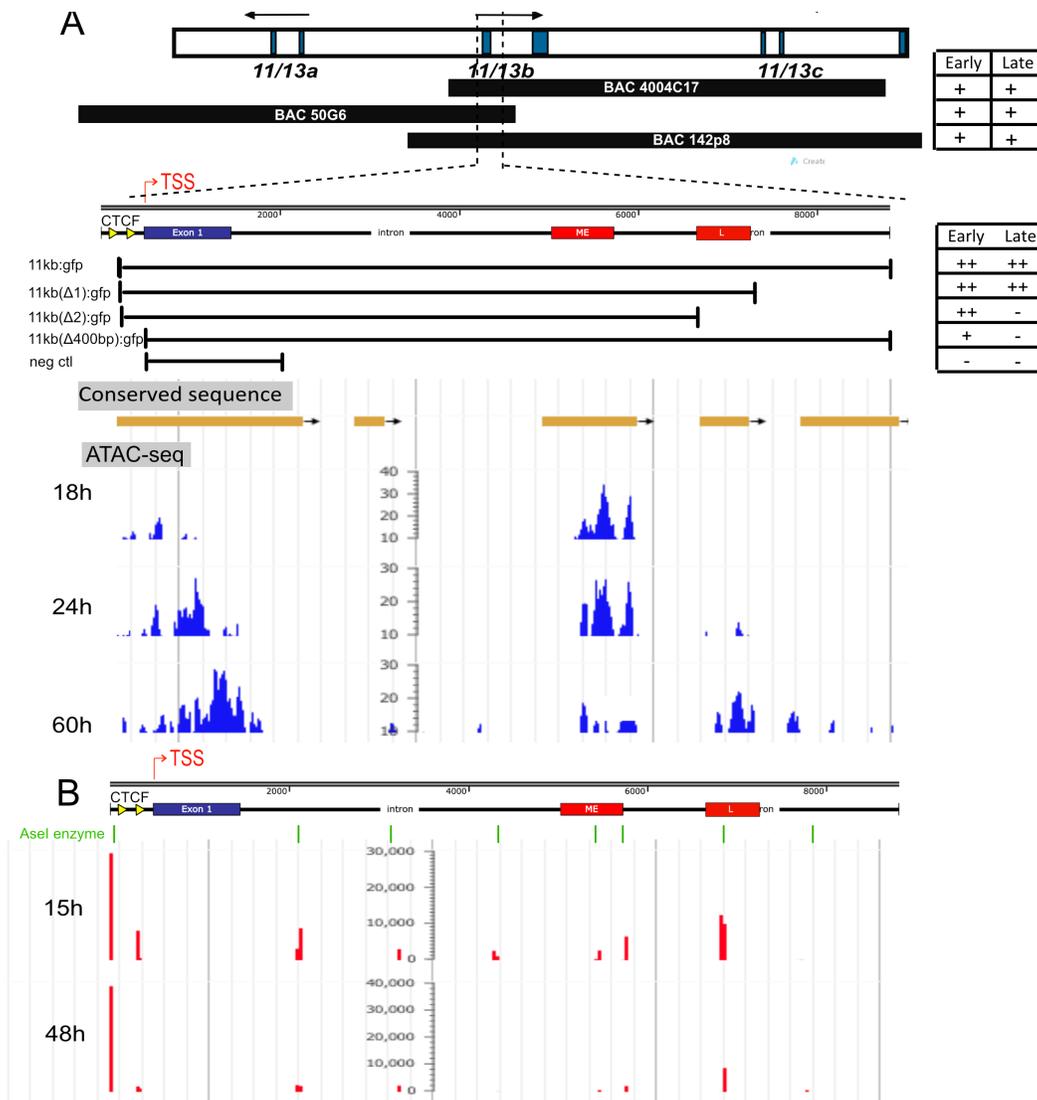


Figure 5: Identification of a second intronic module (L) required for the late expression of *hox11/13b* in the hindgut. (A) Map of BAC clones used for narrowing the late-module-containing region; their capability of driving early and late expression is shown on the right. Deletion on the 11kb region revealed a second intronic sequence (module L) and a gene proximal region containing two CTCF sites (yellow triangles) that are required for the late expression. The regulatory activity of deleted reporter constructs are shown on the right; same symbols as in Figure 4B are used to show the magnitude of expressional changes. Quantitative PCR measurements of expression level of these constructs are shown in Figure S10. A snapshot of genome browser shows ATAC-seq tracks (18h, 24h and 60h) and conserved regions with species *L. variegatus* around the region near the first exon including part of the intron. (B) 4C-seq tracks at 15h and 48h are visualized on genome browser with the bait region designed around the proximal

region. Restriction enzyme cutting sites for AseI used for the primary digestion are shown green lines.

Identification of binding sites required for late *hox11/13b* expression in the hindgut

We next sought to identify the binding sites driving late *hox11/13b* expression in the hindgut. To do so, we generated a construct in which we took a 1.5kb sequence containing both module *E* and *L* out of its genomic context and placed it in front of a 1kb TSS-proximal sequence that contains the two CTCF sites, followed by the *gfp* coding region (Figure 6A). This reporter construct configuration recapitulated the late expression pattern of *hox11/13b* in the hindgut to the similar level as the 11-kb construct (Figure 6D, Figure S9), and was subsequently subjected to sequence mutation analysis. We found that mutation of ETS and Eve/Hox sites located in the early module E not only decreased the early *gfp* expression at 18hpf and 24hpf but also the late expression at 48hpf (Figure 6B). Further mutations of either the ETS or Eve/Hox sites showed that the Eve/Hox sites, but not the ETS sites, are required for the late *gfp* expression, as we observed a ~60% reduction of *gfp* transcripts in the Eve/Hox sites mutant construct but not the ETS sites mutant construct (Figure 6B). These Eve/Hox sites in the early module could be required to interact with some homeo-domain factors, which are yet to be identified, for the late gene expression.

Functional sites within module L have also been identified. Sequence deletion analysis conducted on the late regulatory module *L* identified a 142bp sequence that is essential for regulating late gene expression and that overlaps with our ATAC-seq signal reads. Deletion of this sequence diminished *gfp* expression at 48hpf (Figure 6C). Three predicted binding sites for Homeobox proteins reside in this sequence (Figure S5B). Simultaneous mutation of all three of these sites also decreased *gfp* expression (Figure 6C), suggesting that these sequences possess regulatory activity. In summary, we discovered that the late expression of *hox11/13b* in the hindgut requires Eve/Hox binding sites in the early module *E* as well as Homeobox binding sites in the late module *L*. Ongoing studies characterizing the regulatory states formed in sea urchin gut morphogenesis will uncover potential late regulators of *hox11/13b* expression based on spatial correlations, eventually leading to the discovery of the late regulatory inputs through the use of trans-perturbation analyses.

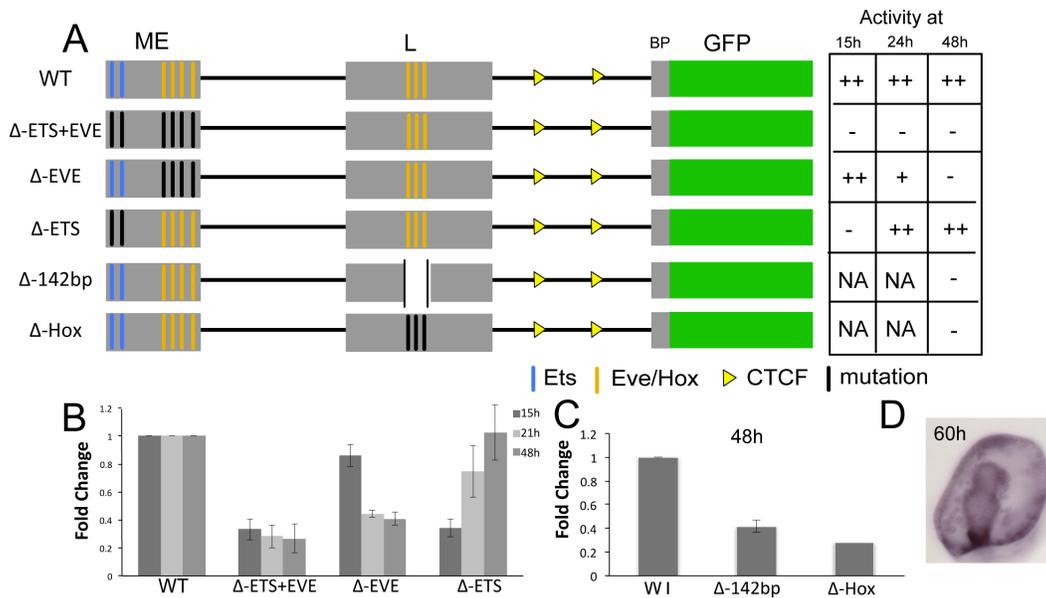


Figure 6: Discovery of regulatory binding sites required for the late expression of *hox11/13b* in the hindgut. The region containing module ME, module L and their intermediate sequence was placed in front of the 700bp gene proximal region and fused into a GFP reporter construct. The reporter activity of this configuration (ME-L:GFP) and other deleted and mutated forms, illustrated in A, was measured by QPCR at 15h, 24h, and 48h, shown in B and C. The hindgut specific expression of ME-L:GFP is shown in D.

DISCUSSION

In this study we have identified and characterized the genomic regulatory information that mediates the dynamic and endoderm-specific expression of *hox11/13b* during the first two days of sea urchin embryogenesis. This regulatory system is encoded in three temporally and spatially distinct modules: the intronic early module *E*, the distal module *D*, and the intronic late module *L*. In order to

coordinate the proper expression pattern of *hox11/13b*, these three cis-regulatory modules must cooperate in distinct spatial domains and at appropriate times to execute specific regulatory operations, as summarized in Figure 7. Below we discuss the inputs, functions, and interactions of these modules during each stage of *hox11/13b* expression.

