THE REGULATORY ORIGIN OF ORAL AND ABORAL MESODERM IN SEA URCHIN EMBRYOS

Thesis by

Stefan Christian Materna

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy in

Biochemistry and Molecular Biophysics

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2012

(Defended September 20, 2011)

ACKNOWLEDGEMENTS

First and foremost I would like to thank Eric Davidson for having accepted me as a student in his lab. Little did I know what I had stumbled into when I first arrived at Caltech. Asking around before leaving Germany what Eric's lab is like, the most common response was 'different from your standard german lab'. It certainly is, but in the best possible way. I could not have hoped for a more stimulating and motivating environment than the one I found. Thank you, Eric, for challenging and encouraging me, for the many enlightening discussions, and, of course, introducing me to good bourbon.

I appreciate the time and energy of Marianne Bronner, Angela Stathopoulos and Lea Goentoro, the members of my thesis committee. Thank you for your support and advice.

I am grateful for the friendship of my fellow graduate students Meredith Ashby, Titus Brown, Sagar Damle, Roger Revilla, Pei Yun Lee, Eric Erkenbrack, and Jon Valencia. Only you truly know what the experience is like and can appreciate the pleasure *and* pain.

Many thanks to all the people that I worked with closely over the years: Jongmin Nam – my congenial bench mate of many years, Andy Ransick – the other half of Team Mesoderm, Joel Smith, Enhu Li, Andy Cameron and Paola Oliveri. Jane Rigg and Deanna Thomas are the glue that keeps the lab together; wherever I might end up, it will be a lesser place because of your absence. Thanks to all the other wonderful people in lab, too numerous to list you all but no less significant. You have contributed to making my time there special.

I am indebted to my friend Celina Juliano who was a great help on the home stretch and the voice of reason in the last few weeks of sheer madness. I would like to express my gratitude for the time spent with Barbara and Andy Cameron. My time in Pasadena would not have been the same without you.

Last but not least I would like to thank my parents, Helga and Reinhard Materna, and my sister, Christa, for their support and encouragement. They have not seen much of me in the last few years but I know they missed me as much as I missed them. Yet, they did not complain once about my prolonged absence.

Thank you all.

ABSTRACT

Gene regulatory networks (GRN) underlie the control processes that are executed during embryonic development. Their constituents are transcription factors that regulate downstream targets, including other transcription factors. The regulatory architecture of a GRN reveals how discrete developmental tasks, such as cell specification, are implemented.

We here expand the GRN underlying development of sea urchin non-skeletogenic mesoderm (NSM). NSM cells are the offspring of the inner ring of the *veg2* cells that lie adjacent to the skeletogenic mesoderm (SM) and receive the Delta signal presented by these cells. Perturbation of Delta reveals that all NSM-specific genes are activated by Delta, but also indicate that Delta has few direct targets. A large number of genes are activated only after *delta* expression in the SM disappears, thus indicating that these genes are indirect targets and downstream of early NSM transcription factors. We show that the second phase of *delta* expresion (in the NSM) activates the *foxY* gene; loss of NSM Delta does not interfere with early NSM specification but instead abolishes development of late mesoderm derivates.

NSM is partitioned into an oral and an aboral segment as a consequence of Nodal signaling. However, Nodal activates the homeobox gene *not,* which represses early NSM genes on the oral side, causing them to become restricted to the aboral side. This allows oral NSM genes to be activated. Oral NSM genes can be expressed throughout the entire NSM if aboral NSM specification is perturbed. This shows that the driver of NSM genes is present throughout the NSM, making SM Delta a likely candidate. We examine the regulatory state of oral and aboral NSM segments and show that a GRN subcircuit on the aboral side locks down its expression. The sets of regulatory genes on both sides of the NSM are entirely distinct and mutually exclusive.

TABLE OF CONTENTS

INTRODUCTION

The instructions for how an animal develops from a fertilized egg are encoded in its genome. But despite an abundance of genomic data, how these instructions are executed during development is still insufficiently understood and remains one of the most exciting questions in biology.

In the nucleus gene expression is regulated by transcription factors. They bind to the cis-regulatory modules of downstream targets where these inputs are integrated and cause changes of transcription. The expression pattern of any gene can thus be understood in terms of the structure/function properties of its cis-regulatory apparatus (Davidson, 2006). However, a single gene cannot by itself cause development; rather, many regulatory genes need to act in concert to drive development forward.

The cross-regulatory relationships among transcription factors can be summarized in gene regulatory network (GRN) maps. GRNs explain how and why spatial regulatory states are set up as development proceeds. They consist of modular entities, or subcircuits, each of which comprises the interactions necessary to achieve a discrete developmental task, for example, the activation of an inductive signal, the lockdown of a transcriptional state, or the expression of differentiation genes (Davidson, 2006). Identification and functional analysis of network subcircuits illuminate the character of the underlying regulatory apparatus.

The goal of this thesis was to unravel the GRN underlying the specification of the non-skeletogenic mesoderm (NSM) in sea urchins. The NSM is born at seventh cleavage when endomesoderm precursors divide and form a ring around the cells of the skeletogenic mesoderm (SM) (Fig. 0.1). The SM is the source of the Delta ligand, a short-range signal that activates early mesoderm genes in the NSM (Sherwood and McClay, 1999; Sweet et al., 2002). Gene expression is initially uniform but, following establishment of the oral/aboral axis, the NSM is divided into an oral and aboral segment (Duboc et al., 2010; 2004). New

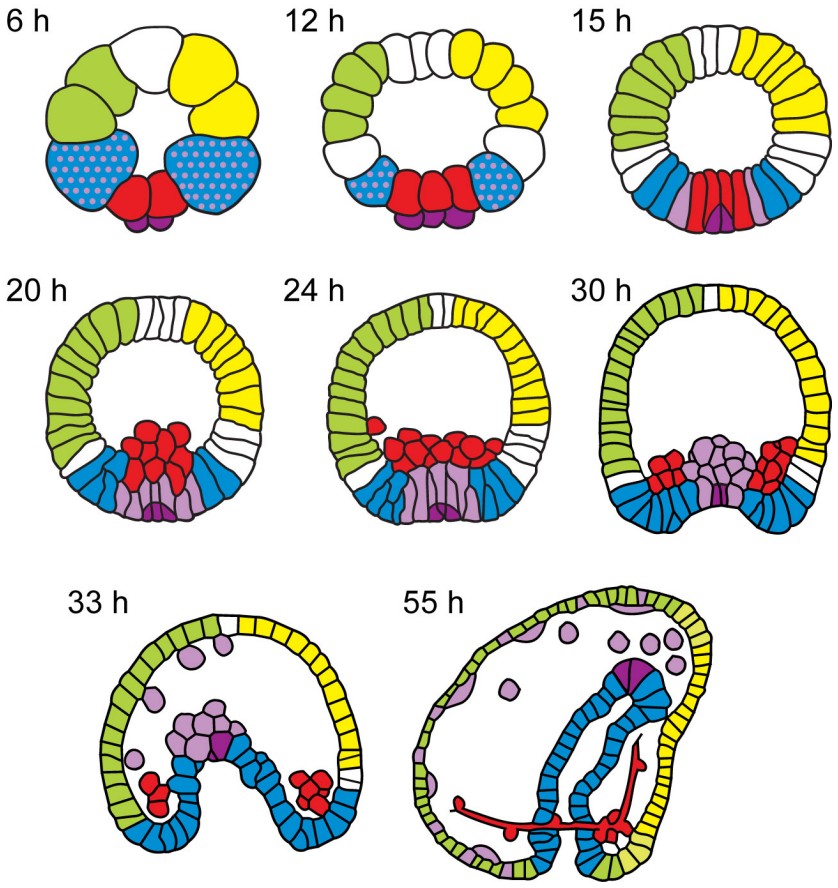


Figure 0.1: Schematic representation of early sea urchin development. The first cleavages divide the embryo into an apical and a vegetal half that are distinguished by the size of their cells. The cells of the apical half, the mesomeres, will develop into the ectoderm. In the vegetal half the fourth unequal cleavage gives rise to macromeres and micromeres. The sixth cleavage of micromeres produces large and small daughter cells (6 h). The large micromeres (red) form the skeletogenic mesoderm (SM). These cells lie at the center of the vegetal plate until they ingress into the blastocoel before gastrulation (20 h). SM cells will eventually fuse and secrete skeletal matrix to form the spicules of the larvae (55 h). The small micromeres (dark purple) are quiescent and up to late gastrulation divide only once more. After their birth, they remain adjacent to the SM but stay behind as these cells ingress. During gastrulation they are located at the tip of the archenteron and will contribute to the coelomic pouches once they form. The macromeres will give rise to mainly endoderm and non-skeletogenic mesoderm (NSM) (6 h, blue with purple dots). They split into two rings of cells during sixth cleavage, an inner tier (*veg2*) and an outer tier (*veg1*, white at 12 h). The *veg1* tier will ultimately contribute to ectoderm and gut. At seventh cleavage the inner tier again splits into an inner and an outer ring separating endoderm precursors (blue) from non-skeletogenic mesoderm (NSM, purple). All mesodermal cell types (except the skeletogenic cells) are derived from this group of cells. These are the pigment cells that intercalate into the aboral ectoderm during gastrulation, the blastocoelar cells, the coelomic pouches, and the muscle cells that surround the esophagus.

genes will turn on in the oral NSM while early mesodermal genes become restricted to the aboral NSM. The descendants of the oral NSM will give rise to blastocoelar cells that have immune function. The cells of the aboral NSM will develop into pigment cells.

A process diagram that lists the developmental tasks that have to be executed for each segment to be specified is shown in Fig. 0.2. Because the NSM is initially uniform, this must include the mutual exclusion of oral and aboral fates before differentiation can occur. Such mutual exclusion functions are typically found in neighboring cells that are the descendants of one common progenitor and which adopt different cell fates (Davidson, 2006). As the NSM and the endoderm also have a common origin, another function of NSM genes has to be the exclusion of endoderm fate (Peter and Davidson, 2011).

To populate the bubbles of the process diagram with the transcription factors that implement these functions, we made wide use of the sea urchin genome. Its sequence provides a full list of transcriptional regulators that made it possible to address the question with a true systems biology approach. However, this required the identification of all transcriptional regulators that are active in early development (Howard-Ashby et al., 2006a, b, c; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). The groundwork, done by myslef and others proved to be a tremendous resource and made it possible to address the regulatory origin of the NSM in a more comprehensive way than we could have otherwise.

A brief review of the GRN approach to development is given in Chapter 1. This chapter outlines the general strategy for building GRNs and contains a detailed protocol for how to unravel linkages between regulatory genes. Since building and revising even a small piece of a GRN is a huge undertaking if it is to be reasonably complete, the steps detailed in the protocol have turned into the following chapters of this thesis.

Chapter 2 describes our effort to identify zinc finger transcription factors in the sea urchin genome and to characterize their expression throughout early development. Zinc finger transcription factors are unique among transcription factors; their DNA binding domain consists of at least two, but usually three or more zinc binding domains that together determine the DNA binding site specificity of the protein. This modular structure may be one

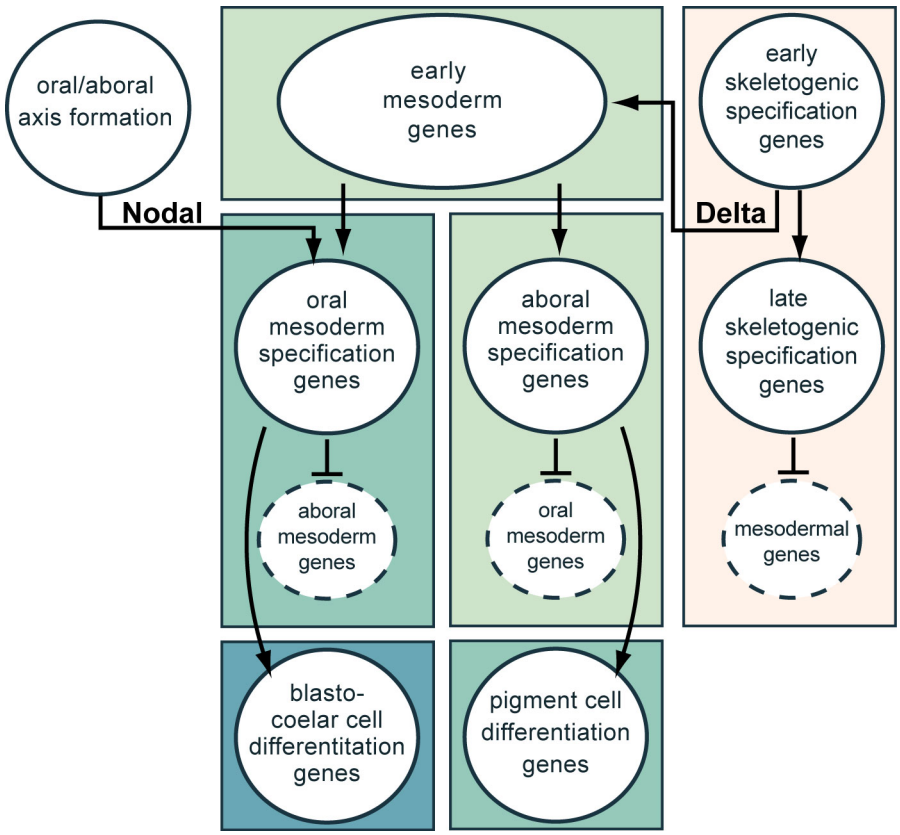


Figure 0.2: A process diagram for mesoderm development. This diagram outlines in broad strokes the functions that need to be executed during mesoderm specification. The early Delta signal from the neighboring skeletogenic mesoderm activates non-skeletogenic mesoderm transcription factors in the entire mesoderm. Nodal signaling establishes the oral/aboral axis along which the NSM is subdivided. The transcription factors expressed in either segment will have to lock down their respective regulatory state as signaling inputs are transient. The NSM is initially uniform, thus one important function is the exclusion of the alternative fate for differentiation to occur.

of the reasons that zinc finger genes are hugely expanded in deuterostomes. This work is part of the larger effort to identify all transcriptional regulators that are active in early development.

The third chapter is an extension of our initial transcriptional profiling effort. The more detailed knowledge of gene expression is available, the better the hypothesis that can be formulated for how these genes may be linked to each other. With the arrival of the Nanostring nCounter, an RNA counting device that achieves high accuracy, we were able to

obtain high-denisty timecourses for 172 mostly regulatory genes in a semi-high-throughput manner. This set includes the majority of transcribed and spatially restricted transcription factors in early development. It allows accurate determination of the onset of gene expression for the genes that are under tight transcriptional control as is the case for many transcriptional regulators.

As mentioned above, the initial, activating signal for NSM specification is Delta expressed in the SM. Chapter 4 is a study of the role of Delta signaling in the early sea urchin embryo and identifies its targets in a comprehensive way, at least with regard to the known and relevant transcriptional regulators. By evaluating gene expression levels in a time resolved manner, we were able to show that Delta has only a few potentially direct targets. We were also able to disentangle the second expression phase of Delta (when it is expressed in the NSM) and could show that the only target of NSM Delta is the forkhead gene *foxY*. The defects caused by this perturbation are limited to loss of late mesodermal cell types, namely coelomic pouch cells and esophageal muscles.

Finally, Chapter 5 addresses the subdivision of the NSM. We expand the set of NSM genes significantly and show that both segments are mutually exclusive. The primary cause of the split into oral and aboral NSM had previously been identified as Nodal signaling from the oral ectoderm. However, making use of the resources we had built, we were able to show that Nodal, although an important upstream activator, is not a direct input into oral NSM genes. Instead, the function of Nodal is to turn on a repressor that shuts down aboral NSM genes orally. We find that expression of oral NSM genes depends on an activator that is present throughout the NSM, and argue that this is the Delta signal originating in the SM.

References

Davidson, E., 2006. The regulatory genome: gene regulatory networks in development and evolution. Academic Press, San Diego, CA.

Duboc, V., Lapraz, F., Saudemont, A., Bessodes, N., Mekpoh, F., Haillot, E., Quirin, M., Lepage, T., 2010. Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. Development 137, 223-235.

Duboc, V., Röttinger, E., Besnardeau, L., Lepage, T., 2004. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. Dev Cell 6, 397-410.

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

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

Howard-Ashby, M., Materna, S.C., Brown, C.T., Tu, Q., Oliveri, P., Cameron, R.A., Davidson, E.H., 2006c. High regulatory gene use in sea urchin embryogenesis: Implications for bilaterian development and evolution. Dev Biol 300, 27-34.

Materna, S.C., Howard-Ashby, M., Gray, R.F., Davidson, E.H., 2006. The C_2H_2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Dev Biol 300, 108-120.

Peter, I.S., Davidson, E.H., 2011. A gene regulatory network controlling the embryonic specification of endoderm. Nature 474, 635-639.

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

Sherwood, D.R., McClay, D.R., 1999. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. Development 126, 1703-1713.

Sweet, H.C., Gehring, M., Ettensohn, C.A., 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. Development 129, 1945-1955.

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

Chapter 1

A PROTOCOL FOR UNRAVELING GENE REGULATORY NETWORKS

Stefan C. Materna [1] & Paola Oliveri [1,2]

1. Division of Biology, California Institute of Technology, Pasadena, CA, 91125

2. Department of Biology, University College London, Gower Street, London, WC1W 6BT

(Published in Nature Protocols 2008, 3:1876 - 1887)

Abstract

Regulatory genes form large networks that are fundamental to the developmental program. The protocol presented here describes a general approach to assemble maps of gene regulatory networks (GRNs). It combines high-resolution spatio-temporal profiling of regulatory genes, strategies to perturb gene expression, and quantification of perturbation effects on other genes of the network. The map of the GRN emerges by integration of these data sources and explains developmental events in terms of functional linkages between regulatory genes. This protocol has been successfully applied to regulatory processes in the sea urchin embryo, but is generally applicable to any developmental process that relies primarily on transcriptional regulation. Unraveling the GRN for a whole tissue or organ is a challenging undertaking and, depending on the complexity, may take anywhere from months to years to complete.

[Keywords: Gene regulatory network, perturbation, development, specification, transcription factor

Introduction

The developmental program of an organism is encoded in its genome. Fundamentally it is constructed of elaborate regulatory processes that are carried out by regulatory genes: Transcription factors bind to the *cis*-regulatory regions of downstream genes to activate, repress, or modulate their transcription and signaling factors allow essential cell-to-cell communication. All regulatory genes are themselves under the control of transcriptional regulators and together form large GRNs (Davidson, 2006; Ben-Tabou and Davidson, 2007). The sum of all transcription factors present in the nucleus of a cell defines its regulatory state and, when the underlying architecture is known, indicates what regulatory processes are currently executed (Materna and Davidson, 2007). To experimentally establish the structure of a GRN means to assemble a logic-map that reveals the cause-effect linkages of the regulatory

factors involved (Davidson et al., 2002; Materna and Davidson, 2007). A thorough understanding of GRNs reveals what drives development forward and offers a mechanistic explanation for developmental events such as specification and differentiation.

Although a whole GRN may exhibit bewildering complexity it can be dissected into modular entities (Ben-Tabou and Davidson, 2006). Each module, or subcircuit, comprises the regulatory interactions that are required to complete a discrete developmental task. This may include, among others, the interpretation of an inductive signal, activation of new regulators, lockdown of the attained regulatory state, and exclusion of alternative fates (Ben Tabou and Davidson, 2006; Oliveri and Davidson, 2007). A subcircuit is also the level that best illustrates the dynamic execution of the genomic regulatory code. Although all cells contain the entire set of genomic information, subcircuits are often active in only a few cells driven by their common regulatory state.

The protocol presented in this article has been developed for, and successfully applied, to GRNs in early sea urchin development (Davidson et al., 2002a; Davidson et al., 2002b; Oliveri et al., 2002; Oliveri et al., 2008). It is applicable to developmental events that are primarily regulated at the transcriptional level, if the system is amenable to gene-specific perturbations. However, morpholino-substituted antisense oligonucleotides (MO) knockdowns are limited to the first expression phase of the targeted gene. As loss of early expression already significantly perturbs development, it is almost impossible to assess the function of the gene during a later expression phase. In other organisms, where perturbation agents can be delivered in stages other than the fertilized egg (e.g. by injection and electroporation into the neural tube as in chickens, or where late expression is specifically disrupted genetically), this is less of a problem. Future developments in the use of expression constructs and antisense technology may help circumvent these problems.