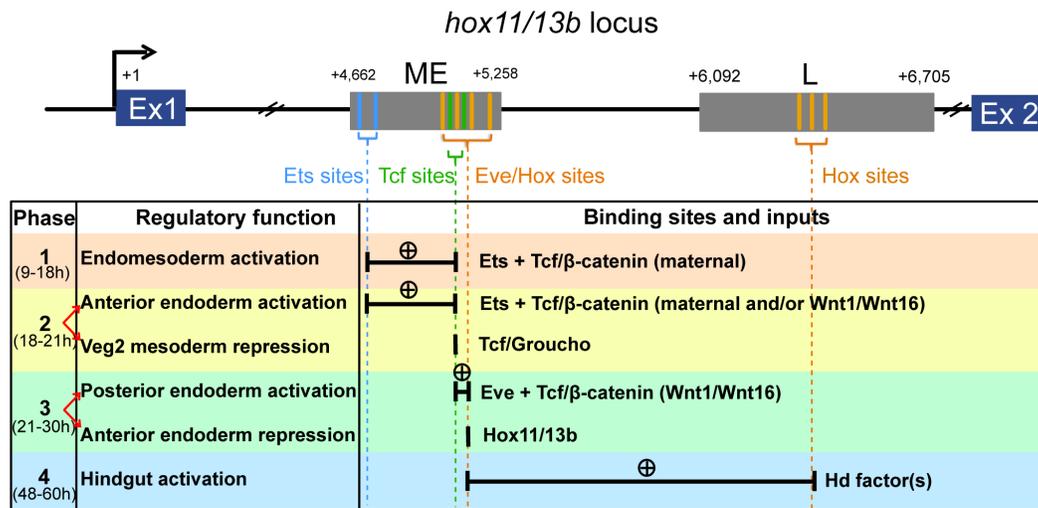


Figure 7: Summary of the combinatory usage of binding sites and their corresponding TF inputs at each phase of *hox11/13b* expression. The binding sites discovered in this study are shown at the *hox11/13b* locus; same functional sites are clustered into color-coded dashed lines. Sites being used at each phase to conduct a particular regulatory function are marked as black lines and are connected by black horizontal lines with other sites, if they exist, that are required for the same regulatory task. “ \oplus ” indicates an AND logic between two connecting sites. The TF inputs are shown after the corresponding sites. The source of nuclear β -catenin is specified based on study (Cui et al 2015).

The activation in veg2 lineages (9-20hpf): Although module *E* is sufficient to mediate proper spatial regulation during this time, module *D* is absolutely required for robust quantitative control of *hox11/13b* expression. Detailed CRA within module *E* has discovered distinct mechanisms underlying this temporal and spatial

regulation. At this time the ubiquitously expressed ETS proteins contribute to the prompt temporal activation, whereas a TCF-Groucho/ β -catenin switch mediates the spatial restriction. TCF likely couples with maternal nuclear β -catenin during this time to form a permissive complex that allows for the transcription of *hox11/13b* in these cells, while forming a repressive complex with Groucho in other cells. Module *D* only contributes to the overall *hox11/13b* expression level, and does not seem to possess spatial regulatory capacity. We did not specifically investigate its regulatory inputs, however it is also likely to be mediated by ETS family TFs, as it contains predicted ETS factor consensus binding sites (data not shown).

Veg2 mesoderm repression (18h-20h): This regulatory function is controlled by the Tcf sites in module *E*. The underlying molecular mechanism is one whereby β -catenin is cleared from the nuclei of cells in this territory, allowing for the establishment of Tcf-Groucho repression. Delta-Notch signaling has been reported to restrict *hox11/13b* expression in the cells that will go on to form the veg2 mesoderm (Sethi AJ 2012), but we were unable to identify any binding site for its nuclear mediator, Suppressor of Hairless, in the intronic early module. This suggests that this may be an indirect effect, likely stemming from interactions with the Tcf toggle-switch mechanism.

Veg1 endoderm activation and establishment of the endo/ecto boundary (21h-30h): The same TCF sites that regulate early spatial expression in veg2 lineages here are

used again to restrict *hox11/13b* expression to the veg1 endoderm. The activation function is however encoded in the Eve binding sites that reside within this same module. Transcripts of *eve* can be detected in cells of the veg1 lineage as early as 12hpf, although Eve proteins only activate *hox11/13b* expression in these cells nearly 10 hours later. This can be explained by the requirement for the permissive TCF/ β -catenin complex. During the mesenchyme blastula stages of about 21-22 hours post fertilization, nuclear β -catenin is located in veg1 endodermal cells and not in the adjacent veg1 ectodermal cells, likely as a result of the canonical signaling pathway mediated by Wnt1 and Wnt16. This asymmetric distribution of nuclear β -catenin within veg1 cells ensures the endoderm-specific expression of *hox11/13b*. Hox11/13b initiates the veg1 endodermal GRN and prevents ectodermal genes, such as *hox7*, *veg3*, and *lim1*, from being ectopically expressed in the endoderm, thus forming a boundary between endoderm and ectoderm (Li, Cui et al. 2014). The establishment of the endoderm/ectoderm boundary is thus a progressive process that first requires asymmetric *hox11/13b* activation in the inner veg1 cells. This activation in turn initiates the veg1 endodermal GRN in those cells and enforces a defensive mechanism that prevents ectodermal genes from being aberrantly expressed in the endoderm.

Veg2 endoderm clearance (21-30h): At approximately the same time as *hox11/13b* is transcribed in the veg1 endoderm, its mRNA is cleared in the veg2 endoderm. Hox11/13b represses its own transcription by binding to the Homeobox binding sites contained within module *E*, thereby contributing to its clearance via a negative

feedback loop.

Hindgut activation (48-60h): Module *E* and module *L* work in a dependent manner in order to activate the late *hox11/13b* expression during this time. Eve/Hox binding sites contained within module *E* are required for this late activation. At present we cannot distinguish between the possibility that these sites are required for the binding of late TFs or that they are solely associated with the early activation and instead contribute to the late expression through epigenetic regulation. It is possible that module *E* contributes to this late regulatory function through multiple mechanisms. Early activation may create a permissive chromatin state that facilitates the late expression through, for example, an enhancer-priming mechanism that induces the chromatin decomposition necessary for later expression. Alternatively, module *E* may be required as a binding platform for unidentified late regulators. It is also possible that module *E* may serve as a tethering system that mediates an interaction between module *L* and the transcriptional apparatus present at the promoter region. While our study did not focus on clarifying the regulatory inputs of module *L*, we have provided the sequence basis and outlined the regulatory logic necessary to conduct such studies.

A combination of broadly expressed transcriptional activators and TCF-mediated spatial restriction governs the dynamic and endoderm-specific early expression of *hox11/13b* in two ways. First, this AND logic gate ensures regulatory specificity, as

only cells containing both the TCF- β -catenin complex and the appropriate TF activators are able to initiate transcription. Second, this mechanism by which a transcriptional activator can be replaced in different embryonic territories and at different times offers flexibility, allowing *hox11/13b* to be associated with multiple distinct transcriptional drivers and to be rewired into different lineage-specifying GRNs. This flexibility is advantageous, allowing *hox11/13b* to gain a new spatial expression domain, thereby permitting the occurrence of co-options of the downstream GRN circuitry. This regulatory principle is unlikely to be unique to the *hox11/13b* gene. During early sea urchin embryogenesis, many endodermal genes exhibit similarly dynamic expression patterns such as that of an expanding torus in the cells of vegetal lineages. Some of these genes have been already shown to directly respond to the canonical Wnt signaling pathway.

It has been previously demonstrated in sea urchin embryos and in other organisms that CRMs work collaboratively to mediate gene expression (Dunipace, Ozdemir et al. 2011, Cochella and Hobert 2012, Barsi and Davidson 2016). This means that CRMs cooperate in a coordinated fashion, contributing to a given phase of developmental expression rather than generating completely distinct phases. Focus on this regulatory cooperation among CRMs serves as a new approach to studying underlying mechanisms of transcriptional regulation. In this study we have clarified a unique case, in which the activation of one CRM is dependent that of another. The discovery of this rarely observed dependent-CRM cooperation, which would be invisible to conventional analyses using isolated short expression constructs, was

possible due to our use of engineered BACs that served as a genomic platform for functional CRA. Our study thus emphasizes the importance of using genomic context as a sequence basis to study the regulatory control of gene expression. This potentially will be promoted as a more generalized approach to CRA with the recent rapid advances in genome editing technologies, such as TALEN and CRISPR/cas9 (Gaj, Gersbach et al. 2013). In summary, our data identifies a new regulatory control paradigm governing gene expression, and lends support to the view that cooperation among CRMs may be a more common and more sophisticated mechanism than previously appreciated.

METHODS

Generation of BAC reporter and deletion constructs by homologous recombination

A 142-kb BAC clone Sp_4005C17 was identified from a *S. purpuratus* genomic DNA library from the sea urchin genome resource (<http://www.echinobase.org/Echinobase/JBrowse>). This clone contains the entire *hox11/13b* coding sequence and 50kb upstream as well as 80kb downstream sequence from the transcription start site. BAC 50G8 and 142p8 were sequenced previously (Cameron, Rowen et al. 2006) and contain at least the first exon and a partial intron of *hox11/13b*. The protocol utilized for the insertion of GFP or RFP marker genes and for deletion of cis-regulatory sequences were adapted directly from published procedures where uses a re-engineered λ phage as a source of recombinase (Lee, Yu et al. 2001, Holmes 2015). For the GFP and RFP inserted BACs, a 129bp of sequence immediately after the start codon is replaced by a cassette containing the fluorochrome coding sequence, an SV40 poly-adenylation site, and the kanamycin gene flanked by FRT sites. After recombination, the kanamycin resistance gene was removed by the induction of the FLP-recombinase, leaving one copy of 45 bp FRT site downstream of the SV40 3'UTR. Similar experimental design was applied for generating deletion on BAC reporters, in which the cis-regulatory sequence was replaced by kanamycin gene flanked by Loxp sites. After recombination, the kanamycin resistance gene was removed,

leaving one copy of 34 bp loxP site which replaces the deleted cis-regulatory sequence.

Generation of cis-regulatory reporter constructs and mutation of putative transcription factor binding sites within reporter constructs

Cis-regulatory reporter constructs were generated by fusing a putative regulatory sequence to *hox11/13b* basal promoter and into individual “barcoded” GFP vectors that were previously developed for high-throughput cis-regulatory analysis (Nam, Dong et al. 2010, Nam and Davidson 2012) using Gibson assembly kit (New England Biolabs, E2611L). Site-specific mutation of reporter constructs was also achieved through Gibson assembly of the synthesized double-strand DNA (IDT) containing the desired mutation or deletion. The general role of disrupting putative binding sites is to exchange “A” with “C” and “G” with “T”, with exceptions when new sites were created by such change. Sanger sequencing was always performed on mutated reporter constructs to confirm the desired sequence changes.

Gene transfer and MASO injection

Sea urchin eggs and sperm were isolated and prepared for injection as described in (Cheers and Ettensohn 2004). For transferring small constructs and linearized BAC reporter into eggs, injection solution contained 120mM KCl, 20ng/ul of carrier DNA, and DNA constructs at a concentration of 100 molecules/pl. In barcoded

GFP reporter experiments, multiple DNA constructs were mixed at equal molar ratio. Injection volume per egg was approximately 10-20 μ l. For delivering MASO into eggs, injection solution contained 120mM KCL and 300 μ M MASO. When co-injecting MASO and reporter construct into eggs, a concentration of 200 μ M for MASO and 50 molecule/ μ l for reporter constructs was used. The sequence for *hox11/13b* and *eve* MASO has been reported previously (Peter and Davidson 2010). Randomized control MASOs (N_{25}) were injected at the same concentration.

Quantitative and spatial measurement of regulatory activity of reporter constructs

Barcoded cis-regulatory reporter construct were mixed into the same injection solution, and injected in fertilized eggs. Embryos were cultured at 15°C. Approximately 300 injected embryos were collected at each time point. Total RNA and the genomic DNA were extracted using the AllPrep DNA/RNA Micro Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. cDNA was synthesized using iScript cDNA synthesis kit (BioRad). GFP "barcode" sequence tags can be detected independently using specific QPCR primers (Nam, Dong et al. 2010) in Power SYBR Green Master Mix (Life Technology). QPCR was used to measure reporter activity of each tag construct quantitatively, and the results were normalized to the number of integrated genomic copies of that tag as described previously (Revilla-i-Domingo, Minokawa et al. 2004). Experiments were

performed on 3-5 independent embryonic batches. An empty GFP construct with only *hox11/13b* basal promoter and GFP reporter was used as a negative control to set a baseline for all expression data.

The spatial activity of reporter constructs was measured by direct visualization of GFP fluorescence using an Axioskop 2 plus (Zeiss) compound microscope or by whole mount in situ hybridization (WMISH) following a standard protocol as described in (Minokawa, Wikramanayake et al. 2005). For lineage scoring purpose, WMISH was performed using anti-GFP Digoxigenin (DIG)-labeled probe on a pool of injected embryos. Cells that are positive for the staining signal were scored by the origin of lineages to get a statistical lineage distribution for the GFP positive cells. Two-color double WMISH was performed using anti-GFP Dinitrophenol (DNP)-labeled probe and anti-*foxa* DIG-labeled probe. DNP-labeled probe was detected with nitro-blue tetrazolium/5-bromo-4-chloro-3'-indolyphosphate (NBT/BCIP)-staining solution, and a second stain was performed on DIG-labeled probe using Fast Red TR/Naphthol AS-MX Tablets (Sigma). The sequences of primers used to generate GFP probes are: 5'-AGCAAGGGCGAGGAACTG-3' (forward primer) and 5'-CAGCTCGTCCATGCCATGTG-3' (reverse primers). Primers sequences for *hox11/13b* and *foxa* probes are acquired from previous study (Peter and Davidson 2010).

ATAC-seq

Embryos at stage 15h, 24h, and 60h were harvested by centrifugation at 1400rpm for 5min in 4 °C. Single cell suspension was prepared by re-suspending embryos in ice-cold dissociation buffer containing 1M Glycine pH 8, 4mM EGTA, and protease inhibitors (Roche mini-complete EDTA free) at room temperature for 10min, followed by 3 times washes with Ca⁺⁺-free Artificial Sea Water. Once embryos are fully dissociated, cells were immediately proceeded to the next step or stored in -80 °C with 10% DMSO before further processing. The transposition reaction and amplification procedures were performed following a protocol described in (Buenrostro, Giresi et al. 2013). Briefly, approximately 500,000 cells were washed with ice-cold PBS and resuspended in 50 ul cold lysis buffer (10mM Tris-HCl, pH 7.4, 10mM NaCl, 3mM MgCl₂, 0.1 % IGEPAL CA-630) for nucleus extraction. The nuclei were gently re-suspended in a 50ul transposition reaction mixture containing 25 ul Tagment DNA buffer (Illumina, Nextera DNA 24 sample prep kit), 2.5ul Tagment DNA enzyme and 22.5ul of Nuclease Free H₂O. The mixture was incubated at 37 °C for 30 minutes, and then proceeded immediately to DNA purification using a Qiagen MinElute Kit. To amplify transposed DNA for sequencing, the following reaction was prepared: 10 ul Transposed DNA, 15 ul of PCR master mix (Illumina, Nextera DNA 24 sample prep kit), 5 ul of PCR primer cocktail and 5 ul of each Index primer (Illumina, Nextera 96 sample Index kit). PCR samples were fragmented into 300bp. The sequencing was done in pair-end 40 millions 50bp reads. The raw reads were aligned to sea urchin reference genome

Spur_v3.1/strPur4 using Bowtie2. Peaks were visualized on EchinoBase Genome Browser with those that have alignment of 10 or greater are shown.

4C-seq

4C-seq assays were performed as previously reported (Stadhouders, Kolovos et al. 2013). Embryos dissociation followed the same procedure as described above in the ATAC-seq assay. Approximately 1 million embryos were dissociated and cross-linked for 10 min in 1% (wt/vol) paraformaldehyde in PBS at room temperature. The reaction was stopped by adding glycine to a final concentration of 125 mM for 5 min at room temperature. Fixed cells were then treated with lysis buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% (vol/vol) NP-40 and 1× protease inhibitor mixture (Complete; Roche; 11697498001) on ice for 10 minutes. The primary nuclei digestion was done with AseI endonuclease (New England Biolabs, R0526L) and ligated with T4 DNA ligase (New England Biolabs; M0202S). The secondary digestion was done with DpnII endonuclease (New England Biolabs; R0543L), and the DNA was ligated again.

Specific primers (with sequences are 5'-GTCTTTCACCCTCTCTCACTC-3' and 5'-AAAGTGCCAGTGGGACTG-3') were designed to target the promoter region of *hox11/13b* and were used to generate library in PCR reaction performed with the Expand Long Template PCR System (Roche; 11759060001). PCR reactions were purified using the QIAquick PCR Purification Kit (Qiagen) and

proceeded to size-selection for 200-800bp fragments using Agencourt AMPure XP (Beckman Coulter, A63881). PCR fragments were then ligated with Illumina adaptors. Libraries were sequenced on the Illumina HiSeq2000. Single-end reads of 100bp length were obtained. 4C-seq data were analyzed as described in (Stadhouders, Kolovos et al. 2013) with some changes. In brief, bait-specific sequences were determined to enrich for the real signal reads. These reads were then trimmed down to 36bp and mapped to the sea urchin genome (Spur_v3.1/strPur4) with Bowtie2. Peaks were visualized on EchinoBase Genome Browser.

ACKNOWLEDGMENTS. We are grateful to Peng He for his help on the 4C-seq data analysis and Say-Tar Goh and Dr. Brian Williams for their helpful discussion as well as experimental assistance on the 4C-seq experiments. We would like to thank Parul Kudtarkar and Dr. Andrew R Cameron for their help on the ATAC-seq data analysis and the maintenance of sea urchin genome browser.

SUPPLEMENTAL FIGURES AND TABLE

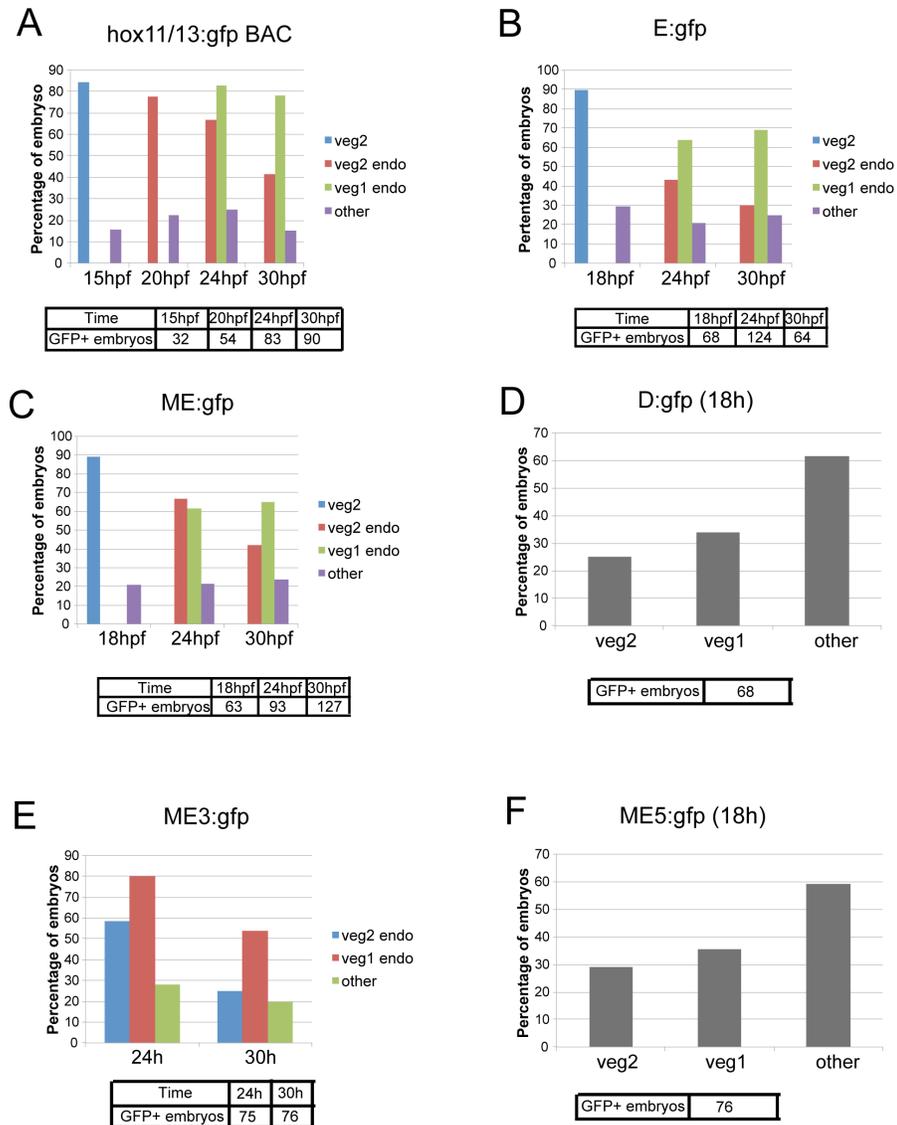


Figure S1: Lineage distribution of GFP reporter expression driven by *hox11/13b* BAC *sp_4005C17* and active modules. Whole mount *in situ* hybridization was performed to detect the localization of *gfp* transcripts in the embryos injected with *hox11/13b:gfp* BAC, *E:gfp*, or *ME:gfp*. Cells expressing *gfp* mRNA were scored for their origin of lineage and is shown in stacked columns bar chart. Total number of *gfp* positive embryos scored at each time point is listed in tables.