Maps of GRNs in early sea urchin development, as well as dorsal-ventral patterning in *Drosophila* (Levine and Davidson, 2005) have reached a high level of completion: They include the majority of regulatory genes active in preparation of a developmental event and explain how a certain regulatory state is established and sustained. Other networks covering

development of the whole *Ciona intestinalis* embryo (Imai et al., 2006), specification of the chicken neural crest (Meulemans and Bronner-Frase, 2004; Sauka-Spengler et al., 2007), specification of the *Xenopus* mesoendoderm (Koide et al., 2005), dorso-ventral patterning of the mouse neural tube (Vokes et al., 2007), specification of sensory neurons in *Caenorhabditis elegans* (Johnston et al., 2005) and mesoderm specification in the *Drosophila* embryo (Sandmann et al., 2007) are work in progress (more examples have been reviewed elsewhere [Davidson, 2006]). These networks are invaluable in explaining how development proceeds in an orderly and irreversible manner. These features become obvious only by studying the regulatory architecture underlying development. The point of assembling a GRN, after all, is to understand an interesting piece of biology: How the regulatory information encoded in the genome is executed to produce an animal from a fertilized egg.

General Strategy

To determine the overall structure of a GRN, the specific linkages of every regulatory gene with a specific function in the network need to be determined experimentally. The general approach to building GRNs can roughly be divided into five phases (Fig. 1.1):

The general layout of the GRN is defined by available embryological information. This information is summarized in a process diagram.

Regulatory genes are identified and characterized at high resolution in space and time. This characterizes the regulatory state of the cells involved, and may suggest, although preliminary, the flow of regulatory information.

Perturbation experiments reveal the cause-effect linkages between regulatory genes. After altering transcription or function of a transcriptional regulator the effects on transcription of other network candidates are closely monitored with a highly sensitive, quantitative method.

Links between two factors are established where the experimental results indicate activation or repression of the downstream factor. Where the data do not discriminate between two possible network architectures, additional and often more elaborate perturbation experiments need to be designed to distinguish between alternatives.

cis-regulatory analysis of network genes is conducted to confirm the inputs established in perturbation experiments.

The result of this analysis will be a detailed, high-resolution map of the regulatory process governing the developmental events under examination. Although *cis*-regulatory analysis is an integral part of elucidating a GRN it is treated only in a cursory manner in this protocol. It has been covered in depth elsewhere (Smith, 2008).

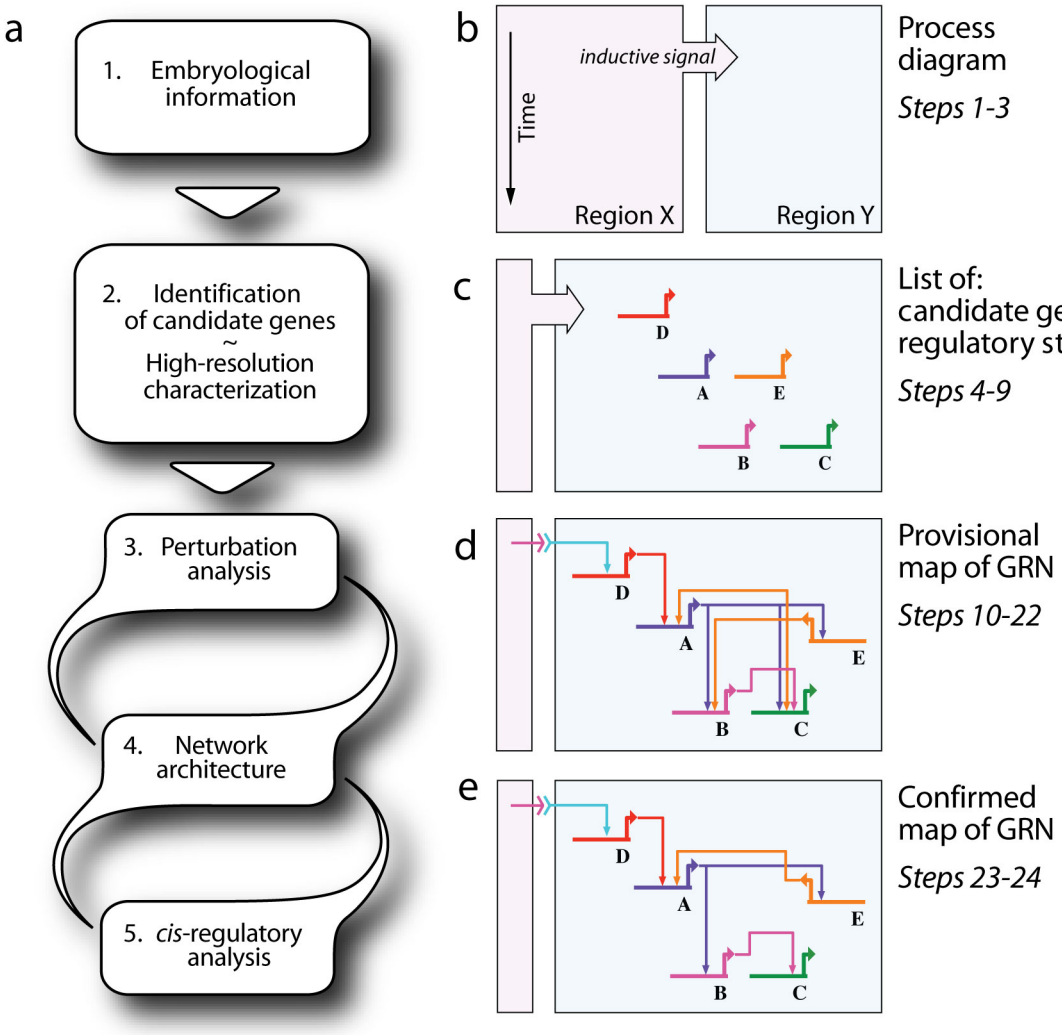


Figure 1.1: General strategy for analysis of gene regulatory networks. (A) Flowchart indicating the major steps of the protocol. The diagram indicates the close link between perturbation experiments, *cis*-regulatory analysis, and the emerging network architecture. The map of the GRN is a product of, and guide for, perturbation experiments and *cis*-regulatory analysis. (B) Initially, a priori knowledge, like embryological information, is summarized in a process diagram. Cells belonging to a particular lineage share the same regulatory state and are represented by one region. In this hypothetical example the cells of region X send out an inductive signal to their neighbors in region Y. (C) Spatio-temporal profiling identifies genes that are active in Region Y. These genes may be downstream of the inductive signal. (D) Perturbation of all candidate genes suggests a preliminary network architecture. The exact wiring between genes A, E, and B cannot be resolved by knockdown experiments only: A

and E form a positive feedback loop and either may be an input into B. (E) Additional perturbation experiments or *cis*-regulatory analysis of gene B resolve the network architecture and confirm that only gene A is a direct input into B. Steps indicated in (B – E) refer to the procedure. Network diagrams were created with the BioTapestry software (www.biotapestry.org).

Construction of a process diagram

The first step in unraveling a GRN is to define the process that is to be explained at the regulatory level (Davidson, 2006; Revilla-i-Domingo, 2003). A long-term goal might be to identify the regulatory processes underlying formation of an entire embryo or organ. But in practical terms it makes sense to confine the network to a more limited set of events in time and space. At a minimum many dozens of genes are involved in transcriptional regulation across the embryo at any given time. And the number of perturbation experiments needed to test their linkages increases dramatically with every additional gene that needs to be considered.

To sketch out the general layout of the GRN, the available information pertaining to the developmental period, or event, of interest is summarized in a process diagram. Basically this includes embryological information about cell lineages and the fate map of the embryo. A simple example of a process diagram can be found in Fig. 1.1 B.

All cells of a certain type and developmental stage are in an almost identical regulatory state: the same regulatory functions are executed in each (Materna and Davidson, 2007). Thus, in the process diagram these regulatory processes can be summarized in a single region. Where detailed gene expression patterns are available, these may help to refine the layout of the process diagram by revealing cells that are in a different regulatory state. This may justify subdividing the region at about the time the distinct regulatory states are established.

Adjacent groups of cells, or territories, may influence development of their neighbors through signaling events, and may induce neighboring cells to adopt a certain cell fate. Cell-to-cell

signaling is also involved in separating two regions and mutual exclusion of cell fates. Downstream of signaling pathways are transcriptional regulators that impinge on the regulatory state of the receiving cell. Thus, information about signaling interactions needs to be incorporated in the process diagram.

Markers are essential in assessing the developmental state. When examining specification events a commonly used marker is morphological change. However, the regulatory processes that underlie these phenotypic changes predate them by potentially long periods of time[1]. A valuable alternative to morphology is the identification of molecular markers, for example, the transcriptional activation of differentiation genes. Such genes are generally structural or metabolic and indicate specialized biological functions. For example, in the sea urchin genes for skeletal matrix proteins and enzymes of the pigment synthesis pathway have been successfully used as markers for skeletogenic and pigment cell specification, respectively (Calestani et al., 2003; Livingston et al., 2006).

In sum, the process diagram outlines the different regions of the emerging GRN and summarizes known developmental events. Many regulatory factors with an important function may already be known and their known interactions can be extracted from the literature. These can immediately be incorporated in the process diagram.

Identification of network genes

The availability of entire genomes makes gene discovery straightforward. Transcription factors have a canonical structure, consisting of primarily a structurally conserved DNA binding domain and a transactivation/repression domain. Computational tools like BLAST (a fast alignment tool; http://blast.ncbi.nlm.nih.gov/Blast.cgi) and HMMER (a protein domain identification program) rapidly identify regulatory molecules using homology criteria and structural knowledge (Howard-Ashby et al., 2006 a,b; Materna et al., 2006; Rizzo et al., 2006; Tu et al. 2006). For transcription factors (and similarly signaling molecules) these searches are,

with few exceptions, comprehensive and essentially produce a complete parts list of regulatory genes encoded in the genome of an animal. For many genomes such information is readily available from genomic databases as in the case of the sea urchin (http://spbase.org). If no sequence information is available, regulatory genes need to be identified by other methods, e.g., cDNA subtraction screens that identify regulatory genes by their involvement in developmental processes (Rast et al., 2002).

Temporal expression data identifies all the genes that are active during the time the regulatory processes under examination are executed. This information is generally easier to acquire than spatial expression data. Temporal expression data, collected with whole genome tiling arrays (Samanta et al., 2006; Stolc et al., 2004) or cDNA microarrays (Arbeitman et al., 2002), often augments genome projects and may be available from genomic databases. For organisms where no high-throughput expression data is available, quantitative expression data may have to be acquired in a semi high-throughput approach, e.g., using quantitative PCR (QPCR) as has been done for the sea urchin (Howard-Ashby et al., 2006 a,b; Materna et al., 2006; Rizzo et al., 2006; Tu et al. 2006). Although methods for expression profiling using pyrosequencing are still in their infancy, they allow analysis of expression levels at a genomic level (Mortazavi et al., 2008).

Accurate spatial expression information is essential for network building. Spatial expression data are collected using in situ detection methods in which the endogenous transcripts are localized by hybridization with a labeled antisense probe. GRNs primarily contain genes that contribute specific information to the regulatory state of cells. Ubiquitous genes do not per se provide spatial information to the developmental process, although their presence may be important[1]. For example, ubiquitous factors may have quantitative effects and enhance the transcriptional level of specific genes, or they may be important for the mechanics of gene transcription, e.g., as DNA bending factors (Yuh et al, 2001). It should be noted that some prominent exceptions exist where ubiquitous factors undergo a local modification to exert specific, local effects. This might be activation via phosphorylation at the end of signaling

cascades or tissue/cell type specific localization to the nucleus. A well-known example is the beta-catenin nuclearization system – a molecular mechanism conserved in many animal species (Henry et al., 2008; Wikramanayake et al., 2003). Nuclearization of beta-catenin in the vegetal pole cells of the sea urchin launches their specification as endomesoderm (Logan et al., 1999). These exceptions notwithstanding, the majority of regulatory genes with a specific function are spatially restricted.

Although this protocol focuses on DNA binding transcription factors (and signaling molecules), certain cofactors may also play an important role in cell specification. These may be treated in the same way as canonical transcription factors: Similar to transcription factors, cofactors are not necessarily expressed in a localized fashion. For example, in early sea urchin development the ubiquitous cofactor Groucho is an important corepressor binding to Transcription Factor 1 (TCF1) (Range et al., 2005) and Suppressor of Hairless (Barolo et al., 2002). It is replaced by beta-catenin or Notch intracellular, domain, respectively. As the GRN maps – as assembled according to this protocol – are driver networks, the emphasis lies on regulatory genes that provide spatial information.

Regulatory genes that are expressed in the region of interest at the right time constitute the shortlist of network candidates. It is important also to include genes that become specifically excluded from a given area, as these might be repressors whose disappearance is required for specification to occur (Revilla-i-Domingo et al., 2007). It is of great value to collect high-density spatio-temporal expression information for these genes. High-density timecourses are instrumental in establishing a preliminary order in which regulatory genes are activated. For the purple sea urchin, with its rather slow transcription kinetics due to life at an ambient temperature of 15°C, a good sampling interval is about 1 hour for temporal, and 3 hours for spatial expression. This may vary greatly in different experimental systems. This data set will allow identification of regulatory states and is important when picking sampling timepoints in perturbation experiments (Fig. 1.2 A, C). Ideally transcripts (and proteins) of the gene targeted in a perturbation should have accumulated significantly in unperturbed control embryos to allow them to exert their effects. The same applies to potential target genes; these should also have accumulated to a significant and reproducible level in the unperturbed embryos. Fig. 1.2

illustrates this with an example from sea urchin mesoderm involving the genes *delta*, *gcm*, and *pks*. Of these three genes the *delta* gene is the first to be expressed, followed by *gcm*. Thus, Delta might induce *gcm* expression and *pks*. The dashed lines in the timecourse diagram (Fig. 1.2 A) indicate appropriate sampling times for the perturbation of *delta* and *gcm*.

Perturbation Analysis

A detailed knowledge of temporal and spatial expression profiles is indispensable for building a GRN. However, only through perturbing transcription, or function, of a regulatory gene and quantitatively monitoring the effects can this knowledge be converted into functional understanding. Otherwise it will remain a mere catalogue of expression profiles. The next section discusses methods available to interfere with gene expression in the sea urchin system and beyond. It is followed by a discussion of how to measure quantitatively and accurately the effect of a perturbation.

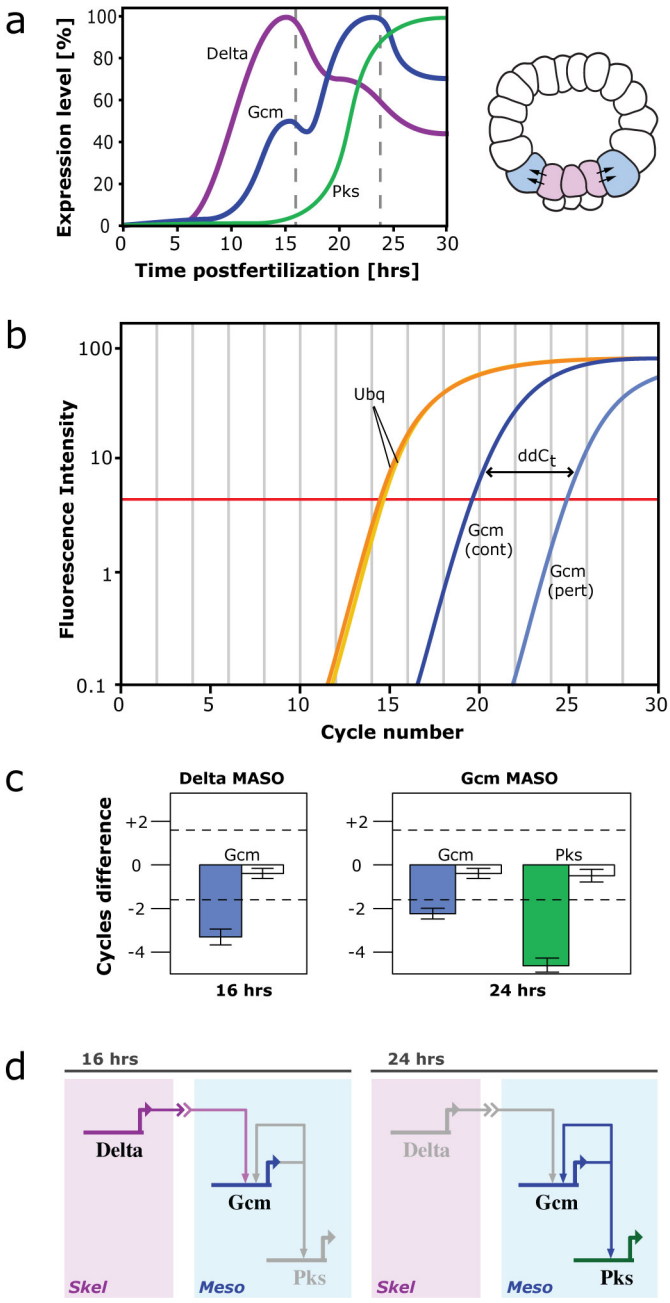


Figure 1.2: Perturbation experiments for network assembly. (A) Temporal and spatial expression data for the genes *delta*, *gcm*, and *pks* reveal their successive activation. The schematic representation on the right shows a sea urchin embryo at early blastula stage (~ 10 hrs postfertilization). *delta* is expressed in the skeletogenic lineage of the sea urchin (pink cells), while *gcm* and *pks* will be turned on in the mesodermal lineage (blue cells). Arrows indicate Delta signaling. The dashed lines in the timecourse diagram indicate optimal timepoints for sampling in perturbation experiments. (B) Amplification plot of a typical QPCR run with cDNA generated from delta-MO-injected embryos and control-MO-

injected embryos. If the same amount of cDNA is used for both treatments the amplification curves for the internal control (ubq – ubiquitin, orange/yellow) are almost identical. The amplification curve for *gcm* (blue) from perturbed embryos crosses the threshold (horizontal, red line) several cycles after the control. The difference, calculated as the ddC_t (see Box 1.1) indicates severely reduced transcription of *gcm* (~ 30 fold). (C) Results from *delta* and *gcm* MO knockdown experiments. Transcript abundance of *gcm* and *pks* was quantified at 16 hrs and 24 hrs postfertilization. Compared to the control, transcription of these genes is downregulated several-fold, as indicated by the ddC_t (colored bars). The difference in transcript level between control-MO-injected and uninjected embryos is marginal (white bars; data from http://sugp.caltech.edu/endomesoderm/qpcr.html, and SCM, unpublished). (D) Network reconstruction based on the data in C. Initially the *gcm* gene is activated in the mesodermal lineage by Delta signaling from the skeletogenic lineage (16 hrs). At 24 hrs the perturbation data indicate autoactivation of Gcm. Thus, Delta is not required any longer for sustained expression of *gcm*. Gcm also activates transcription of pks. The input of Delta into *gcm* via Notch/Supressor of Hairless and of Gcm into *pks* has been confirmed at the *cis*-regulatory level[66]. Network diagrams were created with the BioTapestry software (www.biotapestry.org).

Perturbation methods

The goal of a perturbation experiment is to alter either the endogenous transcription level of a regulatory gene or its function to query its role in the activation, or repression, of downstream genes. The most important factors influencing the success of a perturbation experiment are the effectiveness of the perturbation and its specificity. Although striking effects can be achieved by application of common perturbation agents like zinc sulfate, an animalizing agent in sea urchins, these experiments are generally not gene specific and may have multiple effects that are difficult to control. For network building it is thus preferable to use gene-specific perturbation agents that cause more targeted disruptions. The most essential techniques to query the regulatory wiring are knockdown, overexpression, and ectopic expression experiments that significantly alter transcript abundance or differently localize transcripts of the gene(s) of interest. A multitude of different perturbation methods exists, ranging from genetic mutants, stable knockouts, and expression of transgenes, to the use of shRNA. The choice will depend on the availability in the organism of interest. Table 1.1 summarizes the

most commonly used perturbation methods in the sea urchin; all methods are applicable in a wide range of systems beyond the sea urchin.

Gene knockdown:

Where perturbation agents can be delivered by microinjection (as in sea urchins, Xenopus, and zebrafish) or electroporation (as in chickens), MOs provide high effectiveness and specifity. As MOs do not carry a negative charge on their nucleotide backbone, they bind a specific target sequence with high affinity (Summerton et al., 1997) and elude the endogenous RNA degradation enzymes. MOs can be designed to bind to the translation start site of an mRNA where they sterically inhibit initiation of translation, or may cover a splice junction leading to the loss of an exon. The missing exon should code for a domain with an important function, e.g., the DNA binding domain (Draper et al., 2001), or should cause a frame shift and premature stop of the protein. In either case, the function of the protein is lost, leading to developmental defects.

Gene expression of MO-injected embryos should be compared to embryos injected with a mock control, usually a random mixture of morpholinos (IUPAC sequence: N_{25}). As the concentration of MOs within any one specific sequence is very low (there are 4^{25} possible sequences) its effect on gene expression is essentially negligible. MOs usually slow down development and this mock control minimizes staging errors.

Injecting a control MO with several mismatches to the target can assess the specificity of a MO. Usually, four mismatches are enough to abolish the ability of the MO to bind to its target (Oliveri et al., 2006), however, care has to be taken to ensure that the altered MO does not bind to other targets specific to other genes. Alternatively, specificity of a morpholino may be demonstrated by rescuing the perturbation effects with injection of an mRNA in which the MO binding site has been mutated. However, such overexpression experiments are prone to artifacts (see discussion below). A third way to show that the effects of a gene knockdown are specific is to perturb expression of a gene with two different MOs, ideally a transcription-

blocking MO and a splice-blocking MO. If they cause the same phenotype, the effect is likely to be gene specific.

The efficacy is easiest to assess for splice-blocking MOs. Because the affected transcripts have lost the targeted exon, the abundance of the endogenous and the wrongly spliced transcripts can be examined using QPCR (Oliveri et al., 2008). For translation-blocking MOs the loss of the protein can be examined by immunostaining or Western blot (Angerer et al., 2001; Yaguchi et al., 2008). Alternatively, the targeted region (5' UTR/start codon/first 30 amino acids) can be fused in frame to GFP. Injection of mRNA transcribed from this construct will cause global expression of GFP; coinjection of the MO should abolish fluorescence (Oliver et al., 2001). As a rule, MOs are highly efficient and specific, if care is taken that their sequence is unique to the gene of interest and does not contain polymorphic sequences, which can be an issue for marine animals.

Table 1.1: Common Approaches to Perturb Gene Expression in Sea Urchins.

Perturbation	Method	Effects	Ref.
Gene-knockdown	Injection of morpholino antisense oligonucleotide complementary the translation start codon.	Blocks translation. No protein is produced.	Summer-ton et al., 1997
	Injection of morpholino antisense oligonucleotide complementary to donor or acceptor splice junction.	Blocks splicing. Non-functional protein is produced lacking a domain important for function.	Draper et al., 2001
Overexpression	Injection of synthetic, capped mRNA.	Global expression of a spatially restricted regulatory gene.	Revilla-i-Domingo et al., 2007
	Ectopic expression of regulatory genes under control of a tissue specific promoter.	Ectopic, but localy restricted, expression of a regulatory gene.	Rast et al., 2002
Disruption of function	Overexpression of cofactors with dom.-neg. function (e.g. interaction partners that prevent nuclearization or dimerization).	TF cannot exert its regulatory effects.	Logan et al., 1999; Oliveri et al., 2003
	Overexpression of recombinant transcription factor. Fusion of DNA binding domain, to an Engrailed or WRPW repressor domain turns TF into obligate repressor; VP16 activator domains turns TF into obligate activators.	Repressor domain recruites corepressor Groucho and causes repression of target genes. VP16 activator domain directly interacts with basal transcription apparatus.	Beh et al., 2007; Hall et al., 2002; Oliveri et al., 2003
Disruption of signaling pathways	Overexpression of dom.-neg. signaling pathway molecules. E.g. overexpression of the extracellular domain of signaling receptors depletes pool of signaling molecules.	Disrupts, or constitutively activates signal transduction pathways.	Sherwood et al., 1999
	Small molecule inhibitors, e.g. kinase inhibitors.	Disrupt signal transduction pathways.	Fernandez-Serra et al., 2004

Overexpression

To examine the effect of increased expression in vitro transcribed mRNA may be introduced into the egg or embryo. mRNA can be delivered by injection as in the case of sea urchins, or electroporated like in *Ciona* or chickens. In sea urchins the injected mRNA distributes throughout the whole cell prior to first cleavage, causing ubiquitous expression of the gene. This may cause severe phenotypes because it affects the whole embryo and not only the cells or tissue that endogenously express the target gene. A general drawback of this technique is that it may produce severe artifacts: If the injected mRNA levels are not carefully controlled, the resulting high concentration of transcription factors may cause binding to low-affinity binding sites, resulting in effects that are unrelated to the primary function of the gene. When starting mRNA overexpression experiments several concentrations should be applied, starting with endogenous levels. For reliable results the number of injected mRNA molecules should not exceed the physiological level by more than tenfold. Another assumption that is made in mRNA overexpression is that the regulatory gene in question can work in early development because it acts either independently of other genes or together with other factors that are (already) present. If the synthetic mRNA contains a specific sequence tag different from the endogenous transcript, this can be used for careful quantification of the injected mRNA and to determine its degradation rate. This may be essential to determine that at the time the gene is usually active, enough of the artificial mRNA is still present, and not yet lost due to degradation, to significantly alter the endogenous expression level. Although the mRNA will be translated once it is injected, the protein might not be able to exert its effects without certain cofactors that become available only later.

Genes are commonly overexpressed by artificially placing them under the control of a different, tissue-specific promoter. However, in sea urchins such expression constructs, after injection into the fertilized egg, integrate clonally and cause mosaic expression. As the expression domain may vary in size, this experiment is less well suited for quantitative evaluation. But the spatial effects on downstream genes can be easily assessed by in situ hybridization. This experiment is an excellent way to monitor the effect of ectopic expression in a specific tissue and may be invaluable for confirming network connections.

Disruption of function

A potentially very informative perturbation method is the use of recombinant proteins. Transcription factors may be turned into obligate repressors by fusing their DNA-binding domain with a repressor domain. e.g., the engrailed repressor domain (Li et al., 1999, Oliveri et al., 2002) or the WPRW domain (Beh et al, 2007). Following injection of the mRNA, the recombinant protein will bind to its usual binding sites but will now shut down transcription of the target gene(s). Both engrailed repressor domain and WPRW recruit the corepressor Groucho and lower the transcriptional level of target genes by more than 90%. Recombinant proteins may also be turned into mandatory activators using the VP16 activation domain (Coffman and Davidson, 2001). This domain interacts directly with proteins of the basal transcription apparatus (Hall and Struhl, 2002). Similar precautions as discussed above for mRNA overexpression need to be taken in order to avoid artifacts. DNA binding domains in general, and the above-mentioned activator and repressor domains, are well characterized at the structural level. Mutant versions in which any of these functions are abolished can be used as a control. Use of obligate repressors is informative in connection with mRNA overexpression experiments: If the perturbed gene is a repressor both experiments will yield similar results, otherwise they will have opposing effects[8,44].

Disruption of signaling pathways

Dominant-negative forms of proteins of signaling pathways are an effective way to disrupt cell-to-cell communication. For example, overexpression of a signaling receptor lacking the intracellular domain may lead to disruption of signaling by out-competing the functional receptor for the ligand (Sherwood et al., 1999). Other successful strategies may involve prevention of relocation, as in the case of beta-catenin, by overexpression of n-cadherin in early endomesoderm development in sea urchins (Logan et al., 1999). Signaling pathways often involve cascades of phosphorylations or similar modification of pathway members. Another

perturbation method is the use of small molecule inhibitors that specifically interfere with these reactions, for example kinase inhibitors (Fernandez-Serra et al., 2004). However, the use of these perturbation agents can be problematic because it is difficult to demonstrate their specificity.

Quantitative analysis

The most critical step in network building is the accurate evaluation of perturbation experiments, for it indicates between which genes regulatory links exist. As a regulatory gene activates, or represses, other genes, the effect of a perturbation can be monitored directly at the transcriptional level. In principle this can be done using qualitative techniques like (whole-mount) in-situ hybridization. However, perturbation experiments do not necessarily produce dramatic loss or ectopic expression phenotypes that are easily observed. Effects may be modest but still indicate a significant input, e.g., when knocking down a modulator of transcription. Thus, it is generally favorable to monitor perturbation effects in a quantitative way. In situ hybridization is an invaluable tool to confirm perturbation effects in sea urchins and on occasion may be preferable over quantitative analysis, in particular when ectopic expression of downstream genes is observed as a consequence of a perturbation. In other experimental systems in situ hybridizations may play a more prominent role: In *Drosophila*, for example, the *eve* stripes of the early embryo are under control of different regulatory genes and knockdown of upstream regulators may affect only parts of the *eve* pattern (Clyde et al., 2003); transcript abundance on a per-embryo basis is affected only slightly, and thus is uninformative. Although some genes are subsequently expressed in different tissues in the sea urchin, such complicated expression patterns are rare and usually confined to later stages of embryonic development. As perturbation agents can be delivered only to the fertilized egg, this protocol is limited to early sea urchin development, and thus generally escapes such issues.

Several RNA quantification methods exist but the one most suited for evaluating perturbation experiments is QPCR[27]. QPCR is an extremely sensitive, kinetic method in which abundance

of a specific transcripts is measured by PCR amplification using whole-embryo cDNA as starting material, while monitoring the increase of PCR products with an intercalating fluorescent dye (Fig. 1.2B). The cycle number (C_t) at which the fluorescence crosses a chosen threshold during the exponential phase of the amplification is indicative of the amount of starting material. To quantify the effect of a perturbation, RNA from perturbed and unperturbed embryos is extracted, and expression levels of network candidates are compared. As the quantification by QPCR is sensitive to minute differences in overall transcript abundance, the C_t values for the genes of interest are normalized using an internal, invariant standard (Fig. 1.2B). The value obtained from perturbed embryos can then be subtracted from that of the control. The cycles-difference (ddCt) between the two populations is indicative of the change caused by the perturbation and can be converted into fold-changes (see Box 1.1 for formulae). Suitable internal standards for the sea urchin are housekeeping genes as ubiquitin (Nemer et al., 1991; Oliveri et al., 2002) . Ubiquitin is expressed at a constant level in early sea urchin development by quantification relative to 18S ribosomal RNA and SpZ1 (Revilla-i-Domingo et al., 2007); SpZ12 has been quantified previously using a highly accurate excess titration method (Wang et al., 1995). Although 18S RNA is also a suitable standard, its expression level is several orders of magnitude higher, requiring 1000-fold dilution of the cDNA to fall into the dynamic range of commonly available QPCR reagents. As a rule of thumb, amplification curves should cross the threshold between 15 and 30 cycles. After 32 cycles the variance between technical repeats increases dramatically and thus decreases the confidence in the accuracy of the obtained values.

Several precautions need to be taken when using QPCR. Dyes commonly used in reagent mixes, such as SYBR Green, are not sequence specific and care must be taken that the amplified products are gene specific. Gene-specific alternatives like TaqMan Probes (Applied Biosystems) exist, but their cost may be prohibitive when working with a multitude of genes. Furthermore, to guarantee the accuracy of the calculated perturbation effects, the efficiency of the PCR primers must be determined (QPCR methodology has been discussed elsewhere [Wong et al., 2005]).

A suitable alternative to QPCR may be RNA microarrays (Baugh et al., 2005). An advantage of such arrays is that usually many hundreds or thousands of genes can be queried in a single experiment, enabling genome-wide monitoring of transcription levels. However, this comes at a price, namely a loss in sensitivity compared to QPCR, and thus a less reliable quantification. This is significant because transcription factors are usually present at very low levels; the typical level of an actively transcribed transcription factor lies between a mere hundred and a few thousand transcripts per embryo (from few to hundreds of transcripts/cell) (Howard-Ashby et al., 2006 a,b; Materna et al., 2006; Rizzo et al., 2006; Tu et al. 2006). Another new, powerful, hybridization-based method for multiplex quantification of transcripts may also prove to be adequate: The Nanostring nCounter is a direct RNA sampling method that quantifies transcripts without requiring any enzymatic amplification steps and has been shown to have sensitivity similar to QPCR (Geiss et al., 2008). Pyrosequencing methods should in principle also be suitable for quantitative analysis[31], but the price at present makes them prohibitive as a routine method. Nevertheless, pyrosequencing methods might be employed as a discovery method where genes that may be affected by a certain perturbation are unknown.

Assembly of a provisional network map

The assembly of GRNs is a reiterative process. Each additional perturbation experiment yields new insights that help refine the architecture of the network. This in turn influences which perturbation experiments will have to be conducted. As work in progress, the GRN may have to include potentially superfluous links until further information has been obtained. Ideally, a GRN constitutes a complete map of the regulatory interactions as they are encoded in the genome and executed during development. However, perturbation experiments do not directly prove regulatory inputs but indicate where such inputs may exist; they may then be validated by *cis*-regulatory analysis and the identification of functional binding sites of regulatory factors. Careful interpretation of perturbation experiments will produce a provisional architecture that minimizes the number of false or superfluous links.

Before embarking on a large-scale network-building project, a significance threshold needs to be determined for perturbation experiments that indicates a regulatory input. Transcriptional activators rarely function in an all-or-none manner. Thus, the loss of a regulator may cause a several-fold downregulation of a downstream target gene, but rarely will transcription be lost completely. For the sea urchin a ddC_t of 1.6 cycles, corresponding to a threefold change, has proven adequate[4]. Initially this threshold was arrived at empirically, by validation of a few, known, direct regulatory inputs. It is conservative and may cause some more subtle inputs to be missed in favor of minimizing the number of false positives. The sea urchin system is non-clonal and experiments are performed on wild-type individuals, thus biological variation is considerable. It is reasonable to assume that variation is lower in experimental systems in which the genetic background can be controlled. But the appropriate level of significance must be established for each experimental system by quantitative evaluation of known regulatory interactions.

Where the perturbation data indicate a significant effect, the proposed input must conform to basic transcriptional kinetics. In general there is a lag between subsequent gene activations due to the time required for accumulation of sufficient transcription factor protein (Bolouri et al., 2003; Revilla-i-Domingo et al., 2007). Regulatory genes often activate other regulators before they themselves have reached peak concentrations (Bolouri et al., 2003). In fact, once their concentration plateaus they might already be dispensable, because the regulatory state of the cell has progressed and does not require their presence any more to be sustained. Such kinetic consideration in connection with an appropriate significance threshold may help discriminate between direct and indirect links.

The amount of information that has to be taken into account during assembly of regulatory networks is considerable. Because the sheer volume of data may easily become overwhelming, some computational tools are available to help with maintenance and interpretation of perturbation data and visualization of the results. The most appropriate tool for GRNs is the program BioTapestry (Longabaugh et al., 2005). It was designed specifically to represent gene regulatory networks. In addition to its usefulness as a visualization tool, it aids in interpreting perturbation data and suggests alternative network architectures that in turn may be verified by

additional experiments. However, any computational tool is only as good as the data and the assumptions that are made by the user. Despite their great potential, in particular in association with prediction of transcription factor binding sites, computational methods for the assembly of GRNs are still in their infancy.

Network revision and cis-regulatory analysis

The effect of perturbing a regulatory gene is often strongest for its direct downstream targets. However, perturbation analysis is limited in its ability to distinguish between direct and indirect targets. A case in point in the coherent feedforward loop, a motif that is ubiquitous in GRNs. In this linkage pattern gene A activates gene B, and both A and B are required to activate a third gene, C. Perturbation analysis will likely produce A → B and B → C, but fail to recognize that A is a direct input into C. In lieu of a solution to this problem, the linear arrangement of genes A, B, and C may be desirable for the sake of parsimony, but this limitation of perturbation experiments has to be kept in mind. The solution may lie in double perturbation experiments and undoubtedly in *cis*-regulatory analysis. Other network motifs (e.g., incoherent feedforward loops) pose similar problems (Smith, 2008; Alon, 2007).

In many situations the perturbation data may not favor one particular network architecture over another, for example, where two genes lock down their expression by forming a positive feedback loop. It may not be possible to decide from knocking down either one which of the two is the activating input into downstream genes. In the example in Figure 1.1D, genes A and E form a positive feedback loop. If expression of gene A is eliminated, the transcriptional level of gene E will be reduced, and vice versa. The result is the loss of A and E that, in turn, will cause a decline of B transcription. These experiments do not reveal whether A, or E, or both, are direct activators of B. This situation is not hypothetical since such feedback loops are common network motifs. To resolve the architecture, mRNA for either one of the genes could be co-injected when knocking down the other. If only one factor is a required input, transcription of B will be rescued. However, the ectopic effects of the mRNA injection may be

severe and make the experiment uninformative. In this case the architecture can again only be resolved by *cis*-regulatory analysis of gene B.

Some situations may require evaluation of perturbation by in situ hybridization rather than QPCR; for example, in situations where a gene is activated by ubiquitous inputs but repressed by localized transcription factors (Revilla-i-Domingo, 2007). Knockdown of the repressor will lead to ectopic expression. However, if the gene of interest is already expressed in more than 50% of the embryo, ectopic expression in the remainder of the embryo may not cause a significant change in overall abundance of the mRNA. Of course, the converse experiment, i.e., overexpression of the repressor, may provide an answer. But for reasons discussed above, the results of this experiment may not be interpretable. In situ hybridization is the ideal experiment to observe such dramatic spatial effects invisible to quantitative analysis

The following protocol emphasizes planning, conducting, and evaluating perturbation experiments to reveal the architecture of a GRN. As has been pointed out above, identification of the *cis*-acting, regulatory sequences is an integral part of the building and confirming the map of a GRN. However, we do not deal with *cis*-regulatory analysis in detail here as a detailed protocol has been published elsewhere (Smith, 2008). The present protocol is based on the procedure used for building the network-covering endomesoderm specification in sea urchins (Davidson et al., 2002a, b; Revilla-i-Domingo et al., 2007). It is generally applicable to any developmental process that relies primarily on transcriptional regulation. The protocol makes use of standard molecular techniques such as polymerase chain reaction that can be applied without modification. Where specific techniques are required, they are referenced (if not described) here. It is assumed that prospective users are familiar with microinjection methods for sea urchin embryos. These are described in detail elsewhere (Angerer et al., 2001; Jaffe and Terasaki, 2004).

Materials

Reagents

- Sea urchins (for a list of commercial suppliers see list in Bottger et al., 2004)
- Morpholino antisense oligonucleotides, custom designed (see reagent setup)
- mMESSAGE mMACHINE high-yield capped RNA transcription kit (Applied Biosystems/Ambion, cat. no.AM1340)
- RNeasy Micro Kit – RNA extraction kit (Qiagen, cat. no. 74004)
- Superscript III RT, reverse transcription kit (Invitrogen, cat. no. 18080-085)
- SYBR GreenER qPCR SuperMix, quantitative PCR reaction mix (Invitrogen, cat. no. 11760-02K)
- Smart RACE cDNA Amplification Kit (Clontech, cat. no. 634914)
- Whole-mount-in-situ reagents (for a comprehensive protocol see Ransick et al., 2004).
- Access to genomic information (http://spbase.org, www.genboree.org for sea urchin genomic data; other organism specific databases)
- perturbation agent injection solution (see Reagent setup)

Equipment

- Standard microinjection station (Narishige) mounted on Axiovert 25 inverted microscope (Zeiss) with Picospritzer II injection device (Parker Instrumentation)
- Axioskop 2 plus microscope with Plan-NEOFLUAR x10 and x20 objectives. Image capture with Zeiss AxioCam MRm using the Zeiss AxioVision software
- Quantitative PCR machine (ABI Prism 7900 HT, Applied Biosystems)
- BioTapestry, network assembly and analysis software (www.biotapestry.org)

Reagent setup

MO are designed by GeneTools on request (www.gene-tools.com). For translation-blocking MO, the MO should cover, or lie just upstream, of the translation start site; splice-blocking MO should cover the splice donor site of the targeted exon. Care should be taken that MO do not cover known polymorphisms.

MO injection solution: Defrost MO stock and heat to 60ºC for ≥ 10 min. Injection solution should have a concentration of no more than 300 μM MO in 120 mM KCl solution (injection volume into the egg is ≤ 10 pl). Centrifuge solutions at ≥ 10,000 rpm for 10 min and keep at RT until use. Store leftovers at 4°C.

QPCR primer design: QPCR primers should be designed for a unique stretch of sequence using a common primer design program (e.g., Primer3, http://frodo.wi.mit.edu/). Care should be taken that primers do not cover polymorphisms where such information is available. Primers should be optimized for standard QPCR conditions (60°C annealing/extension) and produce an amplicon of about 110–150 bp. It is critical to assure that amplification efficiency is high. A discussion of how to determine amplification efficiencies can be found elsewhere (Wong et al., 2005). As a shortcut, when QPCR primers do not span introns, new primer pairs can be tested using genomic DNA. The amplification curve should cut the threshold at a C_t value comparable to known, efficient primers.

RNA injection solution: Adjust RNA to desired concentration (usually max 100,000 molecules/10 pl injected volume) in 120 mM RNAse-free KCl solution. Centrifuge solutions at ≥10,000 rpm for 10 min and keep on ice until use.

Procedure

A general outline of the procedure is given in Fig. 1.1. An example of how to build a mini-network of three mesodermal genes in sea urchins is presented in Fig. 1.2.

1| Define the developmental event of interest. Use a priori knowledge to delimit the system in space and time. What cell types, cell lineages, territories, or tissues are involved? Are cells specified autonomously? Do they have known inductive ability? Are there any known signaling events with important developmental function? What regulatory genes are known to be active in the time window in which regulatory processes are executed? Which interactions between regulatory genes are known, or implied? Which molecular markers are known?

2| Summarize the available embryological information in a process diagram and establish a preliminary layout of the regulatory network. The process diagram should provide a description of embryological information without giving a mechanistic explanation. A simple example of a process diagram is given in Fig 1.1B.