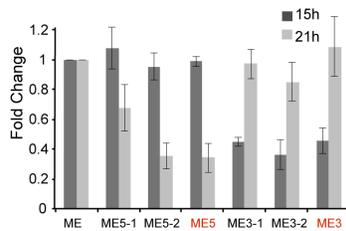
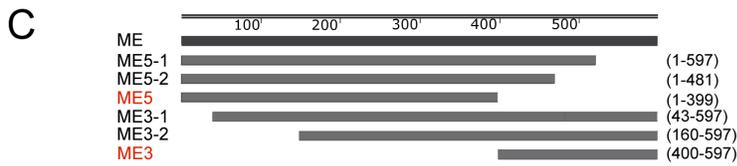
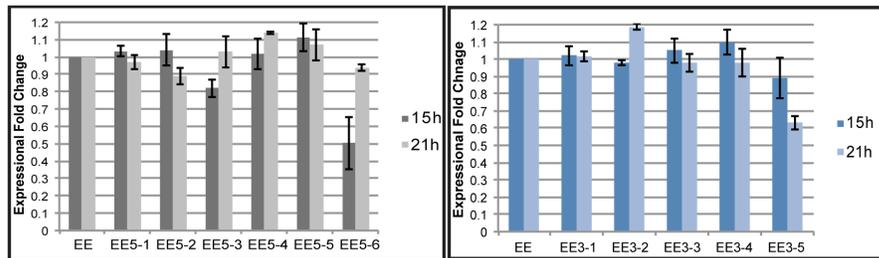
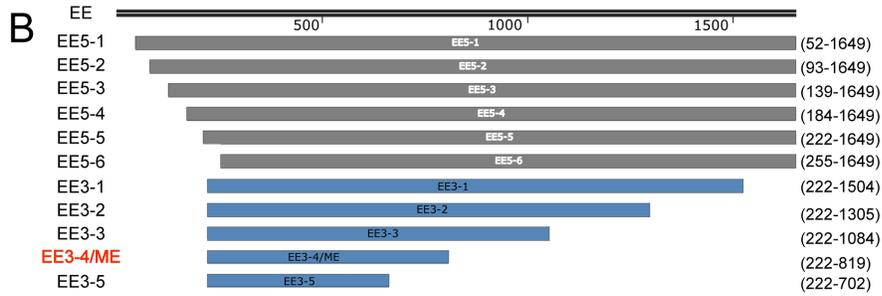
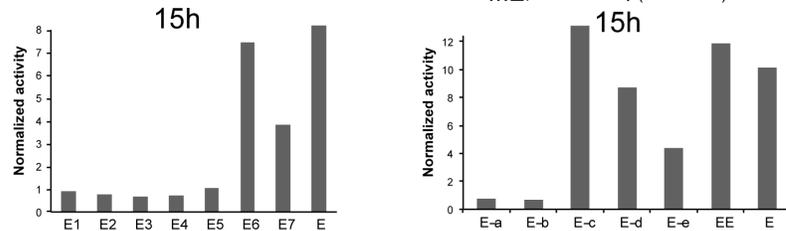
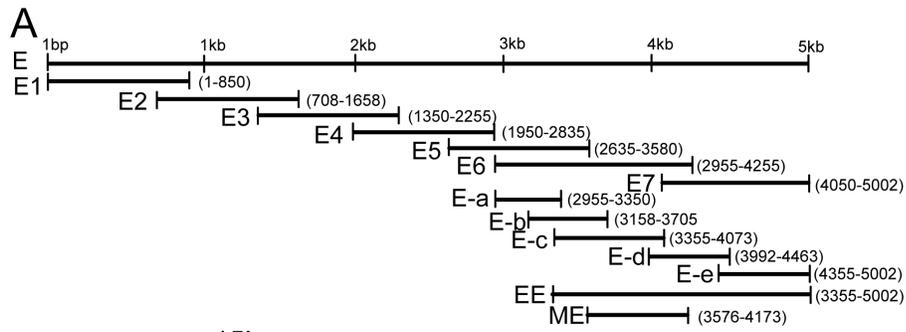


Figure S2: Discovery of the minimal sequence (ME) of the early module E.

(A) The entire region of module E (~5kb) was first dissected into 7 partially overlapping fragments (E1 to E7), whose regulatory activity was measured using QPCR at 15h shown as histogram on the left. The region covering two active fragments E6 and E7 was again dissected into five smaller fragments (E-a to E-e) that also partially overlap with each by 500bp at the ends. QPCR measurement of the regulatory activity of these smaller constructs at 15h identified three active sequences: E-c, E-d and E-e. The combined region of these three active sequences is named EE and was subjected to searching for even smaller sequence with comparable regulatory activity. (B) Progressive trimming from both ends of sequence EE was performed in order to search for the minimal region. QPCR measurement of the reporter activity driven by various deletion forms of construct was performed at 15h and 21h and is shown in histograms. Fold change was calculated by comparing reporter expression driven by deleted constructs to that driven by the intact sequence EE. (C) Sub-modularization within module ME. Deletion series from both ends of ME discovered that the sequence encoded at the 5' end (ME5) and 3' end (ME3) are independently responsible for the reporter expression at 15h and 21h, respectively. The histogram shows quantitative PCR measurement of reporter activity of various deletion constructs normalized to that of the wild-type ME construct.

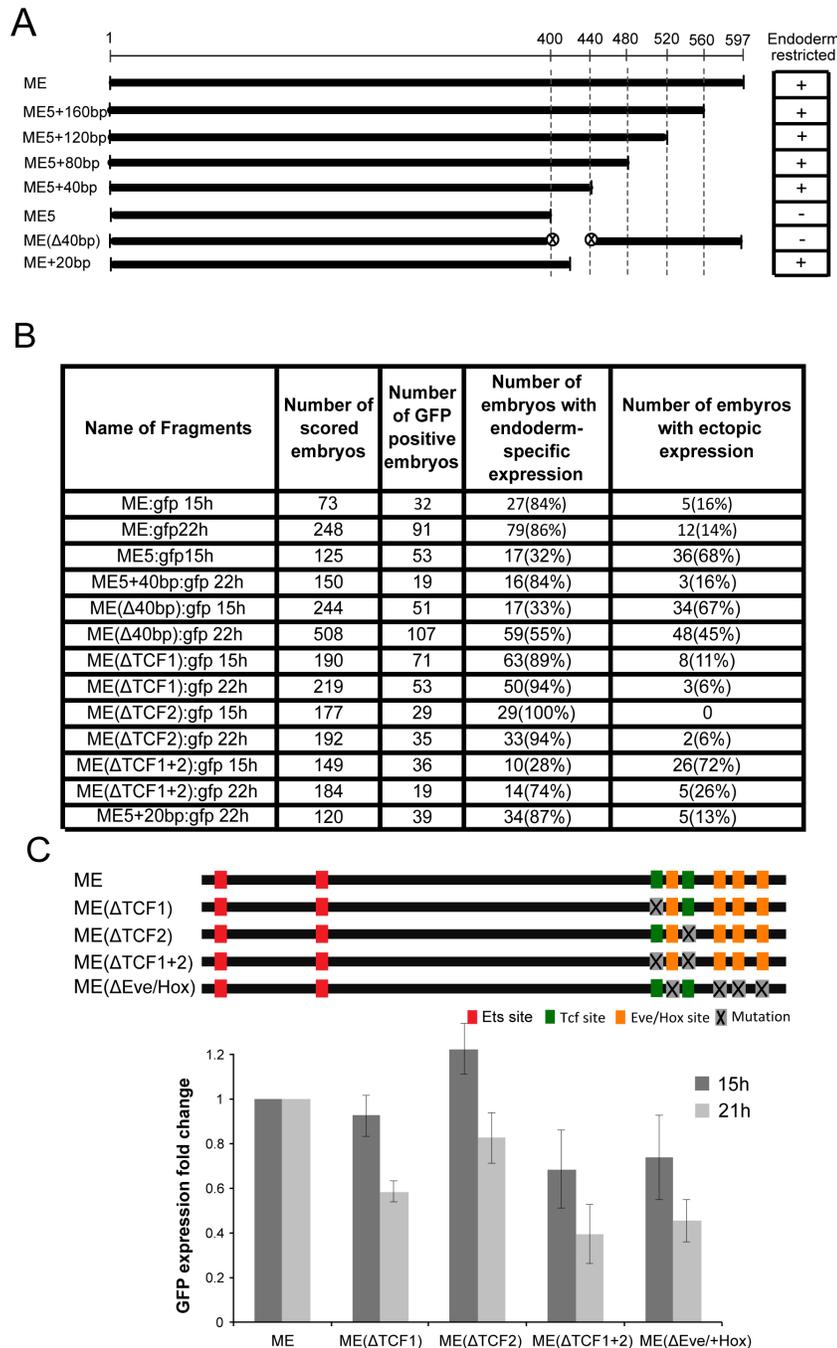


Figure S3: Tcf sites are required for the spatial regulation (A) progressive trimming from the 3' end of ME identified a 40bp sequence that is necessary and sufficient for restricting reporter expression to the endoderm, examined at 20h by fluorescent microscope. (B) Embryo scoring data shows the percentage of GFP positive embryos having endoderm-specific expression for various mutant or deleted forms of reporter constructs. (C) Quantitative change of *gfp* expression in TCF mutant constructs.