3| Based on the process diagram, prepare a draft of the GRN using the BioTapestry editor (Longabaugh et al., 2005). Populate the region(s) with all known regulatory genes and link them according to their known interactions (The network diagrams in Fig. 1.1 were created using BioTapestry. An excellent tutorial on how to use BioTapestry can be found on the BioTapestry website: www.biotapestry.org).

4| Assemble a list of all transcriptional regulators and signaling molecules by extracting this information from genomic databases (for the sea urchin: Spbase, www.spbase.org; for others: databases at NCBI, www.ncbi.nlm.nih.gov/Genomes/, and organism-specific databases). If the genome is not yet available this information needs to be extracted from sequence collections (e.g., EST libraries) or genes need to be identified in RNA subtractions screens (for a detailed protocol see Rast et al., 2000).

5| Gather temporal expression data for all regulatory genes. These may be available from genomic databases. Identify all the regulators that are active during the time period and the cells/tissue under consideration.

6| Where necessary, complete the list by determining the transcription level for genes where such information is missing. QPCR is the most sensitive, and flexible, technique and is recommended. However, in principal other methods like Northern blotting could also be used, although they are labor intensive and less suitable for large numbers of genes.

CRITICAL STEP: QPCR reagents like SYBR Green are not sequence specific. Care has to be taken that only specific sequences are amplified. Minimally, this requires separating the PCR products on a gel to confirm that only one band of the expected size is observed. Most QPCR machines can collect denaturation data after the amplification phase of the PCR program is completed. The melting curve should indicate if only one specific product has been amplified. Denaturation data should be collected routinely to monitor solutions for contaminations.

CRITICAL STEP: PCR products never quite duplicate during each PCR cycle, even under optimal conditions. To ensure that the data for different genes are comparable, care must be taken that the QPCR primers are similarly efficient. Ideally PCR primers should have an amplification efficiency of about 1.95. A discussion of how to determine amplification efficiencies can be found elsewhere (Wong et al., 2003).

7| Gather spatial expression data for the genes where temporal expression data indicates active transcription. Spatial expression data may also be available from genomic databases. Where no information is available, spatial expression data will have to be examined by whole-mount-in-situ hybridizations. Detailed protocols describing this technique for sea urchins have been published elsewhere (Minokawa et al., 2004; Ransick, 2004). (Protocols for most established model systems can be found in the literature). Genes expressed in the region covered by the GRN might have an important regulatory role in the developmental event under consideration and are network candidates.

8|For all network candidates, collect temporal and spatial expression data in short sampling intervals. This high-resolution data set establishes the progression of regulatory states by allowing genes be ordered according to their onset (and termination) of expression. For the sea urchin, appropriate intervals for data collection are 1 hour for temporal, and 3 hours for spatial expression data. In Fig. 1.2A an example of high-density timecourses are given for three genes expressed successively in the skeletogenic and mesodermal cells of the sea urchin.

9| Incorporate the spatio-temporal data into the network model. The BioTapestry program allows importing expression data from simple comma-separated value files (Longabaugh et al., 2007). A step-by-step guide can be found on the website www.biotapestry.org.

10| All regulatory factors that are expressed significantly (i.e., with more than ~ 10 copies per cell) should be targeted in perturbation experiments. For these genes, full-length cDNAs should be isolated. This will reveal the transcription start site and, if not already known, indicate the translation start site. It may also identify potential splice forms. Such information is required for designing MOs. A subcloned cDNA is convenient when preparing mRNA overexpression or dominant repression constructs. For identification of cDNAs, use standard rapid-amplification-of-cDNA (RACE) PCR (e.g. using the Clontech Smart RACE cDNA amplification kit) or conduct cDNA library screens (Rast et al., 2000).

11| Plan perturbation experiments based on the high-density data. Determine sampling timepoints that maximize the information to be obtained from an experiment. Fig. 1.2A illustrates this for three sea urchin genes. *delta* is the first gene to be active, followed by *gcm* and *pks*. To query the role of Delta in *gcm* activation a good timepoint for sampling is 16 hrs postfertilization of delta-MO-injected embryos; to examine the role of Gcm in *gcm* and *pks* activation 24 hrs post fertilization is optimal (Fig 1.2A, and see discussion above).

12| Obtain MOs for the genes to be knocked down. When using translation-blocking MOs, clone the 5'-UTR and up to about 30 amino acids of the coding region in frame into a GFP expression construct; this will be used as a positive control.

13| Inject the MO alone and together with mRNA transcribed from this expression construct. The MO should block translation of the mRNA and abolish fluorescence.

14| CRITICAL STEP: Before conducting any perturbations, determine the significance threshold for perturbation experiments. Conduct test-perturbation experiments where known regulatory interactions are queried. These may be taken from the literature but should be based on *cis*-regulatory analysis. Quantify the effects and determine what constitutes a significant change. When picking a threshold it is important to take the level of biological variance between repeat experiments into consideration. Although it is conservative, a threshold of ddC_t 1.6, corresponding to a threefold change, has been proven adequate in sea urchins. This threshold minimizes false positives but is certain to cause some connections to be missed.

15| Perform perturbations by injecting MO or mRNA.

CRITICAL STEP: The concentration at which MOs are effective varies. Generally, the effective working concentration ranges from 50–300 fmol/embryo (10–15 mole/embryo; this corresponds to injection of 10 pl of a 300 μM MO solution in an egg with a diameter of about 80 μm). At higher concentrations nonspecific side effects can be observed, e.g., a retardation of development. When injecting eggs of a different size (e.g., star fish or zebrafish) the concentration of the injection solution must be adjusted to yield a similar in ovo concentration[64].

CRITICAL STEP: When injecting mRNA, start by injecting an amount of molecules that matches the endogenous expression (on a per cell basis). Increase the concentration in subsequent experiments but do not exceed the endogenous concentration by more than tenfold.

16| Characterize the phenotype of the perturbation. Assert that the phenotype is caused by specific action of the MO. Demonstrate that mRNA injection rescues the phenotype, or demonstrate that a second MO causes the same phenotype.

CRITICAL STEP: Make sure that the effects of the perturbation are specific. E.g., if the perturbed gene is not expressed before late gastrula stage, confirm that early stages of development proceed normally.

17| Collect embryos and extract RNA with a standard RNA extraction adequate for the total amount of RNA of the sample kit (e.g., Quiagen RNAeasy Micro Kit). Although the amount of RNA from expressed genes will significantly exceed contaminating genomic DNA, it is desirable to minimize the amount of gDNA by treatment with DNAse. Generate cDNA using a high-efficiency reverse transcription kit (e.g., Invitrogen, iScript III).

18| Quantify the expression level of all genes that may be affected by a perturbation using QPCR. It is a good idea to include genes from cells/territories that should not be affected by the perturbation as a control. All QPCR reactions should be carried out in triplicate.

CRITICAL STEP: When using SYBR Green QPCR chemistry take care to not cross contaminate wells with different primers. SYBR Green is a non-specific intercalating DNA dye that will indicate accumulation of PCR product, independent of its sequence.

CRITICAL STEP: It is important to stay within the dynamic range of the QPCR reagents. In general, the amplification curve should cross the chosen threshold between cycles 15 and 30. After 32 cycles the fidelity of the PCR decreases and the variation between repeat wells increases dramatically. Determine the right amount of starting material. For sea urchins the RNA equivalent of one to two embryos (~ 3 – 5 ng of total RNA) is sufficient to obtain accurate and reproducible data. When using SYBR Green ensure that the quantitative data are reliable and comparable (see discussion above).

19| Evaluate the QPCR results. Calculate the dC_t value for the genes of interest using an invariant, internal control (for the sea urchin, e.g., ubiquitin). Calculate ddC_t values using perturbed and control embryos (see Box 1.1). In Fig. 1.2B the amplification plot of a QPCR run with *gcm* primers using delta-MO-injected and control-injected embryos is shown. In the perturbed embryos the *gcm* expression level is dramatically lower, as indicated by the shift of the amplification curve to the right. A ddC_t value of ~ 5 cycles corresponds to a roughly 30-

fold difference in expression level. Similarly, injection of a *gcm* MO causes as strong a downregulation of *pks* as of *gcm* itself (Fig. 1.2C).

CRITICAL STEP: To be certain that the injection itself does not cause artifacts, it is advisable to also compare control-injected embryos to uninjected embryos. ddC_t values calculated from these controls should not show any significant changes (Fig. 2c).

20| Confirm perturbation results by examining the effects on spatial expression by whole-mount in-situ hybridization. Although *in situ* RNA detection methods are not quantitative, highly significant effects usually produce recognizable changes in spatial expression.

21| Where the perturbation data indicate regulatory inputs, draw a link between the genes in the network model. Check that the resulting map is in agreement with all the data. Long delays between subsequent activations may indicate that the activation (or repression) is indirect. In the example in Fig. 1.2. quantitative analysis of *pks* expression at 24 hrs would indicate a loss in delta-MO-injected embryos. However, the long time between the start of *delta* and *pks* expression makes this unlikely; and indeed Gcm is an intermediate regulator. QPCR results can be imported into the BioTapestry software for automatized evaluation and network building; detailed instructions on how to import data can be found on the BioTapestry website.

22| Once the network map grows in complexity, find perturbation experiments that have not been conducted. Check if the network makes predictions about the outcome of the proposed experiment. Conduct the experiment and compare the results to the predictions. If these are not in agreement, revise the network architecture or, if necessary, acquire additional data. In the simple example in Fig. 1.2, if a Delta MO perturbation had been conducted, the loss of *pks* at 24 hrs might be attributable to loss of *gcm* expression, as might be suspected due to the long time between initial *delta* and *pks* activation. Perturbation of *gcm* tests this prediction.

23| Where the network architecture remains conflicting or ambiguous, more perturbation data are needed. Design specific perturbation experiments to distinguish between different topologies. This may require double perturbations combining different methods from Box 1.1.

24| Proceed to *cis*-regulatory analysis to confirm the provisional regulatory architecture at the DNA level. *cis*-regulatory analysis will resolve ambiguities that could not be worked out with perturbation analysis alone. *cis*-regulatory analysis is treated in detail elsewhere (Smith, 2008).

Timing

To assemble the map of a GRN is no simple task. The exact timing depends on the complexity of the system under examination. It is possible to establish the regulatory linkages between a few well-characterized regulatory genes in a few weeks. However, where extensive identification of regulatory factors is necessary, the time needed to arrive at a complete map will be considerably longer. To establish the GRN for a whole tissue, organ, or embryo may easily take months to years to complete, as evidenced by the sea urchin gene regulatory network.

Troubleshooting

Troubleshooting advice can be found in Table 2.2.

Average Ct for n replicates	$Ct_{avg} = \frac{Ct1+Ct2+\cdots Ct_n}{n}$
dCt calculation	$dCt = Ct_{goi}\text{-}Ct_{int.\text{-}std.}$
ddCt calculation	$ddCt = dCt_{cont.} - dCt_{pert.}$
Conversion to fold change	$\text{fold change} = \text{Primer Efficiency}^{(ddCt)} \approx 1.95^{(ddCt)}$
Copy number	$N = N_{int.\text{-}std.} 1.95^{(dCt)}$

Standard deviation of Ct:

$$SD_{Ct} = \sqrt{\frac{(Ct_1\text{-}Ct_{avg})^2 + (Ct_2\text{-}Ct_{avg})^2 + \cdots + (Ct_n\text{-}Ct_{avg})^2}{n}}$$

Standard deviation of ddCt:

$$\sqrt{(SD_{ddCt_{(goi,cont))}})^2 + (SD_{ddCt_{(int.std,cont))}})^2 + (SD_{ddCt_{(goi,pert))}})^2 + (SD_{ddCt_{(int.std,cont))}})^2}$$

Box 1.1. Formulae for evaluation of QPCR experiments: C_t – Cycle number at QPCR threshold, avrg. – average, goi – gene of interest, int. std. – internal standard (e.g., ubiquitin), cont. – control embryos (unperturbed), pert. – perturbed embryos (MO, MOE, etc. treated), SE – standard error. Although in theory the amount of material should double during each PCR cycle, in practice this value is lower. Efficiencies of the PCR primers vary and need to be determined individually. For a working primer pair a value of 1.95 is usually a good approximation.

Anticipated Results

The network building process, as outlined here, will produce a detailed map of the regulatory functions as they are encoded in the genome and executed during development. It will give a detailed explanation as to why certain observed phenomena occur. The map will be highly resolved in time and space and make detailed predictions that can be confirmed at the DNA level. Because regulatory genes can be identified exhaustively, the network map could, in theory, contain all links between all regulatory genes, and thus reach completeness. However, not all perturbation experiments may be feasible or interpretable, and, while the protocol aims at minimizing false positives, some false negatives are certain to remain.

Despite these limitations, the map of a GRN will usually provide an accurate picture of the regulatory code as it is hard-wired into the genome. For example, in the GRN covering endomesoderm development of the sea urchin, where *cis*-regulatory analyses have been conducted about 90% of the predicted connections were confirmed (Minokawa et al., 2005; Ransick and Davidson, 2006; Revilla-i-Domingo et al., 2004; Yuh et al., 2004). Thus, a GRN in an advanced state will have enormous explanatory and predictive power. The network map will provide a mechanistic explanation of the regulatory processes executed in development and will reveal what drives development forward.

Acknowledgements

We would like to thank Eric Davidson, in whose lab this protocol was established, for many interesting and insightful discussions. Many thanks also to Jonathan Rast, Isabelle Peter, Joel Smith, and Andy Cameron for their help in preparing this manuscript. We are grateful for helpful suggestions from our anonymous reviewers. This research was supported by NIH grant HD-37105.

Table 2.2: Troubleshooting

Step	Problem	Solution
6	QPCR produces unspecific bands.	Primer solutions, SYBR Green are contaminated. Use fresh solutions. Primers are unspecific. Redesign primers and take care they are unique by checking the genomic sequence for unspecific binding.
	QPCR melting curves show multiple peaks.	Primers form dimers. Redesign primers and increase stringency for self-complementarity.
	QPCR primers do not work.	Double check correctness of sequence. Assert that primers do not lie in polymorphic regions if information on polymorphisms is available.
13	MO does not abolish fluorescence in positive control.	MO was designed to a falsely mapped translation start site. Examine the mRNA sequence for alternative translation start sites.
15	Splice-blocking MO does not cause loss of exon.	Ensure that the exon/intron boundary has been correctly mapped. Use a different splice- or translation-blocking MO.
15/16	MO causes abortive development.	Target gene may be maternally expressed and loss of early expression phase causes abortive development. Coinject MO with an mRNA in which the binding site has been mutated. On rare occasions MOs exhibit genuine cytotoxicity. Replace the MO.
19	Genes that are included as positive controls are affected by the perturbation.	Assert that the staging is right. Severe perturbations may delay development severely. Evaluate perturbation effects only soon after the knocked down gene becomes activated. Late effects may be only secondary.
	Discrepancy between mock injected and uninjected embryos.	Assert that the staging is right. Occasionally individuals may be more sensitive to presence of MOs and show a more pronounced delay in development.

References

Alon, U., 2007. Network motifs: theory and experimental approaches. Nat Rev Genet 8, 450.

Angerer, L.M. et al., 2001. Sea urchin goosecoid function links fate specification along the animal-vegetal and oral-aboral embryonic axes. Development 128, 4393.

Arbeitman, M.N. et al., 2002. Gene expression during the life cycle of *Drosophila melanogaster*. Science 27, 2270.

Barolo, S., Stone, T., Bang, A.G., Posakony, J.W., 2002. Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless. Genes Dev 16, 1964.

Baugh, L.R. et al., 2005. The homeodomain protein PAL-1 specifies a lineage-specific regulatory network in the *C. elegans* embryo. Development 132, 1843.

Beh, J., Shi, W., Levine, M., Davidson, B., Christiaen, L., 2007. FoxF is essential for FGF-induced migration of heart progenitor cells in the ascidian *Ciona intestinalis*. Development 134, 3297.

Ben-Tabou de-Leon, S., Davidson, E.H., 2006. Deciphering the underlying mechanism of specification and differentiation: the sea urchin gene regulatory network. Sci STKE 14, pe47.

Ben-Tabou de-Leon, S., Davidson, E.H., 2007. Gene regulation: gene control network in development. Annu Rev Biophys Biomol Struct 36, 191.

Bolouri, H., Davidson, E.H., 2003. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc Natl Acad Sci U S A 100, 9371.

Bottger, S.A., Walker, C.W., Unuma, T., 2004. Care and maintenance of adult echinoderms. Methods Cell Biol 74, 17.

Calestani, C., Rast, J.P., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. Development 130, 4587.

Clyde, D.E. et al., 2003. A self-organizing system of repressor gradients establishes segmental complexity in *Drosophila*. Nature 426, 849.

Coffman, J.A., Davidson, E.H., 2001. Oral-aboral axis specification in the sea urchin embryo. I. Axis entrainment by respiratory asymmetry. Dev Biol 230, 18.

Davidson, E.H. et al., 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246, 162.

Davidson, E.H. et al., 2002. A genomic regulatory network for development. Science 295, 1669.

Davidson, E.H., 2006. The Regulatory Genome: Gene Regulatory Networks in Development and Evolution. Academic Press, San Diego, CA.

Draper, B.W., Morcos, P.A., Kimmel, C.B., 2001. Inhibition of Zebrafish fgf8 Pre-mRNA Splicing With Morpholino Oligos: A Quantifiable Method for Gene Knockdown. Genesis 30, 154.

Fernandez-Serra, M., Consales, C., Livigni, A., Arnone, M.I., 2004. Role of the ERK-mediated signaling pathway in mesenchyme formation and differentiation in the sea urchin embryo. Dev Biol 268, 384.

Geiss, G.K. et al., 2008. The NanoString nCounter System: A Sensitive, Digital Technology for Direct Multiplexed Measurement of Gene Expression. Nature Biotechnology 36, 317.

Hall, D.B., Struhl, K., 2002. The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. J Biol Chem 277, 46043.

Henry, J.Q., Perry, K.J., Wever, J., Seaver, E., Martindale, M.Q., 2008. beta-Catenin is required for the establishment of vegetal embryonic fates in the nemertean, *Cerebratulus lacteus*. Dev Biol.

Hinman, V.F., Nguyen, A.T., Cameron, R.A., Davidson, E.H., 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc Natl Acad Sci U S A 100, 13356.

Howard-Ashby, M. et al., 2006. Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. Dev Biol 300, 74.

Howard-Ashby, M. et al., 2006. Gene families encoding transcription factors expressed in early development of *Strongylocentrotus purpuratus*. Dev Biol 300, 90.

Imai, K.S., Levine, M., Satoh, N., Satou, Y., 2006. Regulatory blueprint for a chordate embryo. Science 312, 1183.

Jaffe, L.A. Terasaki, M., 2004. Quantitative microinjection of oocytes, eggs, and embryos. Methods Cell Biol 74, 219.

Johnston, R.J., Jr., Chang, S., Etchberger, J.F., Ortiz, C.O., Hobert, O., 2005. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. Proc Natl Acad Sci U S A 102, 12449.

Koide, T., Hayata, T., Cho, K.W., 2005 *Xenopus* as a model system to study transcriptional regulatory networks. Proc Nat Acad Sci U S A 102, 4943.

Levine, M., Davidson, E.H., 2005 Gene regulatory networks for development. Proc Natl Acad Sci U S A 102, 4936.

Li, X., Wikramanayake, A.H., Klein, W.H., 1999. Requirement of SpOtx in cell fate decisions in the sea urchin embryo and possible role as a mediator of beta-catenin signaling. Dev Biol 212, 425.

Livingston, B.T. et al., 2006. A genome-wide analysis of biomineralization-related proteins in the sea urchin *Strongylocentrotus purpuratus*. Dev Biol 300, 335.

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

Longabaugh, W.J., Davidson, E.H., Bolouri, H., 2005. Computational representation of developmental genetic regulatory networks. Dev Biol 283, 1.

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

Materna, S.C., Davidson, E.H., 2007. Logic of gene regulatory networks. Curr Opin Biotechnol 18, 351.

Meulemans, D. Bronner-Fraser, M., 2004. Gene-regulatory interactions in neural crest evolution and development. Dev Cell 7, 291.

Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B., Davidson, E.H., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. Gene Expr Patterns 4, 449.

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

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621.

Nemer, M., Rondinelli, E., Infante, D., Infante, A.A., 1991. Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos. Dev Biol 145, 255.

Oliveri, P., Carrick, D.M., Davidson, E.H., 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev Biol 246, 209.

Oliveri, P., Davidson, E.H., McClay, D.R., 2003. Activation of *pmar1* controls specification of micromeres in the sea urchin embryo. Dev Biol 258, 32.

Oliveri, P., Walton, K.D., Davidson, E.H., McClay, D.R., 2006. Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo. Development 133, 4173.

Oliveri, P., Davidson, E.H., 2007. Development. Built to run, not fail. Science 315, 1510.

Oliveri, P., Tu, Q. Davidson, E.H., 2008. Global Regulatory Logic for Specification of an Embryonic Cell Lineage. Proc Natl Acad Sci U S A 105, 5955.