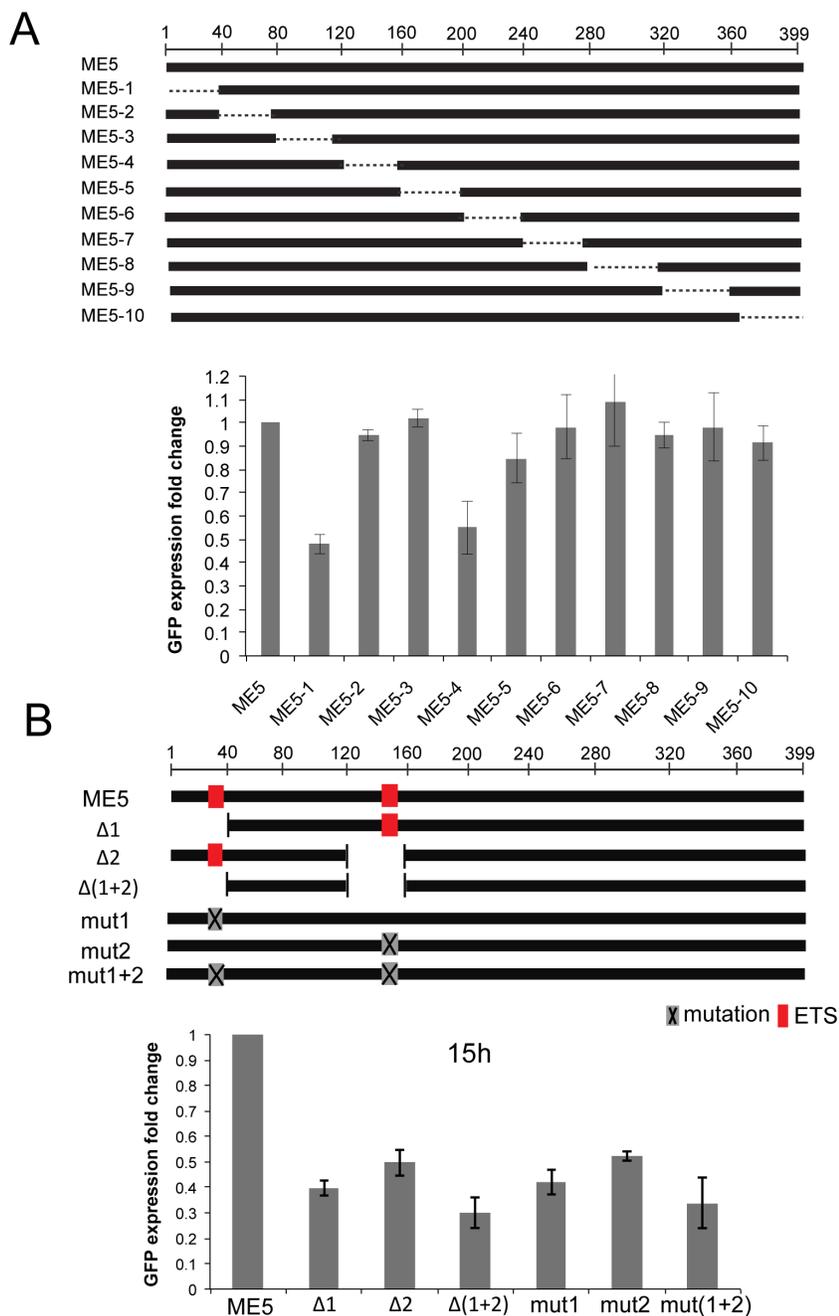


Figure S4: Deletions on ME5 discovered functional Ets binding sites required for driving the reporter expression at 15h. (A) Systematic deletion covering the entire region of ME5 identified two sequences that are required for driving the reporter expression at 15h. The histogram shows QPCR measurement of reporter activity of various deletion constructs normalized to that of the wild-type ME5 construct. (B) Deletion of Ets sites decreased the expression level of ME5 reporter construct at 15h.



Figure S5: Sequence of module ME and module L and the identified functional binding sites within. (A) Sequence and functional binding sites of module ME. Ets sites are shown in blue; Eve sites in orange; Hox sites in pink; Tcf sites in green. Note that three Eve sites overlap with Hox sites, thus in the text we refer them to Eve/Hox site. The mutated sequences in forward strand for each site are shown in parentheses and labeled in red. (B) Sequence and functional binding site of module L Hox sites are shown in pink. The mutated sequences in forward strand for each site are shown in parentheses and labeled in red. The 142bp sequence required for the late activity that was detected for ATAC-seq signal at 60h is shown in yellow.

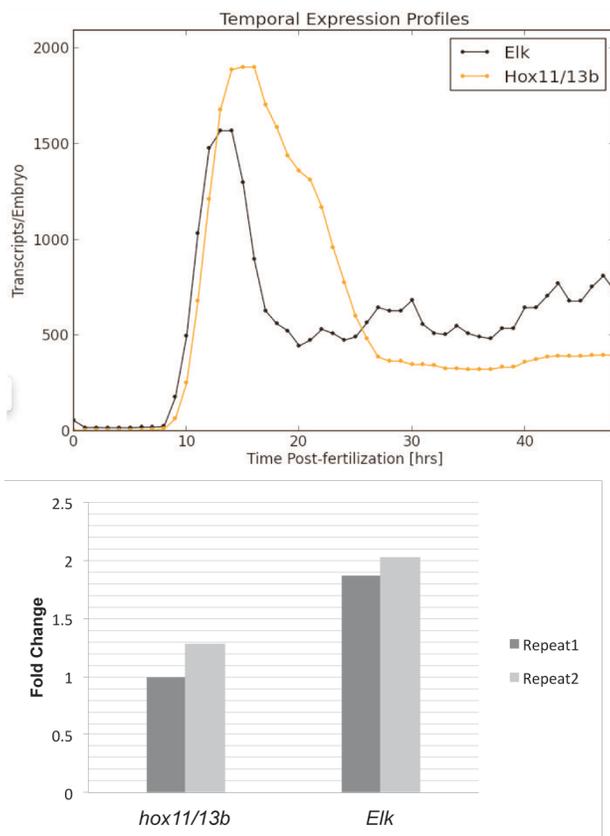


Figure S6: Transcriptional profile of elk and hox11/13b; expressional change of hox11/13b and elk in the embryos injected with Elk MASO at 15h.

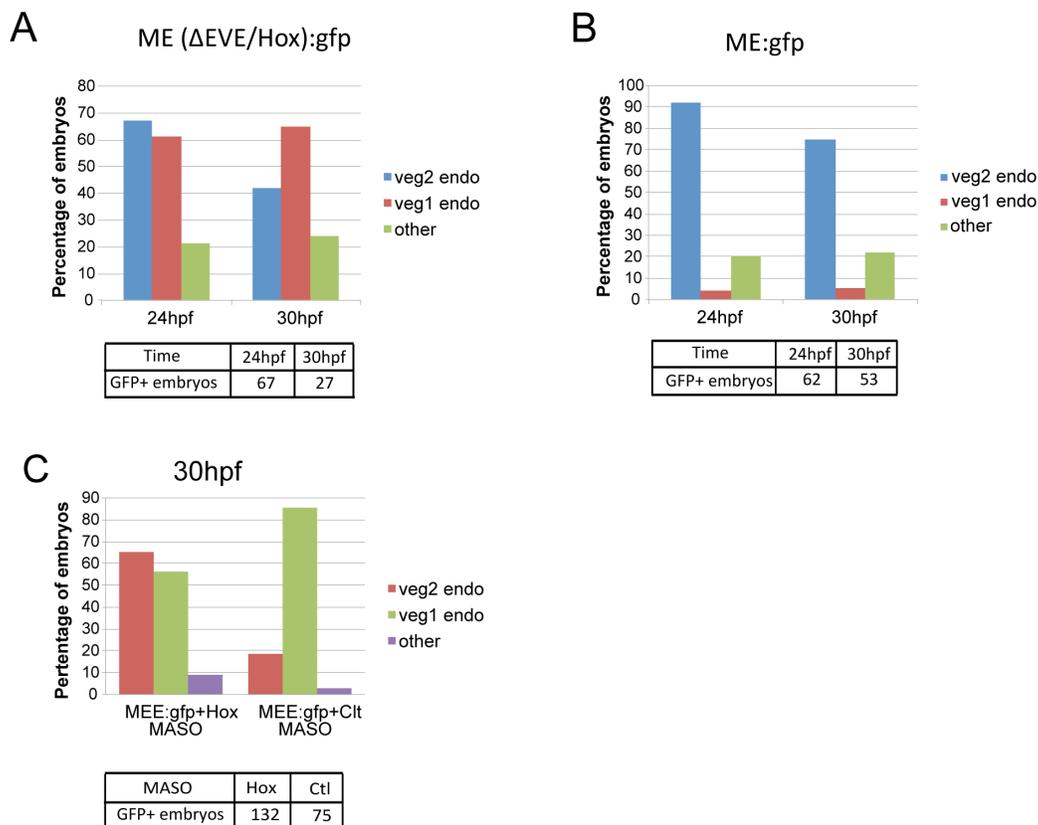


Figure S7: Hox11/13b mediated auto-repression is required for the clearance of its own transcripts in the anterior endoderm. Whole mount *in situ* hybridization was performed to detect the localization of *gfp* transcripts in the embryos injected with either Eve/Hox site mutant *ME:gfp* construct (A), wild-type *ME:gfp* construct (B), or co-injected with *ME:gfp* construct plus Hox11/13b MASO or control MASO (C). Cells expressing *gfp* mRNA were scored for their origin of lineage and is shown in stacked columns. Total number of GFP positive embryos scored at each time point is listed in tables. Percentage of embryos expressing *gfp* in the posterior endoderm is significantly reduced in the embryos injected Eve/Hox site mutant construct (compare the green bars between A and B); percentage of embryos failed to clear *gfp* mRNA in the anterior endoderm at 30h is doubled in the embryos injected with Eve/Hox sites mutant construct compared to that of the wild-type ME construct (compare the red bars between two time points within A and B, as well as at 30h between A and B). In C, failed clearance of *gfp* transcripts in the anterior endoderm of the embryos injected with wild-type *ME:gfp* construct was induced by co-injecting Hox11/13b MASO but not control MASO, shown at 30h.

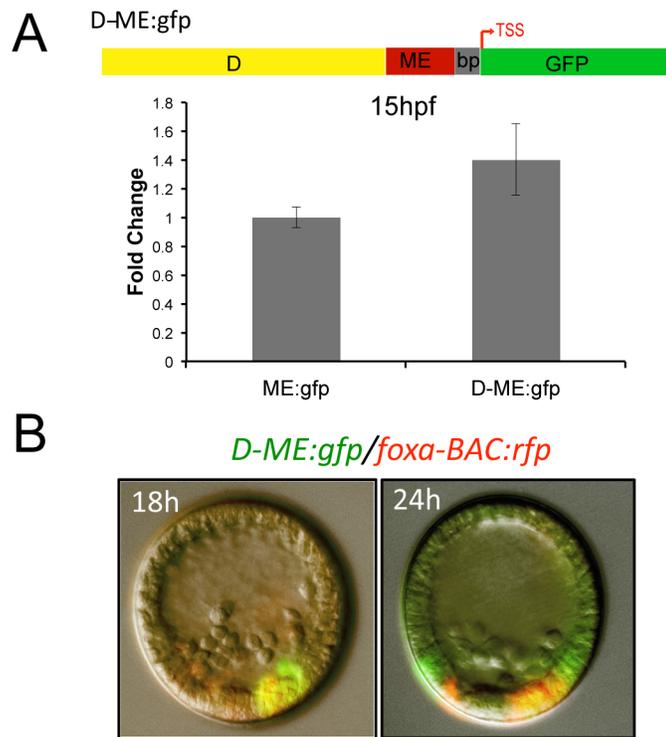


Figure S8: Module D contributes to the robust activation at 15h and listens to the spatial regulation of module ME. (A) Addition of module D to *ME:gfp* construct is sufficient to increase the *gfp* transcripts level by 33%-57% in two biological replicates. (B) The fusion construct *D-ME:gfp* is capable of driving endoderm-specific *gfp* expression, despite *D:gfp* alone is ectopic (shown in Figure S1D). *Foxa-BAC:rfp* was co-injected with *D-ME:gfp* to label endomesoderm at 18h and anterior endoderm at 24h.