Range, R.C., Venuti, J.M., McClay, D.R., 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, 252.

Ransick, A., 2004. Detection of mRNA by in situ hybridization and RT-PCR. Methods Cell Biol 74, 601.

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

Rast, J.P. et al., 2000. Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Dev Biol 228, 270.

Rast, J.P., Cameron, R.A., Poustka, A.J., and Davidson, E.H., 2002. Brachyury Target genes in the early sea urchin embryo isolated by differential macroarray screening. Dev Biol 246, 191.

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

Revilla-i-Domingo, R., Davidson, E.H., 2003. Developmental gene network analysis. Int J Dev Biol 47, 695.

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

Revilla-i-Domingo, R., Oliveri, P., Davidson, E.H., 2007. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. Proc Natl Acad Sci U S A 104, 12383.

Samanta, M.P. et al., 2006. The transcriptome of the sea urchin embryo. Science 314, 960.

Sandmann, T. et al., 2007. A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. Genes Dev 21, 436.

Sauka-Spengler, T., Meulemans, D., Jones, M., Bronner-Fraser, M, 2007. Ancient evolutionary origin of the neural crest gene regulatory network. Dev Cell 13, 405.

Sherwood, D.R. and McClay, D.R, 1999. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. Development 126, 1703.

Smith, J., 2008. A protocol describing the principles of *cis*-regulatory analysis in the sea urchin. Nat Protoc 3, 1.

Stolc, V. et al., 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. Science 306, 655.

Summerton, J. Weller, D., 1997. Morpholino Antisense Oligomers: Design, Preparation, and Properties. Antisense Nucleic Acid Drug Dev 7, 187.

Tu, Q., Brown, C.T., Davidson, E.H., Oliveri, P., 2006. Sea urchin Forkhead gene family: phylogeny and embryonic expression. Dev Biol 300, 49.

Vokes, S.A. et al., 2007. Genomic characterization of Gli-activator targets in sonic hedgehog-mediated neural patterning. Development 134, 1977.

Wang, D.G., Britten, R.J., Davidson, E.H., 1995. Maternal and embryonic provenance of a sea urchin embryo transcription factor, SpZ12-1. Mol Mar Biol Biotechnol 4, 148.

Wikramanayake, A.H. et al., 2003. An ancient role for nuclear beta-catenin in the evolution of axial polarity and germ layer segregation. Nature 426, 446.

Wong, M.L., Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. Biotechniques 39, 75.

Yaguchi, S., Yaguchi, J., Angerer, R.C., Angerer, L.M., 2008. A Wnt-FoxQ2-Nodal Pathway Links Primary and Secondary Axis Specification in Sea Urchin Embryos. Dev Cell 14, 97.

Yuh, C.H., Bolouri, H., Davidson, E.H., 2001. *cis*-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control. Development 128, 617.

Yuh, C.H., Dorman, E.R., Howard, M.L., Davidson, E.H., 2004. An otx *cis*-regulatory module: a key node in the sea urchin endomesoderm gene regulatory network. Dev Biol 269, 536.

Chapter 2

THE C_2H_2 ZINC FINGER GENES OF STRONGYLOCENTROTUS PURPURATUS AND THEIR EXPRESSION IN EMBRYONIC DEVELOPMENT

Stefan C. Materna, Meredith Howard-Ashby, Rachel F. Gray, Eric H. Davidson

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

(Published in Developmental Biology 2006, 300:108–120)

Abstract

The C_2H_2 zinc finger is one of the most abundant protein domains and is thought to have been extensively replicated in diverse animal clades. Some well-studied proteins that contain this domain are transcriptional regulators. As part of an attempt to delineate all transcription factors encoded in the Strongylocentrotus purpuratus genome, we identified the C_2H_2 zinc finger genes indicated in the sequence, and examined their involvement in embryonic development. We found 377 zinc finger genes in the sea urchin genome, about half the number found in mice or humans. Their expression was measured by quantitative PCR. Up to the end of gastrulation fewer than a third of these genes are expressed, and about 75% of the expressed genes are maternal; both parameters distinguish these from all other classes of regulatory genes as measured in other studies. Spatial expression pattern was determined by whole-mount *in situ* hybridization for 43 genes transcribed at a sufficient level, and localized expression was observed in diverse embryonic tissues. These genes may execute important regulatory functions in development. However, the functional meaning of the majority of this large gene family remains undefined.

[Keywords: Zinc finger; Zf-c2h2; Transcription factor; Genome; Sea urchin; Development]

Introduction

Zinc finger motifs are of particular interest in developmental biology because they occur in some prominent transcriptional regulators. Though there are more than seventy classes of zinc binding motifs listed in the PFAM database, the specific transcriptional regulators fall mainly in the C_2H_2 zinc finger class, in which the zinc atom is complexed by two cysteines and two histidines, and in their structural relatives the C4 zinc finger class, in which the zinc is complexed with four cysteines. The latter group consists mainly of nuclear hormone receptors and GATA factors (Krishna et al., 2003). In most animal genomes that have been sequenced, C_2H_2 zinc fingers are among the more abundant protein domains. This applies in particular to mammalian genomes, in which C_2H_2 zinc finger genes have been highly multiplied (Lander et al., 2001; Rubin et al., 2000). The C_2H_2 zinc finger genes usually far outnumber the zinc fingers of the C_4 type, of which most genomes contain only a few dozen. A prominent exception is *Caenorhabditis elegans*, where the nuclear hormone receptors have undergone extensive multiplication (Reece-Hoyes et al., 2005) and outnumber C_2H_2 zinc finger genes. Although sequence-specific DNA-binding proteins can be found in other groups of zinc fingers, specific transcription factors are rare, and these proteins are often part of the basal transcription apparatus or DNA repair machinery.

C_2H_2 zinc finger proteins are commonly viewed as transcriptional regulators, but they may be widely used for RNA binding. This is exemplified by the first known zinc finger transcription factor, *Xenopus* TFIIIa, which binds specifically to both DNA and RNA (Lu et al., 2003). Possibly just as typical for genes with higher numbers of zinc fingers is the *Xenopus xfin* gene, which codes for a protein with 37 zinc fingers. It is localized in the cytoplasm and has been shown to bind to RNA (Andreazzoli et al., 1993). Transcriptional regulatory activity has not been demonstrated for this protein. In addition to DNA and RNA binding, zinc finger domains may also be used for protein-protein interactions (Laity et al., 2001). Several examples are known, including well-known transcriptional regulators. For example, two of the five conserved zinc fingers of the Gli family transcription factor encoded by the *cubitus interruptus* gene are needed for specific interaction with another factor, converting the protein into its

active form (Croker et al., 2006). C_2H_2 zinc finger genes are thought to account for 30% – 50% of all transcription factors in metazoan genomes (Adams et al., 2000; Ruvkun and Hobert, 1998). However, except for a minority which is clearly orthologous to known regulatory factors, there are no canonical criteria which suffice to distinguish those C_2H_2 proteins that are dedicated sequence-specific transcription factors from those that bind RNA or perform other functions.

Regulatory genes of the nuclear hormone receptor and GATA classes in the sea urchin genome have been characterized by Howard-Ashby et al. (2006a). Here, in order to encompass the major remaining class of transcription factors, we identify all C_2H_2 zinc finger genes predicted by the genomic sequence and determine their activity during development.

Materials and Methods

Identification of zinc finger genes

Zinc finger genes were identified in the contig assembly by searching for the C_2H_2 zinc finger motif. We built a calibrated Hidden-Markov model from the PFAM seed alignment (PF00096, www.sanger.ac.uk/Software/Pfam/) and searched the sea urchin genome with hmmsearch (http://hmmer.wustl.edu/), accepting only domains with an e-value < 0.1. After release of the gene predictions, the identified zinc finger genes were mapped onto the GLEAN models (Sea Urchin Sequencing Consortium, 2006) by finding near-perfect matches to the calculated QPCR amplicon. On genes that did not perfectly match a GLEAN model, a BLAST search was performed against the remainder of the GLEAN gene predictions. The results were manually inspected and associations validated. Less-than-perfect matches are due to incorporation of sequences of different haplotype in the scaffold assembly. Presumed zinc finger genes that did not match any GLEAN model were searched by BLAST against novel predictions from the whole genome tiling array (Samanta et al., 2006), identifying additional

gene models for genes that are expressed in early development. Gene models were aligned with the contigs using the Spidey genomic mapping program (Wheelan et al., 2001), and the match was validated through manual inspection.

Phylogenetic analysis

For identification of orthologous genes, we obtained the set of C_2H_2 zinc finger containing protein sequences of *Homo sapiens*, *Mus musculus*, *Ciona intestinalis*, *C. elegans*, *Drosophila melanogaster* and *Nematostella vectensis*. Sequences of *Nematostella* were obtained from Stellabase (www.stellabase.org), *C. elegans* sequences from Wormbase (www.wormbase.org, WSWS156), and all others from Ensembl (www.ensembl.org , v.37 – Feb2006) by motif search for Interpro domain IP:007087. To obtain a nonredundant set of proteins, we kept only the longest protein and discarded shorter isoforms for any given gene. A BLASTP search was performed for each sea urchin protein against this set of C_2H_2 zinc finger proteins. Good hits were confirmed by manual inspection. For such genes the zinc finger region together with surrounding conserved sequence was excised and aligned using the Mafft alignment program (Katoh et al., 2005) using 1000 iterations. Phylogenetic analysis was conducted using the neighbor-joining method with the MEGA program (Kumar et al., 2004). The calculated distance was Poisson corrected, gaps were pairwise deleted, and 1000 iterations were used for calculating bootstrap values.

Transcriptional profiling

We performed transcriptional profiling using quantitative PCR (QPCR). QPCR is a comparative method, in which the accumulation of PCR product is monitored for a gene of interest and in the same sample for a given standard, through the use of a double strand-

specific fluorescing dye. By choosing a threshold and determining CtΔ (the difference in cycle number at which each PCR reaction crosses the threshold) the initial prevalence of a gene can be calculated, since the cycle difference is proportional to the abundance in the original reaction mix (Wong and Medrano, 2005).

For primer design, the sometimes short stretch of sequence containing the zinc finger domains was extended using a BLASTX search against the NCBI database of non-redundant proteins "nr" (www.ncbi.nlm.nih.gov). QPCR primers lying within these regions were obtained using the standalone version of Primer3 (Rozen and Skaletzky, 2000). Primers were chosen to yield an amplicon of between 110 and 140 base pairs. Primers were tested for specificity on genomic DNA by QPCR. It was assumed that all genes dealt with were single copy. In this case, given equal amplification efficiency, all PCR products should accumulate to a given threshold at roughly the same cycle (Ct). Primer pairs that did not produce an acceptable Ct value (one cycle more or less compared to the mean Ct) were not used for transcriptional profiling and were redesigned. A more exact determination of primer efficiency was conducted for a representative set of primer pairs using serial dilutions (Wong and Medrano, 2005), and it confirmed the initial findings. We therefore generally assumed an amplification efficiency of 1.95. The presence of a single specific band was confirmed by gel electrophoresis.

Embryos were grown and harvested at fertilization, and at 6, 12, 18, 24, 36, and 48 h postfertilization. RNA was isolated with the Qiagen RNeasy Mini-Kit. RT reactions were performed with ABI (Foster City, USA) TaqMan cDNA synthesis kit according to manufacturer's instructions. QPCR was conducted on an ABI 7900 HT with ABI SYBR-Green reaction mix, using the following program: 1x (95°C – 10 min,) 40x (60°C – 30 s, 95°C – 1 min). At the end of each program a dissociation curve was collected to confirm that only one product accumulated during the reaction. A no- template control also assured that no primer-dimers had formed. The RNA copy number was determined by calculating the CtΔ for a given zinc finger gene with respect to the poly-ubiquitin gene, which was assumed to be represented by 88000 transcripts per embryo during the developmental stages examined (Nemer et al., 1991). On each plate, each primer-cDNA combination was run in triplicate. The experiment was repeated once with the same cDNA and twice with cDNA from a second

animal. For data analysis, the Ct's of wells that obviously did not amplify were omitted. CtΔ was calculated for the triplicates of the four independent runs. The mean of the averages from the four runs was used to calculate the number of transcripts per embryo. Error bars were calculated from the standard deviation on the mean.

Whole-mount in situ hybridization

In order to identify the spatial domain of expression for the higher expressed genes, we conducted in situ hybridization with digoxigenin-labeled antisense probes. From the general assumption that about ten RNA molecules per cell are needed for sufficient staining, it follows that a minimum of several hundred molecules per embryo is needed to obtain a clear stain, since spatially restricted regions like the endoderm contain no fewer than 60 cells prior to gastrulation.

For successful in situ hybridization using these methods, the probe should be a minimum of 600 base pairs in length. We attempted to obtain primer pairs by using either a conserved sequence that is recognizably located within one exon or, if no sequence of sufficient length could be obtained, by assuming the GLEAN gene models. Templates for in situ probes were amplified from cDNA using primers tailed with Sp6 and T7 promotors, or subcloned into the pGEM-T Easy vector, which contains Sp6 and T7 promotor sites adjacent to the multiple cloning site. Sequencing confirmed the identity of the gene. After digoxigenin labeling through in vitro transcription with Roche Sp6 or T7 polymerase, probes were run on a denaturing gel, confirming the size of the transcript. Whole-mount *in situ* hybridizations were carried out according to Minokawa et al. (2004).

Results

Identification of C_2H_2 zinc-finger-containing genes

We initially set out to identify all transcription factor genes using a BLAST-based approach, searching the trace archive of the sea urchin genome project for domains commonly associated with gene regulatory activity (Howard-Ashby et al., 2006a). Identified traces were clustered by BLASTX against our own reference database. However, this method failed for C_2H_2 zinc-finger-containing genes. Due to their tendency to contain multiple zinc finger domains, and the high degree of structural conservation between domains, trace sequences could not be binned unambiguously. Any two zinc finger domains on average are about 40% identical at the protein sequence level. The amino acids that are important for structural integrity, and a very conserved linker of six to seven amino acids that frequently connects zinc finger domains, are responsible for this high level of identity (Knight and Shimeld, 2001).

By searching the genome for C_2H_2 zinc finger motifs with a Hidden-Markov model (see Materials and Methods), we identified 377 genes, of which most corresponded to predicted GLEAN gene models (Sea Urchin Sequencing Consortium, 2006). Of these 377 genes, 17 match a gene model that emerged through inspection of the whole genome tiling array results (Samanta et al., 2006). For 50 of the genes that were not among the original gene model predictions, we were able to obtain a primer pair with which an authenticated zinc finger amplicon could be generated from genomic DNA. Transcriptional profiling revealed that ten of these genes are transcribed, indicating that they are functional genes. The conserved domain structure of the remaining genes indicates that they too are more-likely-than-not functional genes, and we have therefore included them in our list of identified zinc finger genes. Thus, though provisionally, we conclude that the sea urchin genome contains 377 C_2H_2 zinc finger genes.

Rapidly expanding gene families often contain a high number of pseudogenes (Glusman et al., 2001). However, these usually decay rapidly, accumulating stop codons and small deletions

(Zhang and Gerstein, 2004). If generated through retrotransposition events, they consist of only one exon. Although single-exon genes are common in our dataset, they generally have long open reading frames > 1 kb. We cannot exclude that the list of 377 C_2H_2 zinc finger genes includes some pseudogenes, but consider this unlikely.

The sea urchin genome contains more zinc finger genes than found in other invertebrates. There are 326 C_2H_2 zinc finger genes in the *Drosophila* genome (Chung et al., 2002); 198 in the urochordate *Ciona intestinalis* (Miwata et al., 2006), 211 in *C. elegans* (Reece-Hoyes et al., 2005). In contrast to all other species, *C. elegans* contains more nuclear hormone receptor-like genes than C_2H_2 zinc finger genes. In the current version of the *Nematostella vectensis* genome we identified 170 zinc finger genes. The *Nematostella* genome, however, is at this point still in a provisional state. A search for C_2H_2 zinc finger genes in a non-redundant set of mouse proteins, applying our criteria, yielded 731 genes. A similar search in the human genome identifies 764 genes. These numbers are consistent with other current estimates (Shannon et al., 2003; Tupler et al., 2001). The number of zinc finger genes in vertebrates far surpasses the number of zinc finger genes in the sea urchin genome, but this is a vertebrate not a chordate feature.

Properties of sea urchin zinc finger genes

The PFAM search identified over 3000 individual zinc finger domains, making this one of the most prominent protein domains in the sea urchin genome (Cameron et al., 2006). Generally C_2H_2 zinc finger domains appear in tandem. Often they are separated by only a few amino acids, the sequence of which is frequently TGEKPY/F, as has been described in other genomes (Laity et al., 2001). A few exceptions exist in which the domains are dispersed throughout the entire coding region. In the sea urchin genome the median number of zinc finger domains per gene is eight.

The zinc finger motifs of about 10% of all identified zinc finger genes appear to be extremely closely related at the nucleotide level, giving them a repetitive structure consisting of exact repeated sequence units of > 40 continuous base pairs (the entire zinc finger domain is 66 base pairs long). This causes obvious problems for primer design and transcriptional profiling by QPCR. However, where working primer pairs could be obtained, our results show that some of these genes are indeed expressed (e.g,. *Sp-z410*, *Sp-z265*). We conclude that these genes are not artifacts of the assembly.

In most zinc finger genes the C_2H_2 zinc finger motif is the only recognizable domain. Infrequently (about 16%) zinc finger genes contain a second kind of domain, mostly other types of zinc-binding domains, like the BED zinc finger. Two sea urchin zinc finger genes also encode homeodomains (*Sp-SmadIP1*, *Sp-athf15*) (Howard-Ashby et al., 2006b). In mammalian genomes, zinc finger factors frequently have repressive function, which is attributed to the presence of a Krüppel-associated-box (KRAB) domain. Almost half of all human and mouse C_2H_2 zinc finger factors contain this domain (Urrutia, 2003). A search for the KRAB domain (PF01352) was negative in the sea urchin genome; similarly no convincing SCAN domains (PF02023) (Williams et al., 1999) were found. Although the possibility remains that distant relatives of these domains are present in the sea urchin genome, this finding is consistent with the hypothesis that they arose after the divergence of the tetrapod lineage. About a third of all C_2H_2 zinc fingers in the *Drosophila* genome contain a ZAD domain which we could not identify in the sea urchin genome either (PF07776). It is thought to fulfill similar functions as the KRAB domain (Chung et al., 2002) and supposed to be dipteran specific. The lack of a recognized, repressive domain in sea urchin zinc finger factors might suggest that a novel taxon-specific domain remains to be found in these proteins.

To determine the set of C_2H_2 zinc finger genes that is active in early development, we performed transcriptional profiling by QPCR. Expression of a total of 324 genes for which specific primer pairs could be obtained was monitored at seven time points between fertilization and late gastrulation (48 h postfertilization). We consider genes to be transcribed

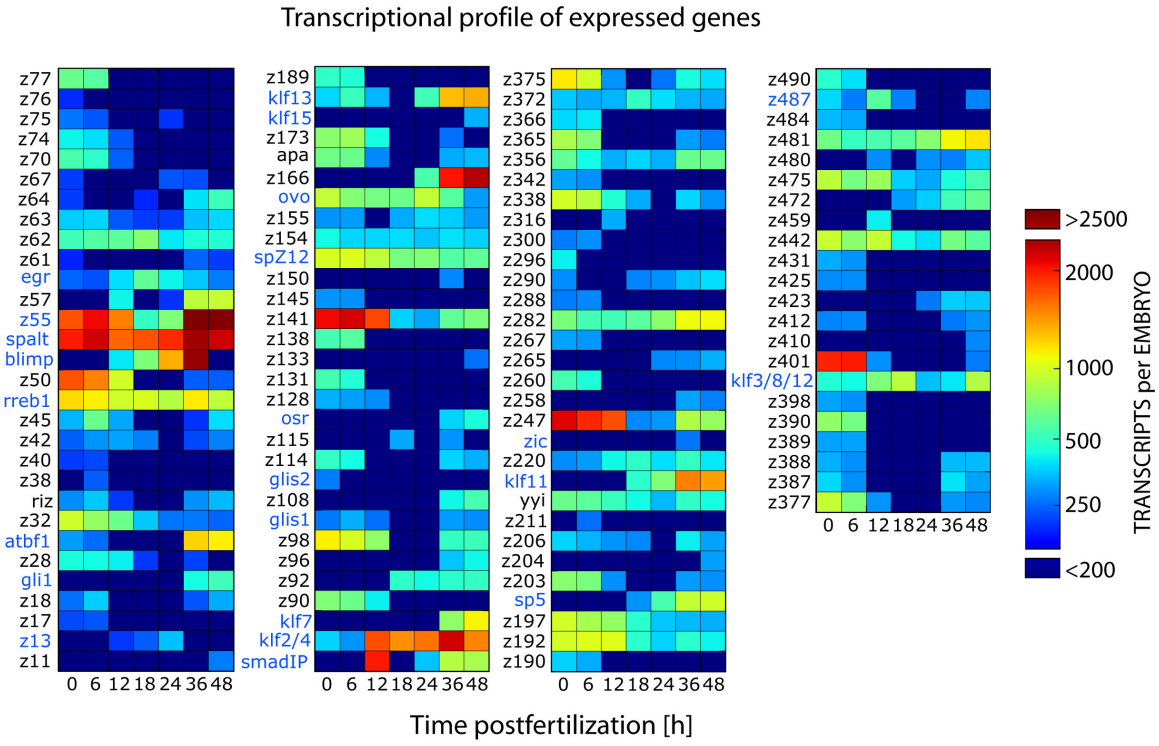


Figure 2.1: Summary of transcriptional profiling data for 112 expressed genes. For each gene the expression level was determined by QPCR at seven time points from fertilization to 48 h, the late gastrula stage. The color at each time point corresponds to the expression level in transcripts per embryo as indicated in the key. Zinc finger genes whose expression exceeds 200 transcripts per embryo at one or more time points were considered to be expressed, and only these genes are included in this figure. The majority are not expressed in this period of development (see text). Most zinc finger genes are expressed at low levels, such that they are represented by no more than several hundred transcripts per embryo at their peaks. Seventy-five percent of expressed zinc finger genes are expressed maternally. *Sp-blimp*, *Sp-z13* (previously *Sp-krl*) and *Sp-SpZ12* are previously known genes. *Sp-atbf1* and *Sp-smadIP* also contain homeobox domains in addition to the zinc finger domains (see Howard-Asbhy et al., 2006a). Names of genes referred to in the text are colored blue.