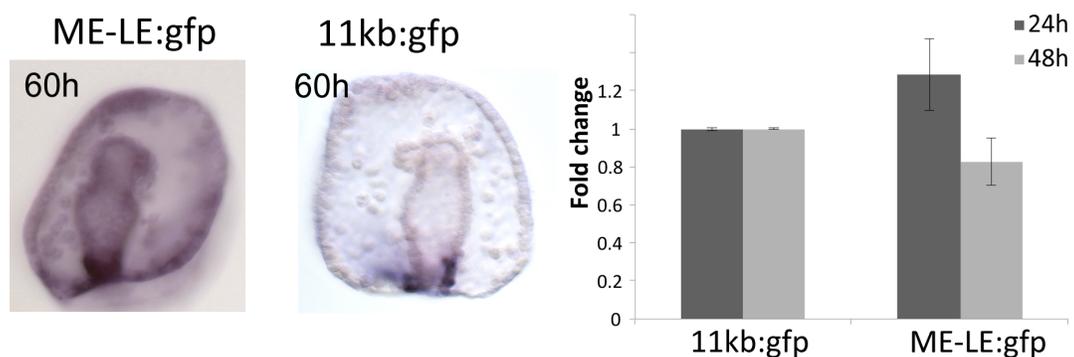


Figure S9: Construct with the configuration of *ME-LE:gfp* is capable of driving the *gfp* expression in the hindgut at 60h; the *gfp* expression level in the embryos injected with *ME-LE:gfp* is comparable to that in the embryos injected with the *11kb:gfp* construct.

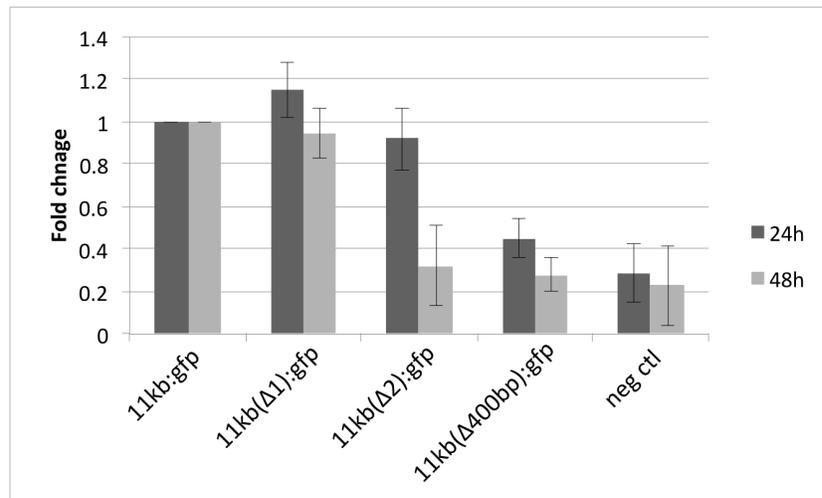
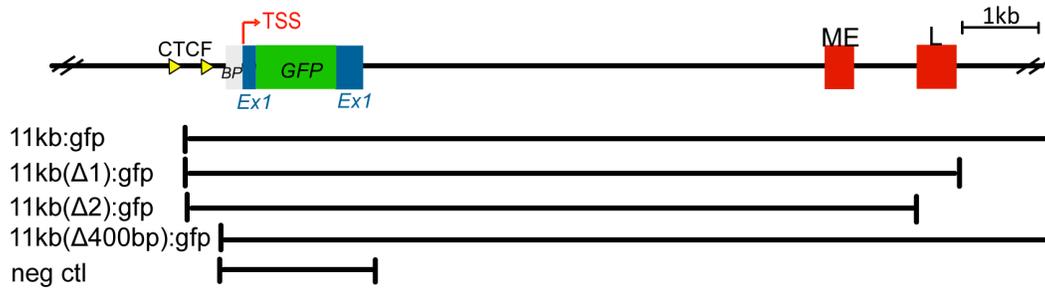


Figure S10: Quantitative PCR measurement of the reporter expression driven by various deletion forms of the *11kb:gfp* construct. Fold change was calculated by comparing reporter expression driven by deleted constructs to that driven by the intact 11kb sequence.

Fragments	Forward Primer	Reverse Primer
1	ATTCAAGGACAACCTCGGCTG	TTTCCACCACCTCATTGAATTGTGC
2	CCGAAAGTCACACGGATTC	CAATGTTAACTACCTAGCAATG
3	AATGACAGCTGATCGTGAGC	GCAGCATTCTACTTCTCTACT
4	AACGGTCAACAATGTGTTACGT	GCACTCACTTATGTCTGATTCG
5	CTGGATACACCAATCTCCGT	ATGGGAGTTCCTGACTGACC
6	CATTCAAGTTGGCAGAGAGCT	CAACCCAGTCTGTAGTACTAGT
7	CAATAGCGCCATCTAGTGAC	TTGGCGAGTCGGTTTACAGC
8	AATCGAATAAGCCTGCCTCA	CCTAATAAGCAACGCATGAC
9	GCCATACAACATTGAACAGC	AGAATTCGACATAGGCCTGC
10	TCACTGCGCTTGTATCGAGCA	TCTCATAGTGCACGATGACCT
11	AATCGCCTTGAATCGTCTGG	CTGCCTTTAACCGATGAATGT
12	GCCTCTCGCCAAAACAAAT	ACTGCCAATACGTTCTCACT
13	GTCGATCTGTATGCTGTATGTC	TGGATGTACTTCGACTGTCGG
14	GCTGTCTTTAAGGAACAGGT	AACATTCCTTGTGTCTCCAA
15	CCAACCTCCCGCTTGATG	GAGTCTCAAGCGCAAATTTATG
16 (E)	ACATCAACCGGTTATATCACC	AATTGGCATGCCTTGAGAG
17	GGCCAGATAGGATAGAGATCA	CTTGCCCGCTTTGTTATCTCT
18	GTTGGGCACAGACAAACT	TATAGGCGCTCGTCTGACA
19	TTAATGACGCAGTAACAGCGAG	TTTGACAGCCTCAAAGCACC
20	GATTGTGTATCGGATATGTGC	CTGCCAGGTTCAATGACAT
21	CAAACGAGTAGTGTGAAGTTT	TAGATGGCCACTTGCTCTTCA
22	GACAAAAGCTAGGCTCCTTG	ACGATGCATTGAATCCGTTTCA
23	GGCAATGGTGTATGACCTTAG	GGGTCTAATATCGACCAAAA
24	TGGTAGAGAGGGCTACAGAG	GATGCTTTGATCAGTCGTCG
25	TGTCCTATACGTCACAATGCA	AGTAATGATAAACCGGCACC
26	CACGTGGGGCAAGTTCTTGG	TTCGACGGCAGTAAGTCTG
27	CATCATGACGACTTGAATGAC	TCTCACAATGCTATGACAGCT
28	ACATAAGAGATAGGAGACTGCA	AGTTGACTGGGGGTGGTT
29	TCAGCTAATGACTGCAATCAA	AGCTCTTACCTTGTCTTATC
30	AAACAATCGAGTAGCATCCCA	CAGATCATCACGCAGAGAC
31	TGGGAAGGTATAGCAGAATGTTG	TGTAGCATGCAAGGCTACCC
32	GGATACCAAACGGAGCGTCA	GGGTTTGTCTGCACCTTATC
33	GCATTGGCTAGGCTTGCATC	GCTATCACCCCTTCAAGG
34	CTTCGTTTATCATCATTGACA	GTCTCAGCTGAATCAACTG
35	TGGGCGACGAACCTGTTAC	GCTTGCCTGTATCCATCTCAG
36	ATGGGTCAGAACCATAAGTCA	AGGGGACAGTTATATGCGATC
37	GTATTCTGCATGCTCGGAGC	ATCATGTGCGCTACGAATCC
38	GAAATCTCTGACGTTACTGGTAA	GATGTCAAATGGACATCTTCTT
39	GGTTGTGACTGCTGTTGAC	TGACGTTACATGACCACAAC
40	TTCCGAGCAAAGTATCAAGA	GAATTGTCAACCATGAACGTC
41	CCAAGGCACGGAATTAGAAG	GCCGTTGAGATGAATGGTTG
42 (D)	AATAACGAAGGAGCGTAGAC	AACTCAATCGGGATAAGTG
43	AACCGTGGGATGTGGATAAGT	AATTTCCACCAGTCTGCTTCTT
44	GCAGTGCATCGAACTCAC	GTTGGTGTATGGTGTATATCGT
45	TCAGTTGGATCCTGAAGATTG	ACATAAAATTTGCACAATATGGC
46	ATATCATGATTGCCTAACATGC	CATGTGTTAAATCTGAGATACAG
47	GTATTACGACACAGAACATTGTTG	AGCTATACAATTTTGACAGAGG
48	ATGGGCCCTGCCCTTAT	TCGCTCATGAACCTTTGGCA
49	CGGGTACCCATTAACT	GTGGATTAGCTACTTTACGGA
50	GTCGGATCTGTACTAGTGAGT	ATGTGGAGCATTGTAGAAGC
51	TCTTCCCATCACAATCGACC	AGTGTATGTACGGTCTACAGT
ME	TTTAAGCAGATTGAATTACCC	TTTTCTTCCCTTCAATTGTGACACCG
ME5	TTTAAGCAGATTGAATTACCC	GCAGCGTTCGATTTCAATCA
ME3	TTGTCTTCTCAATACGGATA	TTTTCTTCCCTTCAATTGTGACACCG
L	TTGATACAGATCCAACCTGAGG	CGTTTGGCTTCTCAATGACG

Table S1: primers used to amplify putative cis-regulatory modules

REFERENCES

Arenas-Mena, C., A. R. Cameron and E. H. Davidson (2000). "Spatial expression of Hox cluster genes in the ontogeny of a sea urchin." Development **127**(21): 4631-4643.

Arenas-Mena, C., R. A. Cameron and E. H. Davidson (2006). "Hindgut specification and cell-adhesion functions of Sphox11/13b in the endoderm of the sea urchin embryo." Dev Growth Differ **48**(7): 463-472.

Arenas-Mena, C., P. Martinez, R. A. Cameron and E. H. Davidson (1998). "Expression of the Hox gene complex in the indirect development of a sea urchin." Proc Natl Acad Sci U S A **95**(22): 13062-13067.

Barsi, J. C. and E. H. Davidson (2016). "cis-Regulatory control of the initial neurogenic pattern of onecut gene expression in the sea urchin embryo." Dev Biol **409**(1): 310-318.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang and W. J. Greenleaf (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." Nat Methods **10**(12): 1213-1218.

Buenrostro, J. D., B. Wu, H. Y. Chang and W. J. Greenleaf (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide." Curr Protoc Mol Biol **109**: 21 29 21-29.

Cameron, R. A., L. Rowen, R. Nesbitt, S. Bloom, J. P. Rast, K. Berney, C. Arenas-Mena, P. Martinez, S. Lucas, P. M. Richardson, E. H. Davidson, K. J. Peterson and L. Hood (2006). "Unusual gene order and organization of the sea urchin hox cluster." J Exp Zool B Mol Dev Evol **306**(1): 45-58.

Cheers, M. S. and C. A. Ettensohn (2004). "Rapid microinjection of fertilized eggs." Methods Cell Biol **74**: 287-310.

Cochella, L. and O. Hobert (2012). "Embryonic priming of a miRNA locus predetermines postmitotic neuronal left/right asymmetry in *C. elegans*." Cell **151**(6): 1229-1242.