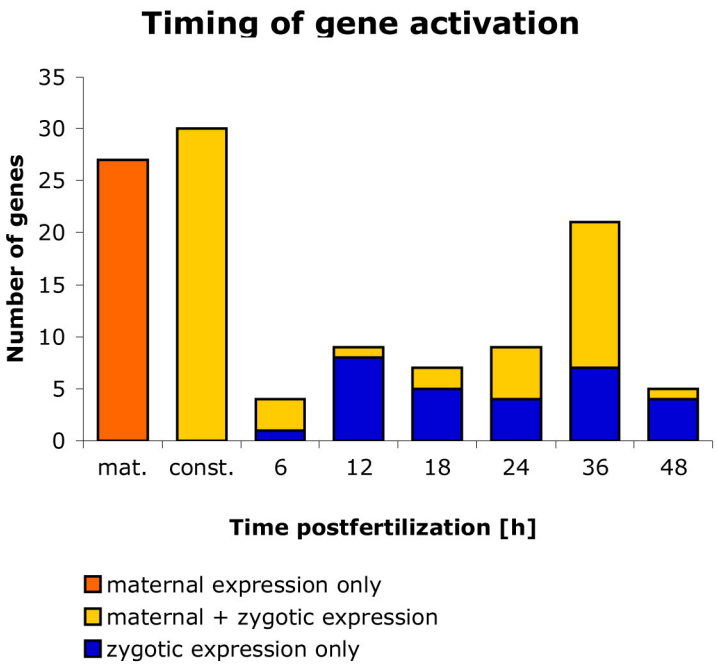


Figure 2.2: Stages of initial activation of C_2H_2 zinc finger genes. Genes are classified by the start of zygotic expression (6 to 48 h postfertilization). Genes that are only zygotically expressed are depicted in blue, while genes that are maternally and zygotically expressed are yellow. Genes that are maternal and expressed throughout early development (allowing for some variation in expression level) are captured by the "constant" category (*const.*). Genes that are maternal and fall below our threshold of 200 transcripts per embryo, i.e., are not expressed, but later have a phase of zygotic expression, are classified by the start of zygotic expression. Genes that are maternal only and not expressed zygotically are depicted in orange (*mat.*). The six hour category represents genes that display a clear increase in the transcript number at this point. Although "constant genes" may begin to be transcribed as well at this or at later times, unless there is net transcript accumulation this will not be visible.

significantly if they are represented by at least 200 transcripts per embryo. This excludes genes that, even if expressed in as few as 40 cells in the 800-cell late gastrula will be represented by only 5 mRNAs per cell, and if expressed ubiquitously, 0.25 mRNAs per cell. According to this classification, only 112 of the genes are significantly transcribed, i.e., 35% of all zinc finger genes tested. A colorimetric summary of the expression dynamics for these 112 genes is

displayed in Fig. 2.1 (individual timecourses can be viewed at our website, http://sugp.caltech.edu). The overall transcriptional level of expressed zinc finger genes is relatively low, except for a few cases (red and yellow in Fig. 2.1). Typically, at peak levels of expression, there are but several hundred molecules per embryo. More than 95% of genes we classify as transcribed were also identified as transcribed in the whole genome tiling array analysis (Samanta et al., 2006). Furthermore, the present QPCR analysis of three previously known C_2H_2 zinc finger genes – i.e., *Sp-Z12/z151* (Wang et al., 1995), *Sp-z13* (previously *Sp-krl*; (Howard et al., 2001), and *Sp-blimp/krox/z51* (Livi and Davidson, 2006) – was consistent with previously determined transcriptional profiles.

Eighty-three genes, that is, 74% of significantly expressed genes, are represented in maternal RNA, indicating a strong bias toward oogenetic transcription for zinc finger genes. Twenty-seven genes (24%) are expressed maternally only and are not reused during early development (Fig. 2.2). Thirty genes (27%) are expressed maternally and continue to be transcribed throughout early development, albeit showing fluctuations in transcript level (e.g. *Sp-spalt/z54*, *Sp-rreb/z48*, Fig. 2.1). Twenty-six maternal genes (23%) initiate a second expression phase after falling below our threshold of 200 molecules and are likely to have a zygotically regulated activation in addition to maternal expression (e.g., *Sp-athf1/z30* or *Sp-z50*, Fig. 2.1). Twenty-nine genes (26%) are not maternally expressed and begin to be transcribed only in the course of development. Together, the "maternal only" and "constantly expressed" genes account for more than half of all expressed genes (51%; Fig. 2.2). Only four genes are activated for the first time at six hours, but genes which are maternally expressed may well be transcribed at this point too. A plurality of the expressed genes, discounting those that are constantly expressed, begin zygotic expression in the 24–36 h window (Fig. 2.2). This corresponds to the time prior to and overlapping with gastrulation. By 48 h, only five additional genes have begun to be transcribed.

Table 2.1:
Summary of Spatial Expression Patterns.

Name	Exp. Start	Mat.	In Situ Result					
			7h	12h	18h	24h	36h	48h
z13	12		EM, MPS	SM	SM	E	-	ND
z30	36	M	-	-	-	-	FG	-
z48	c	M	UBIQ	-	MPS	SM	-	ND
z54/spalt	c	M	UBIQ	UBIQ	UBIQ	MPS,E,SM	FG,MG,HG	FG,MG,HG
z55	c	M	UBIQ	-	-	OE, ABO	FG,MG,HG,OE	FG,MG,HG,OE
z60/egr	c	M	UBIQ	UBIQ	UBIQ	OE, ABO	A	A
z67	24	M	-	-	-	MPS	B	-
z81/smadIP	12		-	ABO and OE	-	SM	A, FG	A, FG
z85/klf2/4	12	M	-	ABO or OE	ABO or OE	ABO or OE	ABO and/or OE	-
z86/klf7	18		-	-	-	ABO or OE	-	-
z92	18		UBIQ	UBIQ	UBIQ	MPS,E,SM	OE	UBIQ
z121/osr	36		-	-	-	-	MG, B	ND
z133	36		-	-	-	A	A	A
z141	6	M	UBIQ	UBIQ	UBIQ	MPS,E,SM	B	-
z157/ovo	c	M	UBIQ	UBIQ	A,MPS,E,SM	MPS,E,SM	-	-
z166	24		-	-	-	SM	OE or ABO	ND
z173	36	M	-	ABO or OE	ABO or OE	-	-	ND
z188/klf13	c	M	-	-	E, A	E, A, OE or ABO	A, OE	A, OE
z199/sp5	18		-	-	ABO or OE	ABO or OE	B, OE	HG
z204	48		-	-	-	-	B	B
z244 zic	36		-	A	A	A	A	A
z487	c	M	UBIQ	UBIQ	ABO, OE	-	-	-
z18	c	M	UBIQ	-	-	-	-	-
z28	c	M	UBIQ	-	-	-	-	-
z32	c	M	UBIQ	-	-	-	-	-
z38	6		UBIQ	-	-	-	-	-
z45	6	M	UBIQ	-	-	-	-	-
z62	c	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	ND
z65	48	M	UBIQ	-	-	-	-	ND
z70	n	M	UBIQ	-	-	-	-	-
z74	n	M	UBIQ	-	-	-	-	ND
z77	36	M	UBIQ	-	-	-	-	-
z90	n	M	UBIQ	-	-	-	-	ND
z98	24	M	UBIQ	-	-	-	-	-
z114	24	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	ND
z197	c	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ
z212	c	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	ND
z214	18		-	-	UBIQ	UBIQ	-	ND
z247	c	M	UBIQ	-	-	-	-	-
z338	36		UBIQ	-	-	-	-	-
z401	48	M	UBIQ	-	-	-	-	-
z425	n	M	UBIQ	-	-	-	-	ND
z442	c	M	UBIQ	UBIQ	UBIQ	-	-	-

Genes that show a localized expression are listed in the upper half, genes with ubiquitous expression in the lower half. For each gene the start of zygotic expression (*Exp. Start*) is given (c — constant expression, n — no zygotic expression) and whether or not it is maternal (M — maternal expression). EM — endomesoderm, EC — ectoderm, MPS — micromeres/PMC/skeletogenic tissue, A — apical ectoderm/apical plate, SM — secondary/mesenchyme/mesoderm, E — endoderm, OE — oral ectoderm, ABO — aboral ectoderm, B — blastopore, FG — foregut, MG — midgut, HG — hindgut, UBIQ — ubiqitous.

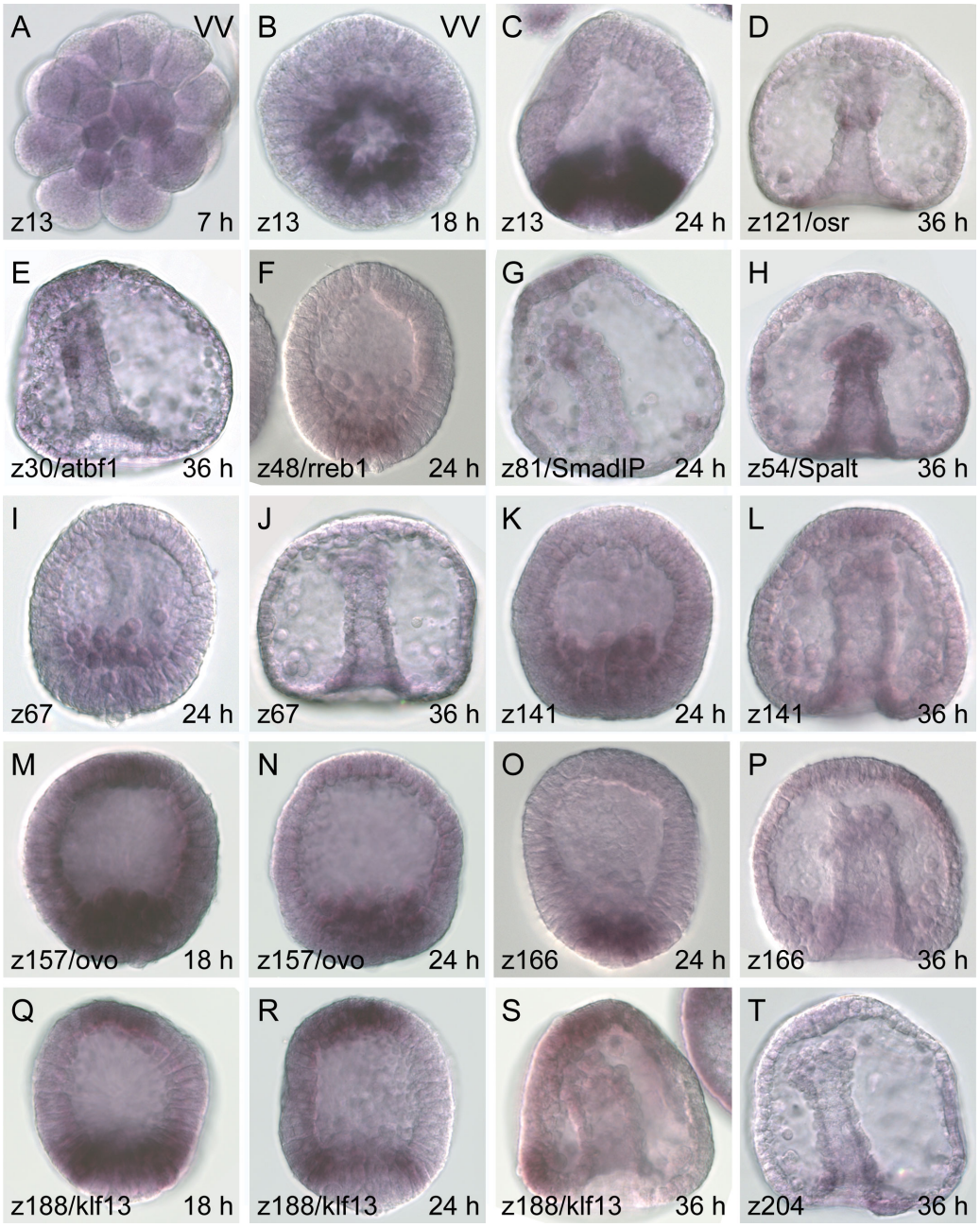


Figure 2.3: Endomesodermal expression of C_2H_2 zinc finger genes. Whole-mount *in situ* hybridizations are shown at stages listed in the lower right corner of each panel. A full summary of spatial expression patterns is given in Table 2.1. Gene names are given in the lower left corners. Embryos were recorded in lateral view unless stated otherwise. vv — vegetal view.

In situ hybridization

We attempted to perform in situ hybridization for all the more highly expressed genes, and succeeded in collecting spatial expression data for 43 genes (summarized in Table 2.1). Localized expression was found for 22 genes during at least one phase of development, while 21 genes were observed to be expressed ubiquitously throughout. Patterns of expression for genes utilized specifically in the endomesodermal domains are shown in Fig. 2.3, and for genes transcribed detectably in ectodermal and apical domains in Fig. 2.4.

A highly confined spatial expression pattern is presented by *Sp-osr/z121*. It is expressed during gastrulation and is localized in what will become the midgut, possibly only at the midgut-hindgut boundary (Fig. 2.3D). Many other genes are also expressed in particular regions of the archenteron (Figs. 2.3H, J, L, P, T), though some are also expressed in ectodermal or apical domains. *Sp-klf13/z188* is expressed in the endodermal region at swimming blastula stage (Figs. 2.3Q, R), but becomes localized to the ectoderm after the beginning of gastrulation (Fig. 2.3S). The *Sp-z13* gene, previously known as *Sp-krl* (Howard et al., 2001), displays a particularly dynamic pattern of expression. It is first transcribed at ~ 7 h postfertilization, in the micromeres. During the 12 to 18 h period, its transcripts disappear from the cells of the skeletogenic lineage and become localized first to the endomesoderm and then in the mesoderm proper (Figs. 2.3A, B). After ingression of the primary mesenchyme cells, expression disappears from the mesodermal tissue and extends to the endodermal region (Fig. 2.4C).

Genes that are expressed in mesodermal tissues include *Sp-z166* which is localized in the vegetal plate after the ingression of primary mesenchyme cells (Fig. 2.4O). Similarly, after an initial phase of ubiquitous expression, *Sp-ovo/z157* is transcribed in the vegetal plate (including the mesoderm) at swimming blastula stage (Figs. 2.3M, N).

Expression of several genes localizes to different regions of the ectoderm. This includes genes expressed in the apical region (Figs. 2.4A–D, F) and genes expressed in either the oral or aboral ectoderm (e.g., Figs. 2.4H–K, M–T). *Sp-z487* is a gene that is expressed only transiently

in embryonic development. It is initially expressed ubiquitously, and then becomes restricted to the ectoderm only, without showing any oral or aboral bias (Fig. 2.4L).

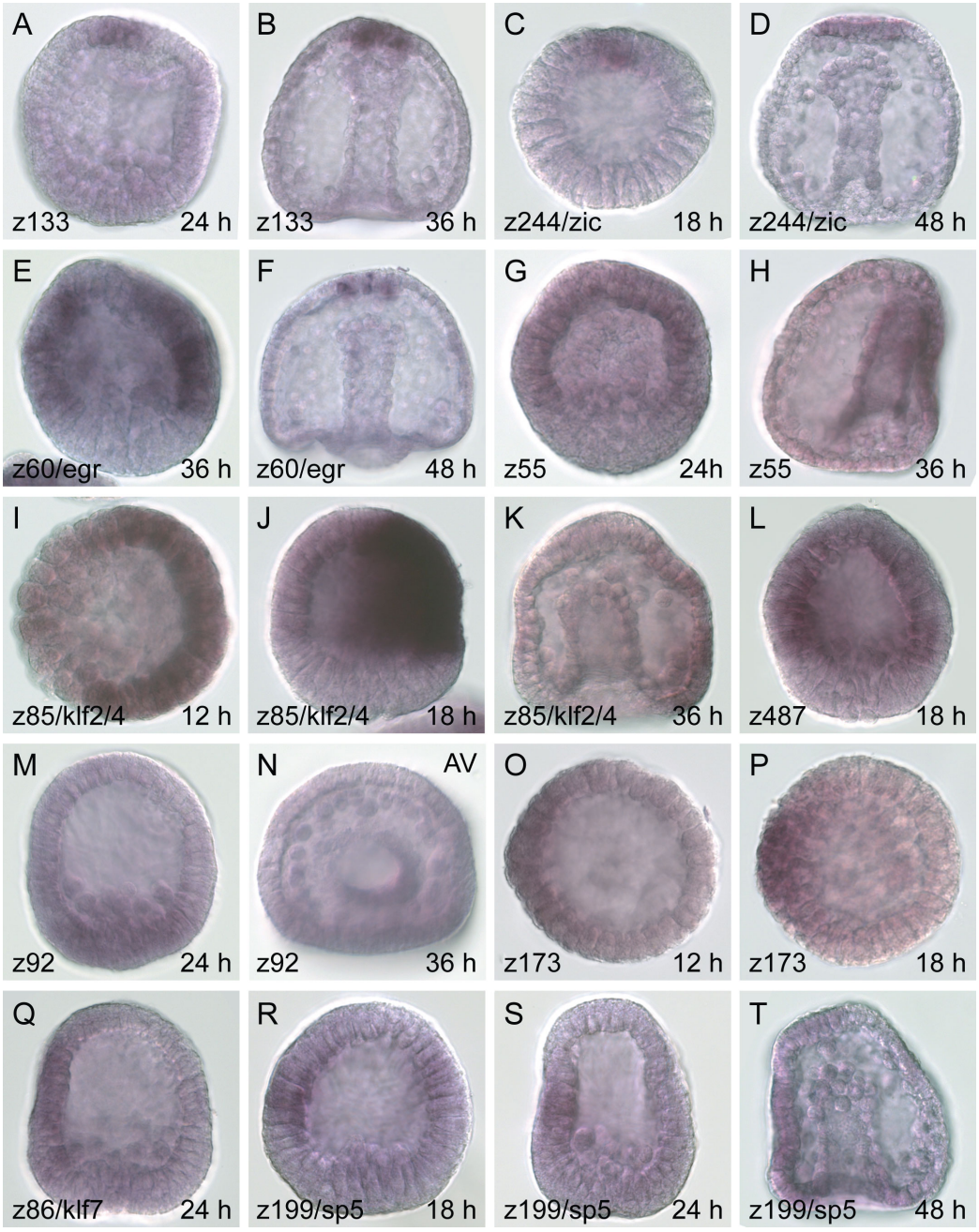


Figure 2.4: Ectodermal expression of C_2H_2 zinc finger genes. This includes genes with expression in the apical plate (e.g., panels A–D), and oral and aboral ectoderm. A full summary of spatial expression patterns is given in Table 2.1. Embryos were recorded in lateral view.

Due to the rapid evolution of zinc finger genes it is often difficult to recognize orthology between different proteins even where it does exist. We identified zinc finger genes that belong to orthologous groups of highly conserved and well studied genes (Table 2.2). Even for these groups, only the zinc finger region can be aligned confidently, sometimes including some flanking sequence. The number of zinc finger domains per gene is known to be variable even between fairly closely related zinc fingers (Shannon et al., 2003). However, within the orthologous gene sets presented in this section the number of zinc finger domains is invariant. All genes described here for which we identified sea urchin homologues are known transcription factors, involved in various developmental processes.

Table 2.2: Comparison of Number of C_2H_2 Zinc Finger Genes in Sea Urchin and other species

Gene Family	**S. p.**	**H. s.**	**M. m.**	**C. i.**	**D. m.**	**C. e.**	**N. v.**
ZF total	377	764	731	198	326	211	170
zic	1	5	6	2	1	1	n.d.
gli	3	5	5	2	3	1	2
klf/sp1	9	24	24	n.d.	n.d.	n.d.	n.d.
snail	1	3	3	1	4	1	1
scratch	2	3	2	0	3	1	0
egr/krox	1	4	4	0	1	1	1
ovo	1	2	3	1	1	1	0
spalt	1	4	3	0	2	1	0
odd	1	2	2	1	3	2	0

S.p. — *Strongylocentrotus purpuratus*, H.s. — *Homo sapiens*, M.m. — *Mus musculus*, C.i. — *Ciona intestinalis*, D.m. — *Drosophila melanogaster*, C.e. — *Caenorhabditis elegans*, N.v — *Nematostella vectensis*, n.d. – not determined. The genome of *Nematostella* is still in a provisional state; absence as indicated in this table, may not mean absence from this genome.

Gli (Glioma associated oncogene) proteins are known to play important roles in development, for example in promoting neuronal differentiation (Mayor and Aybar, 2001). They are transcriptional mediators of the hedgehog (Hh) signaling cascade. Gli proteins contain five zinc finger domains that are highly conserved. Human and mouse genomes both contain three *gli* genes. *Drosophila* contains one, the segment polarity gene *cubitus interruptus*, as does *C. elegans* (*Tra-1*). Two groups of *gli-similar* genes (*glis*) have been described. Mouse and human contain two genes of the *glis1* subfamily to which the *Drosophila* gene *lmd* also belongs. The *glis2* subfamily contains one group each of mouse, human, and fly (*Dm-sug*) genes. According to the phylogenetic tree in Fig. 2.5, the sea urchin contains one gene from each group (*Sp-gli/z22*, *Sp-glis1/z113*, *Sp-glis2/z107*) indicating that these subgroups had already appeared before the divergence of the vertebrate lineage. Transcriptional profiling shows that *Sp-gli* is transcribed at 36 h postfertilization. The two *glis* genes are expressed maternally and are transcribed at low levels in the embryo.

Human and mouse genomes contain at least five *zic* (zinc finger genes of the cerebellum) genes. In Drosophila only one such gene can be found (*odd-paired/Dm-opa*). Ascidians contain at least two genes in this group, *Ci-macho1* and *Ci-zicL* , that seem to have diverged in this lineage

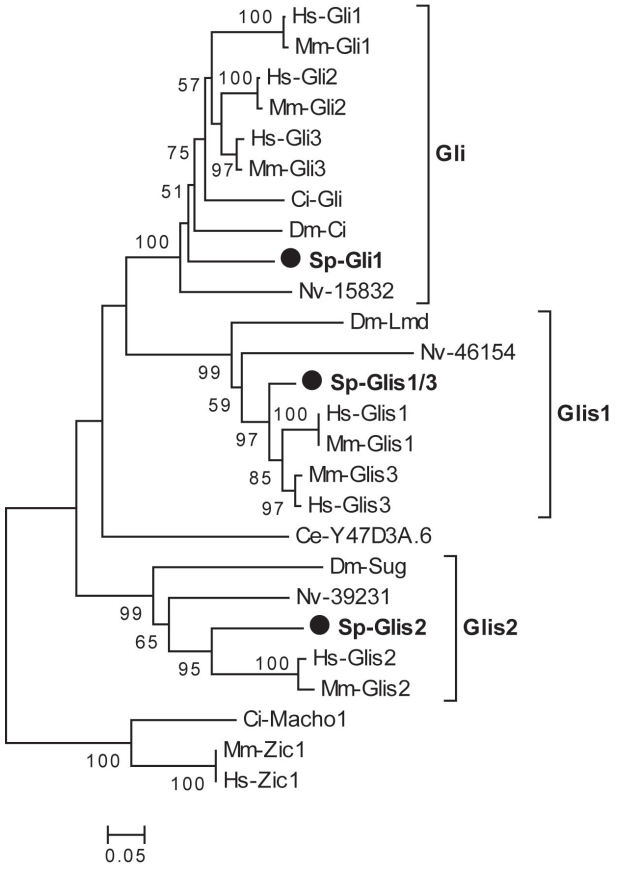


Figure 2.5: Phylogenetic tree of Gli and Glis proteins, based on alignment of the zinc finger regions. Sea urchin proteins can be found in all three major classes (Gli, Glis1, Glis2), supported by high bootstrap values (> 95). Only values higher than 50% are shown. Proteins of the Zic family were used as the outgroup. The scale bar indicates an evolutionary distance of 0.05 amino acid substitutions per position. Ce — *Caenorhabditis elegans*, Ci — *Ciona intestinalis*, Dm — *Drosophila melanogaster*, Hs — *homo sapiens*, Mm — *Mus musculus*, Nv — *Nematostella vectensis*, Sp — *Strongylocentrotus purpuratus*.

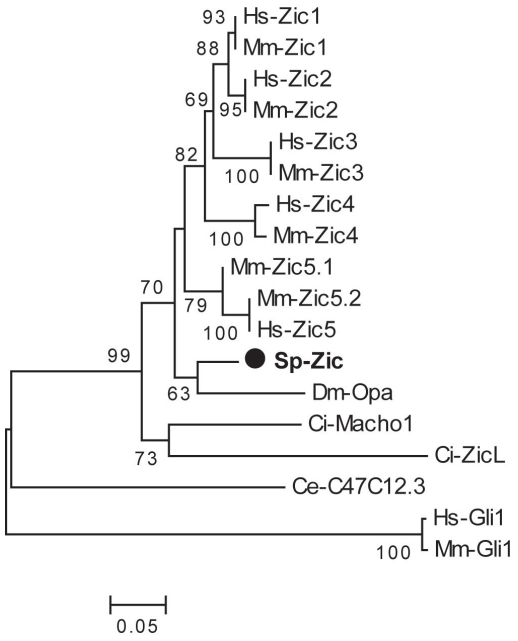


Figure 2.6: Phylogenetic tree of Zic proteins, based on alignment of the zinc finger regions. Only one sea urchin Zic sequence could be identified. Gli proteins were used as the outgroup. Annotations are as in Fig. 2.5.

(Yamada et al., 2003). *zic* genes in vertebrates are located in a cluster, underlining the close relationship of these genes. The Zic proteins, like the proteins of the Gli family, are important regulators of neural development and interact with these (Aruga, 2004). A phylogenetic tree (Fig. 2.6) in which the *gli* genes are used as the outgroup identifies only one sea urchin *zic* gene (*Sp-zic/z244*). The sea urchin *zic* orthologue begins to be expressed at around 18 h postfertilization, and its transcripts are localized in the neurogenic apical plate (Figs. 2.4C, D).

Zinc fingers of the Krüppel-like/Sp1 *family*

The genes of the *krüppel-like/Sp1* family encode a diverse group of transcriptional regulators. In mammals, the genes of this family are dispersed throughout the entire genome. Sp1 was one of the first mammalian transcription factors to be cloned and named by its sequence similarity

to the *Drosophila* gap gene *krüppel*. While this gene contains four zinc fingers, the genes of the *krl/Sp1* family only contain three, which are linked by a highly conserved linker that frequently connects two adjacent zinc finger domains (Kaczynski et al., 2003). The human genome contains eight genes of the Sp1 and 16 genes of the *krl* subfamily. All *Sp1* genes are transcriptional activators, whereas some *krl* genes are repressors and can counteract Sp1-mediated gene activation (Lomberk and Urrutia, 2005; Urrutia, 2003). In mice *krl/Sp1* genes are expressed in a wide variety of tissues controlling various processes in development (e.g. *klf1* and *klf2* are involved in eryothropoiesis and in blood vessel and lung development) or controlling cell growth and proliferation. Several of them are known as tumor suppressor genes (e.g., *klf4* and *klf7*). Phylogenetic analysis (Fig. 2.7) reveals that the sea urchin genome contains six genes of the *klf* subfamily (*Sp-klf2/4/z85*, *Sp-klf3/8/12/z400*, *Sp-klf7/z86*, *Sp-klf11/z214*, *Sp-klf13/z188*, *Sp-klf15/z174*) and three genes of the *Sp1* subfamily (*Sp-sp2/z168*, *Sp-sp5/z199*, *Sp-sp8/z177*). All sea urchin klf genes and *sp5* are expressed in the early embryo. *Sp-klf2/4*, *Sp-klf7* and *Sp-sp5* are localized in the ectoderm (Fig. 2.4I–K, Q–T) and *Sp-klf13* is expressed in the ectoderm and the endoderm (Figs. 2.3Q–S).

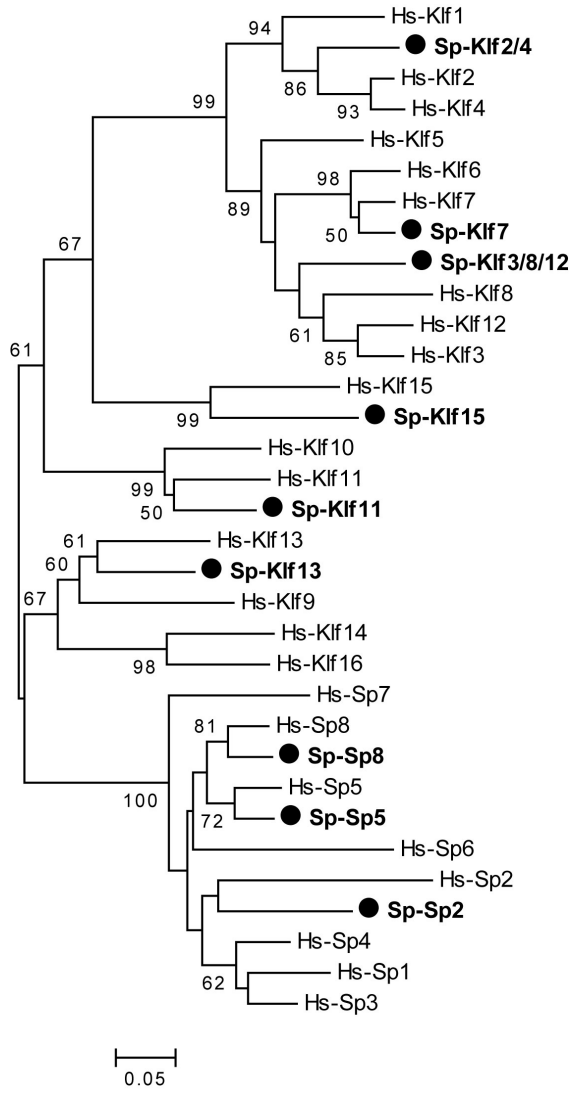


Figure 2.7: Phylogenetic tree of human and sea urchin Klf and Sp1-like proteins based on alignment of the zinc finger regions. In the sea urchin six Klf sequences and three Sp1-like sequences could be identified. Annotations are as in Fig. 2.5.

A sea urchin zinc finger gene described previously as Sp-krl (Howard et al., 2001) does not seem to be a close relative of the genes of the *krl/Sp* family identified by phylogenetic analysis (Fig. 2.8). In fact, it is among the zinc finger genes for which no unequivocal homologue can be identified. In our data set this gene is designated *Sp-z13*.

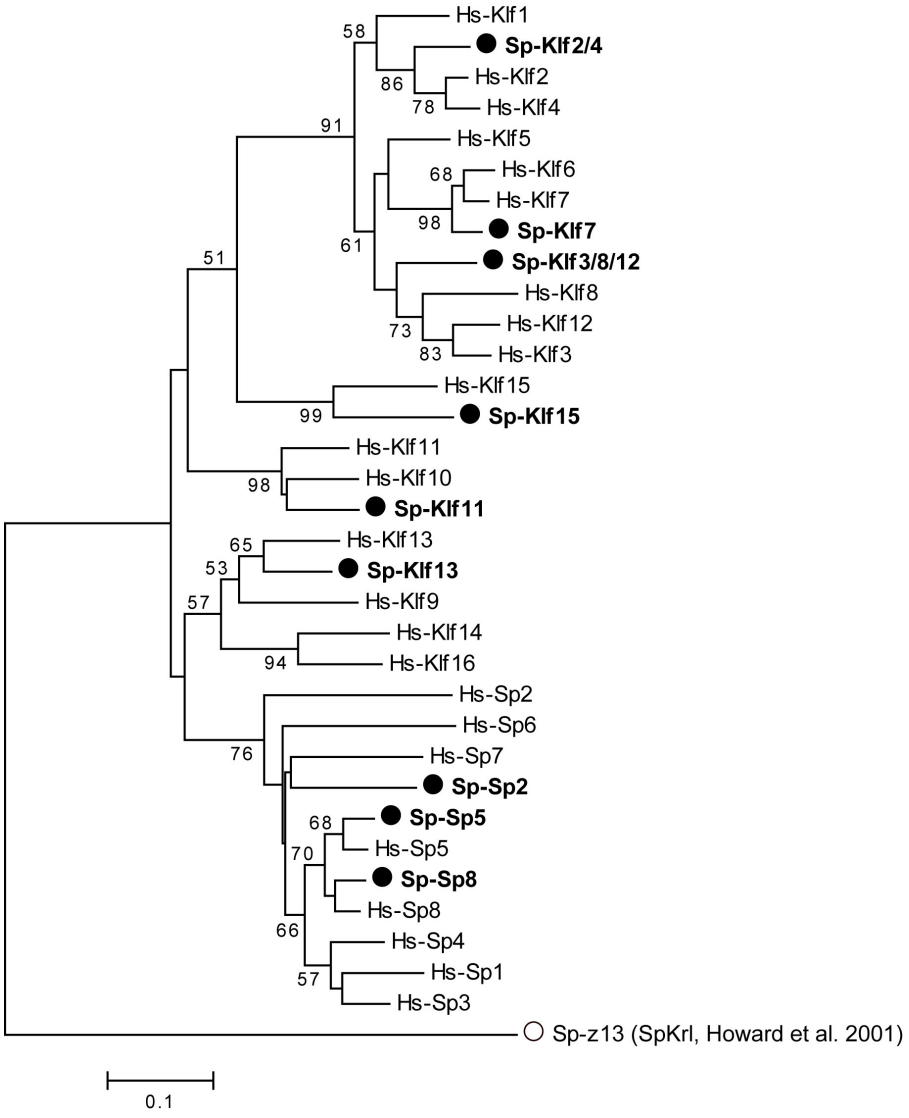


Figure 2.8: The previously described zinc finger gene *Sp-z13* is not part of the Sp/Krl family of transcription factors. Annotations are as in Fig. 2.5

Zinc fingers of the Snail *family*

The *snail* gene family encodes a highly conserved group of transcriptional repressors. In most genomes this family comprises at least one member of both the snail and the scratch subfamilies.

Snail is often involved in epithelium-mesenchyme transitions, for example in neural crest formation in mice (Manzanares et al., 2001). In vertebrates *snail* is thought to have given rise to two additional genes, *snail2/slug* and *snail3* (Manzanares et al., 2004). In *Drosophila* the three members of the family are *snail*, which provides spatial control of gene expression along the embryonic dorso-ventral axis, and *escargot* and *worniu*, which are employed in formation of the central nervous system (CNS). *Ciona* and *C. elegans* have one *snail* gene each. A clearly orthologous *snail* gene has even been found in *Nematostella vectensis* (Corbo et al., 1997). The sea urchin contains one *snail* gene (*Sp-snail/z88*) that is most closely related to vertebrate *Snail1* and *Snail2* (Fig. 2.9).

The genes of the scratch subfamily in mice and fly are employed in neural crest formation and CNS development. In contrast to mammals, which contain two or more of these genes, the sea urchin genome contains only one homologue of *scratch* (*Sp-scratch/z213*). *Drosophila* contains two but they seem to be paralogous and they fall outside the mouse/human/sea urchin group. We identified an additional gene of the scratch subfamily in the sea urchin, that clusters with a novel transcript from Drosophila (Bootstrap value of 99), which we named *Sp-scratchX* (*Sp-scratchX/z191*). This may indicate an ancestral duplication of the *scratch* gene, one of which was lost in the vertebrate lineage.

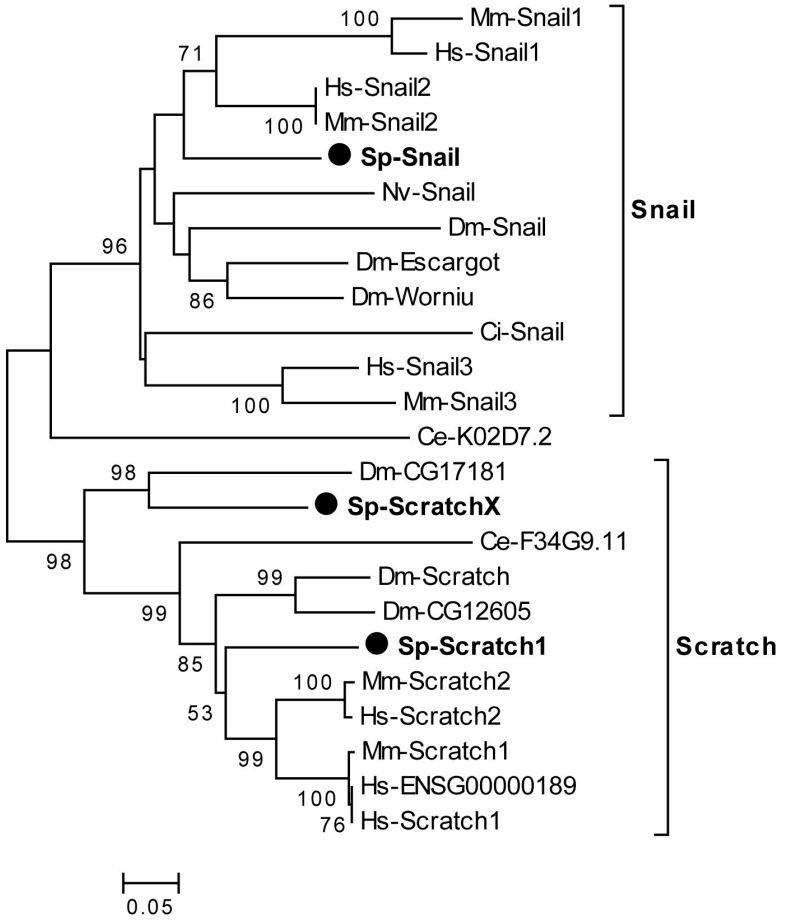


Figure 2.9: Phylogenetic tree of proteins of the Snail family based on alignment of the zinc finger regions. The Snail subfamily clearly separates from the Scratch subfamily. The sea urchin genome encodes one Snail and one Scratch orthologue. A second protein clusters with a novel protein from Drosophila within the Scratch subfamily, but is clearly separated from other Scratches. We named this protein Sp-ScratchX. Annotations are as in Fig. 5.

Zinc finger genes of the Spalt*,* Egr*,* Ovo *and* Odd *families*

spalt genes were first identified in *Drosophila*, where they are involved in diverse processes such as homeotic specification of the embryonic termini, wing patterning, and sensory organ development. In vertebrates four *spalt* genes can be found that are involved in similarly diverse

processes ranging from limb development to development of the nervous system and organs like the kidney and heart (Sweetman and Münsterberg, 2006). They contain three or four pairs of zinc finger domains that are located in the C-terminal half of the protein. The sea urchin contains one gene of this family, which we named *Sp-spalt* (*Sp-spalt/z54*, Fig. 2.10D). This gene is expressed throughout embryonic development and its transcripts are localized to the endoderm (Fig. 2.3H).

The genes of the *egr/krox* family in mammals are mainly involved in brain development, with *egr1/krox-24* being expressed in the sensory cortex (Herdegen and Leah, 1998). Recently a role in learning has been invoked for this gene. In *Drosophila* this gene is expressed in the epidermis, and mutants show defects in recognition of myotubules by epidermal cells. Although the structure of Egr1/Krox-24 has been determined, and its interaction with DNA is one of the best understood, the binding sites seem to show considerable variability and not many direct targets have been identified. The sea urchin contains one member of this family (*Sp-egr/z60*), which seems to be most closely related to *egr1* (Fig. 2.10A). The gene is expressed throughout early development. Whole-mount *in situ* hybridization shows localization of the transcript to the ectoderm beginning with late blastula stage (Figs. 2.4E, F)

The *Drosophila ovo* gene is primarily expressed in the female germ line. Loss of ovo activity leads to sterility in female flies. In mice two *ovo*-like genes are known. Both seem to be involved in spermatogenesis (Dai et al., 1998). *ovol*-1 is expressed in a range of different tissues such as kidney and epidermis, and mutants show defects in hair morphogenesis. In sea urchin we identified one gene of this group (*Sp-ovo/z157*, Fig. 2.10C). *Sp-ovo* is expressed constantly during embryonic development.