Cui M, S. N., Li E, Davidson EH, Peter IS. (2014). "Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo." Proc Natl Acad Sci U S A **111**(47): E5029-5038.

Davidson, E. H. (2006). "The Regulatory Genome: Gene Regulatory Networks in Development and Evolution." Academic Press/Elsevier, San Diego.

Davidson, E. H., R. A. Cameron and A. Ransick (1998). "Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms."

Development **125**(17): 3269-3290.

de Santa Barbara, P. and D. J. Roberts (2002). "Tail gut endoderm and gut/genitourinary/tail development: a new tissue-specific role for Hoxa13."

Development **129**(3): 551-561.

de-Leon, S. B. and E. H. Davidson (2010). "Information processing at the foxa node of the sea urchin endomesoderm specification network." Proc Natl Acad Sci

U S A **107**(22): 10103-10108.

Duboule, D. (2007). "The rise and fall of Hox gene clusters." Development **134**(14): 2549-2560.

Dunipace, L., A. Ozdemir and A. Stathopoulos (2011). "Complex interactions between cis-regulatory modules in native conformation are critical for *Drosophila* snail expression." Development **138**(18): 4075-4084.

Erwin DH, D. E. (2009). "The evolution of hierarchical gene regulatory networks." Nat Rev Genet **10**: 141-148.

Ferrier, D. E. and P. W. Holland (2001). "Ancient origin of the Hox gene cluster." Nat Rev Genet **2**(1): 33-38.

Gaj, T., C. A. Gersbach and C. F. Barbas, 3rd (2013). "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering." Trends Biotechnol **31**(7): 397-405.

Geyer, P. K. (1997). "The role of insulator elements in defining domains of gene expression." Curr Opin Genet Dev **7**(2): 242-248.

Guo, Y., Q. Xu, D. Canzio, J. Shou, J. Li, D. U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M. Q. Zhang, B. Ren, A. R. Krainer, T. Maniatis and Q. Wu (2015). "CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function." Cell **162**(4): 900-910.

Holmes, S., Lyman, S., Hsu, J. K., Cheng, J. (2015). "Making BAC transgene constructs with lambda-red recombineering system for transgenic animals or cell lines. ." Methods Mol. Biol. **1227**: 71-98.

Kim, C. H., T. Oda, M. Itoh, D. Jiang, K. B. Artinger, S. C. Chandrasekharappa, W. Driever and A. B. Chitnis (2000). "Repressor activity of Headless/Tcf3 is essential for vertebrate head formation." Nature **407**(6806): 913-916.

Krumlauf, R. (1994). "Hox genes in vertebrate development." Cell **78**(2): 191-201.

Laurell T, N. D., Hofmeister W, Lindstrand A, Ahituv N, Vandermeer J, Amilon A, Annerén G, Arner M, Pettersson M, Jääntti N, Rosberg HE, Cattini PA,

Nordenskjöld A, Mäkitie O, Grigelioniene G, Nordgren A (2014). "Identification of three novel FGF16 mutations in X-linked recessive fusion of the fourth and fifth metacarpals and possible correlation with heart disease." Mol Genet Genomic Med **2**(5): 402-411.

Lee, E. C., D. Yu, J. Martinez de Velasco, L. Tessarollo, D. A. Swing, D. L. Court, N. A. Jenkins and N. G. Copeland (2001). "A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA." Genomics **73**(1): 56-65.

Lemons, D. and W. McGinnis (2006). "Genomic evolution of Hox gene clusters." Science **313**(5795): 1918-1922.

Li, E., M. Cui, I. S. Peter and E. H. Davidson (2014). "Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo." Proc Natl Acad Sci U S A **111**(10): E906-913.

Lin, N., X. Li, K. Cui, I. Chepelev, F. Tie, B. Liu, G. Li, P. Harte, K. Zhao, S. Huang and L. Zhou (2011). "A barrier-only boundary element delimits the formation of facultative heterochromatin in *Drosophila melanogaster* and vertebrates." Mol Cell Biol **31**(13): 2729-2741.

Livant, D. L., B. R. Hough-Evans, J. G. Moore, R. J. Britten and E. H. Davidson (1991). "Differential stability of expression of similarly specified endogenous and exogenous genes in the sea urchin embryo." Development **113**(2): 385-398.

Logan, C. Y., J. R. Miller, M. J. Ferkowicz and D. R. McClay (1999). "Nuclear beta-catenin is required to specify vegetal cell fates in the sea urchin embryo." Development **126**(2): 345-357.

Matharu, N. and N. Ahituv (2015). "Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease." PLoS Genet **11**(12): e1005640.

Minokawa, T., A. H. Wikramanayake and E. H. Davidson (2005). "cis-Regulatory inputs of the wnt8 gene in the sea urchin endomesoderm network." Dev Biol **288**(2): 545-558.

Nam, J. and E. H. Davidson (2012). "Barcoded DNA-tag reporters for multiplex cis-regulatory analysis." PLoS One **7**(4): e35934.

Nam, J., P. Dong, R. Tarpine, S. Istrail and E. H. Davidson (2010). "Functional cis-regulatory genomics for systems biology." Proc Natl Acad Sci U S A **107**(8): 3930-3935.

Peter, I. S. and E. H. Davidson (2010). "The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage." Dev Biol **340**(2): 188-199.

Peter, I. S. and E. H. Davidson (2011). "A gene regulatory network controlling the embryonic specification of endoderm." Nature **474**(7353): 635-639.

Peter IS, D. E. (2009). "Genomic control of patterning." Int J Dev Biol **53**: 707-716.

Peter, I. S. a. D., Eric H (2015). "Genomic Control Process: Development and Evolution." Academic Press , San Diego ISBN 978-0-12-404729-7.

Range, R. C., J. M. Venuti and D. R. McClay (2005). "LvGroucho and nuclear beta-catenin functionally compete for Tef binding to influence activation of the endomesoderm gene regulatory network in the sea urchin embryo." Dev Biol **279**(1): 252-267.

Revilla-i-Domingo, R., T. Minokawa and E. H. Davidson (2004). "R11: a cis-regulatory node of the sea urchin embryo gene network that controls early expression of SpDelta in micromeres." Dev Biol **274**(2): 438-451.

Rizzo, F., M. Fernandez-Serra, P. Squarzoni, A. Archimandritis and M. I. Arnone (2006). "Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*)." Dev Biol **300**(1): 35-48.

Scotti, M., Y. Kherdjemil, M. Roux and M. Kmita (2015). "A Hoxa13:Cre mouse strain for conditional gene manipulation in developing limb, hindgut, and urogenital system." Genesis **53**(6): 366-376.

Sethi AJ, W. R., Angerer RC, Range RC, Angerer LM. (2012). "Sequential signaling crosstalk regulates endomesoderm segregation in sea urchin embryos.

. " Science **335**(6068): 590-593.

Smith, A. B., D. Pisani, J. A. Mackenzie-Dodds, B. Stockley, B. L. Webster and D. T. Littlewood (2006). "Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata)." Mol Biol Evol **23**(10): 1832-1851.

Smith, J. and E. H. Davidson (2008). "Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo." Proc Natl Acad Sci U S A **105**(51): 20089-20094.

Smith, J., E. Kraemer, H. Liu, C. Theodoris and E. Davidson (2008). "A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo." Dev Biol **313**(2): 863-875.

Splinter, E., H. Heath, J. Kooren, R. J. Palstra, P. Klous, F. Grosveld, N. Galjart and W. de Laat (2006). "CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus." Genes Dev **20**(17): 2349-2354.

Stadhouders, R., P. Kolovos, R. Brouwer, J. Zuin, A. van den Heuvel, C. Kockx, R. J. Palstra, K. S. Wendt, F. Grosveld, W. van Ijcken and E. Soler (2013).

"Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions." Nat Protoc **8**(3): 509-524.

VanderMeer JE, L. R., Sun M, Xue Y, Daentl D, Jabs EW, Wilcox WR, Ahituv N. (2014). "A Novel ZRS Mutation Leads to Preaxial Polydactyly Type 2 in a Heterozygous Form and Werner Mesomelic Syndrome in a Homozygous Form." Hum Mutat **35**(8): 945-948.

Warot, X., C. Fromental-Ramain, V. Fraulob, P. Chambon and P. Dolle (1997). "Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts." Development **124**(23): 4781-4791.

Zacchetti, G., D. Duboule and J. Zakany (2007). "Hox gene function in vertebrate gut morphogenesis: the case of the caecum." Development **134**(22): 3967-3973.

*Chapter 4***CONCLUSION**

In this thesis, I pursued two methods for validating and refining the GRN models of sea urchin embryonic development. First, through perturbation of Wnt protein secretion and translation, I investigated the regulatory functions of the Wnt signaling system in the specification processes of all embryonic lineages during pre-gastrular development. This study extends previous individual assessments of Wnt signaling to the genomic level by taking into account the regulatory functions of all Wnt ligands on the expression of all expressed transcription factors. As the function of these regulatory genes in sea urchin embryonic GRNs is well known, we can directly conclude from this analysis, to an unprecedented degree, which cell fate specification GRNs are under the control of Wnt signaling and which ones are not. Second, I performed a cis-regulatory analysis on the *hox11/13b* gene, which is an important regulatory factor for endoderm specification, particularly for the posterior endoderm. I showed that the early expression of this gene is directly activated by broadly expressed TFs (Ets and Eve) and is spatially restricted by the dual role of Tcf. I identified that the regulatory logic responsible for its dynamic spatial expression was controlled by a lineage-specific response to Wnt signaling, changes in activators between domains, and its auto-repression. I also discovered an AND logic gate in which the late expression of *hox11/13b* in the hindgut requires two regulatory elements, and both are necessary but are not sufficient on their own.

Below I summarized some of the main insights discovered in this thesis

Wnt signaling is not required for the specification of mesoderm lineages and the early expression of endoderm genes due to the presence of maternal beta-catenin.

The study described in Chapter 2 determined that the Wnt signaling inputs are required for three general aspects of embryonic specification: broad activation of endodermal GRNs, regional specification of the immediately adjacent vegetal ectoderm, and the restriction of the apical neurogenic domain. A surprising discovery is that Wnt signaling is not required for the specification of mesodermal cell lineages nor is it necessary for the early specification of the endoderm. Two distinct lines of evidence support this: interfering with the secretion of all Wnt ligands using the Porcupine inhibitor C59 had no effect on the mesoderm specification or on the initial expression of endoderm genes, and this same effect was also observed upon the perturbation of individual Wnt ligands. These observations are in contrast with previous assumptions that Wnt8 is required for the earliest Tcf-dependent expression of endodermal regulatory genes and also for the initiation of the skeletogenic mesoderm GRN (Smith and Davidson 2008). By comparing the effect observed upon blocking Wnt protein secretion to that of disrupting the nuclear localization of beta-catenin (using a Dn-cadherin construct), we were able to separate the function of maternal nuclear beta-catenin from that of zygotic Wnt ligand signaling. We conclude that maternal beta-catenin is sufficient

to drive the activation of the mesodermal specification GRNs and the early expression of endoderm genes. Thus, despite the early expression of *wnt* genes start at 7h and in mesodermal cells, Wnt signaling is dispensable for the expression of mesodermal genes and is not required for the initial expression of endodermal genes.

The earliest function of Wnt signaling is to restrict the neurogenic apical plate

The earliest effect we observed with C59 treatment was the up-regulation and spatial expansion of apical gene expression at 12h. This effect is due to the functional loss of Wnt8, which we confirmed via Wnt8 morpholino perturbation. Thus, interestingly, although the expression of *wnt* genes is restricted to vegetal cells throughout sea urchin embryogenesis, they are first employed to regulate the patterning of the animal domains. Using timed pulse-treatment with inhibitor C59, we discovered a window of time, between 12h to 18h, when apical genes are responsive to Wnt signaling perturbation. No change in apical gene expression was observed in the embryos treated with C59 after 18h post-fertilization.

The spatial restriction of the apical domain is therefore a successive process that eventually results in the confined local activation of the neurogenic apical specification GRN in the most anterior region of the embryo. Three sequential steps are involved in this process. The first step happens between 7h and 12h post-fertilization and is mediated by maternal nuclear beta-catenin. The second step,

between 12h and 18h, requires zygotic Wnt signaling, specifically that of Wnt8 which may function through the non-canonical Wnt/Fzl5/8-JNK and Fzl1/2/7-PKC pathways (Range, Angerer et al. 2013). Repressors expressed in the adjacent non-neurogenic ectodermal cells are also required for this process, as shown by a dramatic increase in *foxq2* expression in embryos injected with *not* or *emx* morpholinos (Li, Cui et al. 2014). Thus, between 12h to 18h, the spatially confined separation of the neurogenic apical plate from the rest of the ectoderm requires both Wnt8 signaling as well as a locally operated gene repression mechanism. At the third step, after 18h, this apical restriction no longer depends upon Wnt signaling. This may be due to the presence of multiple Wnt signaling antagonists including Dkk1, Dkk3, and SFRP1/5 in the apical domain, or it may be because the GRN structure is stabilized and self-maintained, “locking in” the developmental fate of the neurogenic apical plate at this time.

Functional specificity of Wnt signaling

We observed functional diversity among individual *wnt* genes, despite their partially overlapping expression patterns. *Wnt8* is specialized to restrict the expression of apical genes. *Wnt1* is generally utilized to activate endoderm genes, although it also activates a selection of genes expressed in the vegetal ectoderm including *nk1*, *sp5*, and *unc4.1*. *Wnt16* is specifically required for the activation of early endodermal genes including *blimp1b*, *eve*, and *hox11/13b*. *Wnt4* is only employed to activate the expression of a subset of ectoderm genes (i.e. *hox7* and

msx) in the vegetal ectoderm.

An important question is how the specificity of Wnt signaling is mediated, and this is a broadly applicable question given the multiplicity of *wnt* and *fzd* genes encoded in all animal genomes. This functional diversity of different *wnt* genes in the sea urchin cannot be explained by the utilization of diverse Frizzled receptors, because only one receptor -Fzd9/10 - is present in vegetal cells, including the endoderm and vegetal ectoderm where multiple *wnt* ligands are received. Future experiments are necessary to understand the mechanism underlying this functional specificity of *wnt* genes. Potential research directions include exploring the differential expression in the diverse embryonic domains of Wnt co-receptors such as those of the low-density lipoprotein receptor-related protein (LRP) family (van Amerongen and Nusse 2009).

This specificity is also reflected by the differential activation of target genes by the same Wnt ligand. A related question is how Wnt signaling in different embryonic domains or different developmental phases leads to the regulation of specific yet distinct sets of target genes. For example, Wnt1 activates endodermal genes only in the endoderm and ectodermal genes only in the vegetal ectoderm. One possible explanation for this is that there is combinatorial regulatory control for the genes expressed in response to Wnt signaling. We know that Tcf, the transcription factor activated by Wnt signaling, operates in concert with other transcription factors, some of which are exclusively expressed in cells of a given fate at a particular

developmental time. For example, most Wnt1 target genes activated in endoderm cells are also activated by Hox11/13b, while *nkl* is activated by Wnt signaling in the vegetal ectoderm and is also a target of the Not and Lim1 transcription factors expressed in these cells (Li, Materna et al. 2012). Thus the impact of Wnt signaling on gene expression in cells of different fates causally depends on the regulatory state in the cells receiving the signal.

The regulatory logic for *hox11/13b* expression

As described in Chapter 3, the cis-regulation of *hox11/13b* expression employs AND logic gates both within and between regulatory modules. For the early expression, before pre-gastrular stage, the early module E utilizes a combination of activators (Ets and Eve) and a spatial regulator (Tcf/beta-catenin) to drive the dynamic and endoderm-specific expression of *hox11/13b*. A lack of these activators leads to insufficient expression, whereas a lack of spatial regulation induces ectopic expression. This AND logic gate, in which both regulatory players are required and each executes a specialized function, is clearly optimized to drive this specific and dynamic expression. One clear advantage to this system is the assurance of specificity: only when two players are both present can *hox11/13b* be expressed. A second advantage of this regulatory logic is that it allows some level of flexibility, as different TFs with same regulatory function can be utilized in multiple combinations in different cells to create more combinations of regulatory players, leading to many dynamic expression patterns. This is the exact strategy used to

control the different expression phases of *hox11/13b*. When *hox11/13b* is expressed in the endomesoderm, a ubiquitously expressed Ets factor is employed for the activation function. This function is however replaced by Eve a few hours later in Veg1 cells.

An AND regulatory logic gate also exists between modules for the late expression of *hox11/13b* in the hindgut: two cis-modules are both required but neither one is sufficient. The mechanism for this is not yet understood; however, we have discovered that the Eve/Hox sites in the early module E, which are required for the early posterior endoderm expression, are needed again for the late expression. Curiously, Eve transcripts are not present in the hindgut expressing *hox11/13b* (Jonathan Valencia, unpublished data), so it is unlikely that Eve also activates this late expression. At present, we think there are two potential explanations for this requirement of early Eve/Hox binding sites for the late activity of *hox11/13b*: i) they may contribute to the direct binding of unidentified late activators; ii) they do not directly contribute to the late TFs binding per se, but instead create a favorable chromatin state (chromatin decompression, H3K4m1 modification, etc.) by activating the early module which may then facilitate late expression. To distinguish between these two scenarios, more experiments focused upon identifying the late regulators and understanding the epigenetic signatures of these sequences in their native context and in the early module mutant background would be helpful. A straightforward experiment to test the second scenario would be to replace the early module with one from a different gene that is also active in the

hindgut at the right time to see if this chimeric construct is able to drive the late expression. The discovery of this cooperative regulation between early and late enhancers tells us the enhancer cooperation is much more comprehensive than we had previously thought. One precaution in interpreting these data is that data collected from any high-throughput enhancer screening system, e.g. STARR-seq and MPRA, is only an incomplete representation of enhancer activity and additional result confirmation is needed.

The establishment of the endoderm/ectoderm boundary is controlled at the *hox11/13b* locus

The boundary between the endoderm and the ectoderm arises within the veg1 lineage, with the inner cells specifying the posterior endoderm and the outer cells giving rise to the vegetal ectoderm. *Hox11/13b* is the first gene differentially expressed within the veg1 lineages, and it is specific to the posterior endoderm. At the protein level it represses the expression of ectodermal genes in the presumptive posterior endoderm, thus defining the boundary between the endoderm and ectoderm. The genomic code that controls the specific expression of *hox11/13b* in the posterior endoderm is therefore essential for the establishment of this boundary. As described in Chapter 3, we discovered that the spatial repression of *hox11/13b* in the presumptive ectoderm is regulated by the dual function of Tcf. When Tcf sites were mutated in a reporter construct this spatial restriction was lost, as evidenced by ectopic expression of the reporter gene. From the study described in Chapter 2, we

also know that the Wnt1/Wnt16 signaling pathway is required for the expression of *hox11/13b* in the posterior endoderm. The expression of *wnt1* and *wnt16* is also restricted to the posterior endoderm and thus it is likely to activate *hox11/13b* via autocrine signaling or in a community effect. One unresolved question is that of why the veg1 ectodermal cells, which directly abut the cells expressing *wnt1/wnt16* and *hox11/13b* and also express the same receptor *fzd9/10* as well as the same activator *eve*, do not activate the expression of *hox11/13b*. Additional regulation must be involved such that the veg1 ectoderm cells become immune to the activation function of Wnt1/Wnt16 signaling on the *hox11/13b* expression, but the nature of this regulation remains uncertain.

Context-dependent multi-functionality of binding sites in the *hox11/13b* regulatory modules

The cis-regulatory analysis carried out in Chapter 3 focuses on an extensive developmental period covering multiple phases of *hox11/13b* expression across several different embryonic domains. This allows us to study the functional dynamics of TF binding sites in different domains and at different times. What we see repeatedly is that same functional sites adopt distinct regulatory functions in different development contexts. For example, the Tcf sites in the early module first mediate the activation of *hox11/13b* in the veg2 cells, yet are meanwhile required for the repression of *hox11/13b* in all other domains. These same Tcf sites are later employed to represses *hox11/13b* expression in the inner veg2 cells that are

precursors to the mesoderm, all while continuing to activate the expression of *hox11/13b* in the outer veg2 cells that become the anterior endoderm. Functional dynamics are also observed with the Eve/Hox sites residing in the early module. These sites are required for the activation of *hox11/13b* through Eve in the posterior endoderm, and they concomitantly mediate the auto-repression of *hox11/13b* in the anterior endoderm. In the later phase, these sites are required again for the expression in the hindgut. Therefore, unlike most other cis-regulatory studies which focus on the functions of TF binding sites in a particular context, here we use a systematic approach to demonstrate the functional dynamics of TF binding sites in the cells of different regulatory states.

REFERENCES

Li, E., M. Cui, I. S. Peter and E. H. Davidson (2014). "Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo." Proc Natl Acad Sci U S A **111**(10): E906-913.

Li, E., S. C. Materna and E. H. Davidson (2012). "Direct and indirect control of oral ectoderm regulatory gene expression by Nodal signaling in the sea urchin embryo." Dev Biol **369**(2): 377-385.

Range, R. C., R. C. Angerer and L. M. Angerer (2013). "Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos." PLoS Biol **11**(1): e1001467.

Smith, J. and E. H. Davidson (2008). "Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo." Proc Natl Acad Sci U S A **105**(51): 20089-20094.

Van Amerongen, R. and R. Nusse (2009). "Towards an integrated view of Wnt signaling in development." Development **136**(19): 3205-3214.