In *Drosophila*, *odd-skipped* (*odd*) was originally identified as a pair-rule gene because mutations at this locus lose portions of odd numbered segments in the embryo. Two closely related genes, *bowl* and *sob*, were identified, of which Bowl has an important function in the development of the terminal segments and the gut. Two orthologues are known in mammals. Mouse *osr1* is expressed in embryonic mesoderm and at later stages in the branchial arms and limb buds (Wang et al., 2005). It has also been shown to be involved in heart and kidney development.

osr2 is also expressed in limbs and kidneys, but mice deficient for *osr2* show no defects in these organs. We identified one *odd*-like gene in the sea urchin (*Sp-osr/z121*, Fig. 2.10B). This gene is expressed late in embryonic development.

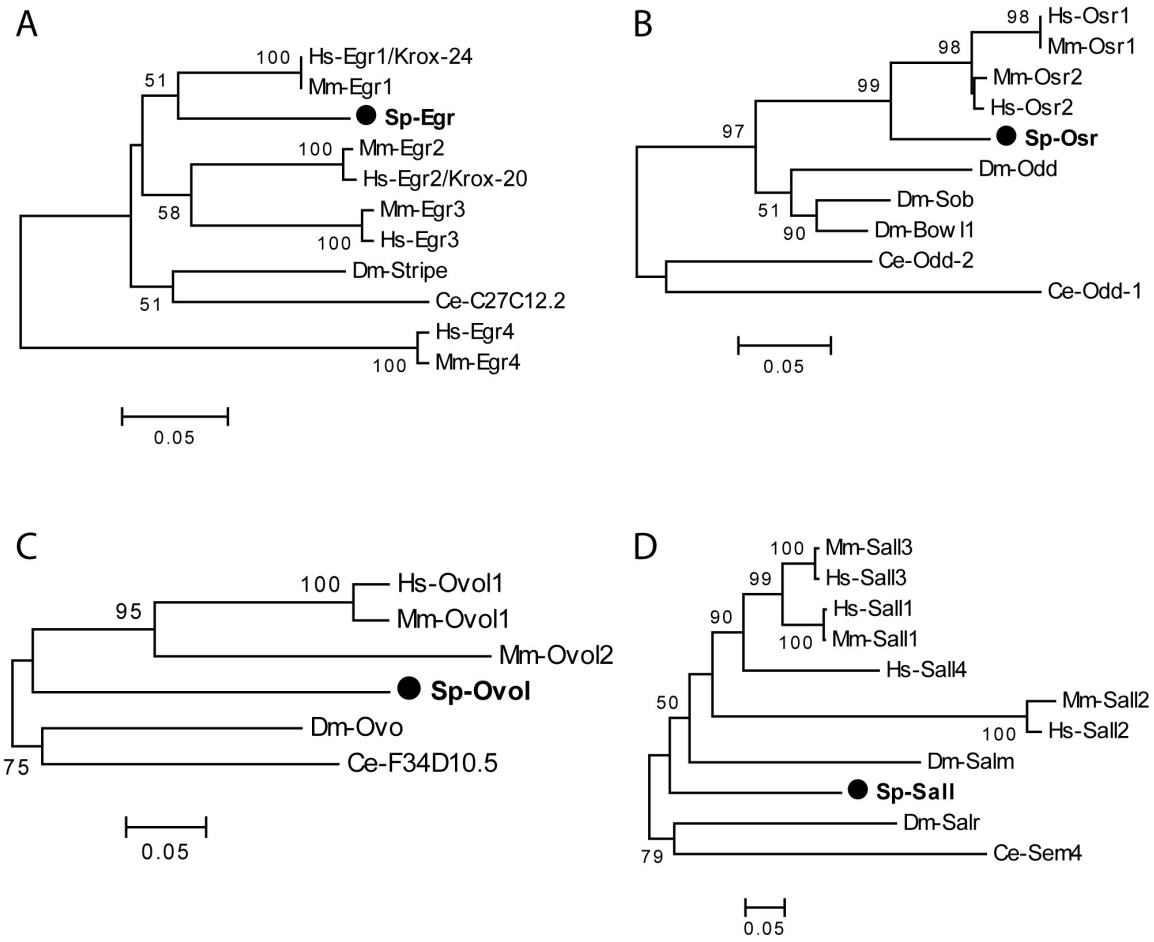


Figure 2.10: Phylogenetic analysis of zinc finger transcription factor families. (A) Egr/Krox, (B) Osr, (C) Ovo, (D) Spalt. Annotations as in Figure 2.5.

Discussion

The zinc finger genes thus far identified account for at least 1.5% of the total genes in the sea urchin genome (Sea Urchin Sequencing Consortium, 2006). This result parallels findings for other genomes in which zinc finger-containing genes are among the most abundant classes (Adams et al., 2000; Lander et al., 2001; Tupler et al., 2001). Yet the meaning of this large set of genes remains largely enigmatic. Only a minor fraction of the *Strongylocentrotus purpuratus* C_2H_2 zinc finger genes are orthologous to known regulatory genes of human or *Drosophila*. As detailed above these include members of the *zic*, *gli*, *snail* , and *krl/Sp1* families, all of which are known for their roles in development. The relatively low number of clearly identifiable orthologues is consistent with results of Knight and Shimeld (2001), who report that only about 25% of fly, worm, and human C_2H_2 zinc finger genes can be assigned to orthologous groups.

Several features of the large class of genes defined by the presence of sequence encoding C_2H_2 zinc finger domains distinguish them from most classes of DNA-recognizing regulatory genes, and these features are unlikely to be independent. First, the lack of extensive interspecific conservation, such as permitting unequivocal assignment of sequence orthology, contrasts sharply with what is seen for virtually every other prominent class of regulatory gene. In the *S. purpuratus* genome, for example, except for rare stragglers almost every gene encoding homeodomain factors (Howard-Ashby et al., 2006b), ets factors (Rizzo et al., 2006), forkhead factors (Tu et al., 2006), bHLH factors, nuclear hormone receptor factors, and factors of many other smaller groups (Howard-Ashby et al., 2006a) can be related to one or another known regulatory gene subfamily in other animal genomes. Partly the problems in orthology assignment of zinc finger genes are due to internal structural aspects, which confound the algorithms that order phylogenetic similarity: there is a generally high level of sequence identity within zinc finger regions, due to structural requirements of the zinc finger structure itself. On average 40% of amino acids in the zinc finger region are identical (Knight and Shimeld, 2001). An additional difficulty is posed by the internally repetitive modular structure of many zinc finger proteins. However, other classes of regulatory genes, such as those encoding the bHLH

proteins, are subject to processes such as domain shuffling (Morgenstern and Atchley, 1999), but these nonetheless retain identifiable orthologous relationships. The example of those zinc finger genes which can be assigned to orthology groups show that the general difficulty of doing this for most of the zinc finger genes is not just due to their canonical internal properties.

A second distinctive feature of the sea urchin zinc finger genes is that they appear to have been evolving rapidly and in a clade-specific way. This is the most obvious explanation for their lack of orthology with the zinc finger genes of other animal genomes. Their number per genome is extraordinarily flexible: Table 2.2 shows there are almost twice as many zinc finger genes in *S. purpuratus* as in *Ciona*, another deuterostome; slightly more than in *Drosophila*; and half as many as in mammalian genomes. Rapid evolutionary change in zinc finger genes is their prominent characteristic. A good example is given by a cluster of zinc finger genes in the human genome on chromosome 19 (Shannon et al., 2003). Genes in this cluster were shown to be duplications of each other, but the number of zinc finger domains they contain varies widely – between seven and eighteen per gene. Another example, as we report here, is that no *S. purpuratus* zinc finger gene possesses either a KRAB or SCAN domain though these occur frequently in mammalian zinc finger genes. The expanded sea urchin zinc finger gene family, like those of other animals, is among the more evolutionarily flexible, lineage-specific families to be found in this genome.

Thirdly, the zinc finger genes are used differently in development than any other family of regulatory genes, or the set of all such genes (Howard-Ashby et al., 2006b). Up to the 48-hour late gastrula stage, 75 to 78% of this total set is significantly transcribed, while the expression measurements we report here, which were carried out with the same technology, showed that out of 324 zinc finger genes assayed only 112 are expressed by 48 hours. Most of the expressed genes zinc finger genes exceeded the threshold set arbitrarily for significant expression by several-fold at least at one or more time points, and most in the non-expressed category are well below this threshold. For example, in the whole genome tiling array analysis of the embryo transcriptome, expression of only an additional 58 of the 377 zinc finger genes was detected, and these genes are usually represented by about 50 to 100 transcripts in the

whole embryo (800 cells at 48 h). The large fraction of silent or essentially silent zinc finger genes is not the only difference. In the set of total regulatory genes, only 20% of the 192 that are significantly expressed are represented in maternal RNA of the unfertilized egg (Howard-Ashby et al., 2006c), while about 75% of the 112 expressed zinc finger genes are maternally expressed. The high maternal utilization of zinc finger genes is emphasized by the observation that 27 of the zinc finger genes are only expressed maternally (i.e., up to 48 h of development), while the corresponding number for genes encoding all other types of transcription factors is but two (Howard-Ashby et al., 2006c). We can not exclude the possibility that these maternal RNAs are not fully processed and, hence, are nonfunctional (Davidson, 1986). In order to show full maturity, the maternal transcripts would have to be cloned and sequenced. However, it is difficult to imagine that zinc finger genes produce such transcripts while all other transcription factors do not (Howard-Ashby et al, 2006c).

These three major attributes of the zinc finger gene family, which distinguish it from the regulatory gene set as a whole, can be interpreted in alternative ways. The most likely explanation is simply that C_2H_2 zinc finger domains do not per se constitute evidence sufficient to assume regulatory gene function, so that the comparison is to some extent between apples and oranges. On the other hand, some zinc finger proteins are most certainly transcription factors and the prevailing view is that even though they are in most genomes poorly annotated, they generally possess DNA-binding capability and are involved in transcriptional regulation (Knight and Shimeld, 2001; Krishna et al., 2003; Miwata et al., 2006). In addition to the orthologues of known zinc finger regulatory genes identified in this work, the specific expression patterns summarized in Table 2.1 and shown in Figs. 2.3 and 2.4 are typical of bona fide regulatory gene products in their spatial specificity, their relatively low copy number, and their dynamic quality. These genes are very likely to execute important regulatory functions in specific tissues of the embryo. Nonetheless, there is accumulating evidence that zinc finger domains are employed in processes other than DNA binding. Many of the C_2H_2 zinc finger domains may belong to proteins the primary function of which is RNA binding or protein-protein interactions (Croker et al., 2006; Laity et al., 2001; Lu et al., 2003). Relative prevalence of RNA binding function could help to account for the high representation of zinc finger sequences in the maternal transcript stockpile. Perhaps such other, nontranscriptional

functions are very heavily represented in the majority set of C_2H_2 zinc finger genes that remain silent during embryogenesis, while the ones that do encode transcription factors are utilized more or less as are other regulatory genes (Howard-Ashby et al., 2006c).

An alternative is that most zinc finger genes do indeed have regulatory function, but that they are clade-specifically specialized to execute the regulatory functions required to generate clade-specific features of the organism. Since they appear to be specific additions to most genomes, they may be involved in processes that are not shared between organisms, and these are primarily the processes of the terminal stages of development (Davidson and Erwin, 2006). This would explain why they are poorly represented in the canonical gene regulatory networks of the early to mid-stage embryo. Even the set of zinc finger genes of mice and humans seem to have diverged significantly, and this difference has been imagined to contribute to the differences between these species (Shannon et al., 2003). The flexibility and diversification of zinc finger gene use in the tip of the iceberg so far known to us is reminiscent of terminal differentiation processes, where, for example, alternative splicing and deployment of paralogous gene relatives are often most extreme. Though crystal structures indicate that three zinc fingers often suffice for recognition of sequence-specific DNA binding sites (Choo et al., 1997), many of the zinc finger genes of the sea urchin have a much higher number of zinc fingers (the same is true for zinc fingers in the human genome; data not shown). Different finger domains may be used alternately for recognizing different binding sites, as in the transcription factor CTFC, an 11 zinc finger protein (Filippova et al., 2002). Alternative splicing may indeed also contribute versatility in the highly specific utilization of zinc finger regulators. As a rule, evolutionarily expanded gene families display reduced numbers of alternative splice forms, but there is a known exception: the zinc finger genes (Kopelman et al., 2005). The mysteries surrounding this prominent, rapidly evolving, and ubiquitous gene family are likely to resolve only when at last we have learned something of the functional roles in development of the majority of C_2H_2 zinc finger genes.

Acknowledgements

The authors would like to thank C.T. Brown for help with computation, T. Minokawa for help with WMISH, and J. Nam for critical support of phylogenetic analysis. We are grateful for helpful comments and suggestions on the manuscript by V. Hinman and J. Nam. This research was supported by NIH grant HD37105, and by the Office of Science (BER), US Department of Energy, grant DE-FG02-03ER63584.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H.C., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Miklos, G.L.G., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, A., de Pablos, B., Delcher, A., Deng, Z.M., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F.C., Gorrell, J.H., Gu, Z.P., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, D.A., Howland, T.J., Wei, M.H., Ibegwam, C., et al., 2000. The genome sequence of Drosophila melanogaster. Science 287, 2185–2195.

Andreazzoli, M., Delucchini, S., Costa, M., Barsacchi, G., 1993. RNA-binding properties and evolutionary conservation of the *Xenopus* multifinger protein Xfin. Nucleic Acids Res 21, 4218–4225.

Aruga, J., 2004. The role of Zic genes in neural development. Mol Cell Neuro, 26, 205–221.

Cameron, R.A., Materna, S.C., Berney, K., 2006. The *Strongylocentrotus purpuratus* genome: A Comparative Perspective. Dev Biol, 300, 485–495.

Choo, Y., Castellanos, A., Garcia Hernandez, B., Sanchez Garcia, I., Klug, A., 1997. Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. J. Mol. Biol. 273, 525-532.

Chung, H.R., Schäfer, U., Jäckle, H., Böhm, S., 2002. Genomic expansion and clustering of ZAD-containing C_2H_2 zinc-finger genes in Drosophila. Embo Rep 3, 1158-1162.

Corbo, J.C., Erives, A., DiGregorio, A., Chang, A., Levine, M., 1997. Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. Development 124, 2335-2344.

Dai, X., Schonbaum, C., Degenstein, L., Bai, W.Y., Mahowald, A., Fuchs, E., 1998. The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. Genes Dev 12, 3452-3463.

Davidson, E.H., 1986. Gene activity in early development, third edition. Academic Press, Orlando, FL.

Davidson, E. H., Erwin, D. H., 2006. Gene regulatory networks and the evolution of body plans. Science 311, 796-800.

Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., Feinberg, A.P., Cleton-Jansen, A.M., Moerland, E.W., Cornelisse, C.J., Suzuki, H., Komiya, A., Lindblom, A., Dorion-Bonnet, F., Neiman, P.E., Morse, H.C., Collins, S.J., Lobanenkov, V.V., 2002. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. Cancer Res 62, 48-52.

Glusman, G., Yanai, I., Rubin, I.,Lancet, D., 2001. The complete human olfactory subgenome. Genome Res 11, 685-702.

Herdegen, T., Leah, J. D., 1998. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. Brain Res Rev 28, 370-490.

Howard, E.W., Newman, L.A., Oleksyn, D.W., Angerer, R.C., Angerer, L.M., 2001. SpKrl: a direct target of β-catenin regulation required for endoderm differentiation in sea urchin embryos. Development 128, 365-375.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Cameron, R.A., Davidson, E.H., 2006a. Identification and characterization of transcription factor families in early *Strongylocentrotus* development. Dev Biol 300, 74-89.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Davidson, E.H., 2006b. Identification and characterization of homeobox transcription factor genes in *S. purpuratus*, and their expression in embryonic development. Dev Biol 300, 74-89.

Howard-Ashby, M., Materna, S.C., Tu, Q., C.T., Oliveri, P., Davidson, E.H., 2006c. High regulatory gene use in sea urchin embryogenesis. Dev Biol 300, 27-34.

Kaczynski, J., Cook, T., Urrutia, R., 2003. Sp1-and Krüppel-like transcription factors. Genome Biol 4.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33, 511-518.

Knight, R.D., Shimeld, S.M., 2001. Identification of conserved C_2H_2 zinc-finger gene families in the Bilateria. Genome Biol 2:research0016.1-16-8.

Kopelman, N.M., Lancet, D., Yanai, I., 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nature Genet 37, 588-589.

Krishna, S.S., Majumdar, I., Grishin, N. V., 2003. Structural classification of zinc fingers. Nucleic Acids Res 31, 532-550.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief in Bioinf 5, 150-163.

Laity, J.H., Lee, B.M., Wright, P.E., 2001. Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. 11, 39-46.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Livi, C.B., Davidson, E.H., 2006. Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. Dev Bio 293:513-525.

Lomberk, G., Urrutia, R., 2005. The family feud: turning off Sp1 by Sp1-like KLF proteins. Biochem J 392, 1-11.

Lu, D., Searles, M.A., Klug, A., 2003. Crystal structure of a zinc-finger–RNA complex reveals two modes of molecular recognition. Nature 426, 96-100.

Manzanares, M., Blanco, M.J., Nieto, M.A., 2004. Snail3 orthologues in vertebrates: divergent members of the Snail zinc-finger gene family. Dev Genes Evol 214, 47-53.

Manzanares, M., Locascio, A., Nieto, M.A., 2001. The increasing complexity of the Snail gene superfamily in metazoan evolution. Trends Genet 17, 178-181.

Mayor, R., Aybar, M.J., 2001. Induction and development of neural crest in *Xenopus laevis*. Cell Tissue Res 305, 203-209.

Miwata, K., Chiba, T., Horii, R., Yamada, L., Kubo, A., Miyamura, D., Satoh, N., Satou, Y., 2006. Systematic analysis of embryonic expression profiles of zinc finger genes in *Ciona intestinalis*. Dev Biol 292, 546-545.

Morgenstern, B., Atchley, W.R., 1999. Evolution of bHLH transcription factors: Modular evolution by domain shuffling? Mol Biol Evol 16, 1654-1663.

Nemer, M., Rondinelli, E., Infante, D., Infante, A.A., 1991. Polyubiquitin RNA Characteristics and Conditional Induction in Sea-Urchin Embryos. Dev Biol 145, 255-265.

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., Walhout, A.J., 2005. A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol 6, R110.1-110.12.

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

Rozen, S., Skaletzky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S., Eds., Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365-386.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Miklos, G.L.G., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., Cherry, J.M., Henikoff, S., Skupski, M.P., Misra, S., Ashburner, M., Birney, E., Boguski, M.S., Brody, T., Brokstein, P., Celniker, S. E., et al., 2000. Comparative genomics of the eukaryotes. Science 287, 2204-2215.

Ruvkun, G., and Hobert, O., 1998. The taxonomy of developmental control in *Caenorhabditis elegans*. Science 282, 2033-2041.

Sea Urchin Sequencing Consortium, 2006. The sequence of the sea urchin genome. Science 314, 941-952.

Samanta, M.P., Tongprasit, W., Istrail, S., Cameron, A., Tu, Q., Davidson, E.H., 2006. A high-resolution transcriptome map of the sea urchin embryo. Science 314, 960-962.

Shannon, M., Hamilton, A.T., Gordon, L., Branscomb, E., Stubbs, L., 2003. Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. Genome Res 13, 1097-1110.

Sweetman, D., Münsterberg, A., 2006. The vertebrate spalt genes in development and disease. Dev Biol 293, 285-293.

Tu, Q., Brown, C.T., Davidson, E.H., Oliveri, P., 2006. Sea urchin forkhead gene family: phylogeny and embryo expression. Dev Biol 300, 49-62.

Tupler, R., Perini, G., Green, M.R., 2001. Expressing the human genome. Nature 409, 832-833.

Urrutia, R., 2003. KRAB-containing zinc-finger repressor proteins. Genome Biol 4, 231.

Wang, D.G.W., Kirchhamer, C.V., Britten, R.J., Davidson, E.H., 1995. SpZ12-1, a negative regulator required for spatial control of the territory-specific cyiiia gene in the sea-urchin embryo. Development 121, 1111-1122.

Wang, Q.R., Lan, Y., Cho, E.S., Maltby, K.M., Jiang, R.L., 2005. Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development. Dev Biol 288, 582-594.

Wheelan, S.J., Church, D.M., Ostell, J.M., 2001. Spidey: A tool for mRNA-to-genomic alignments. Genome Res 11, 1952-1957.

Williams, A.J., Blacklow, S.C., Collins, T., 1999. The zinc finger-associated SCAN box is a conserved oligomerization domain. Mol Cell Biol 19, 8526-8535.

Wong, M.L., Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. Biotechniques 39, 75-85.

Yamada, L., Kobayashi, K., Degnan, B., Satoh, N., Satou, Y., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis* – IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. Dev Genes Evol 213, 245-253.

Zhang, Z.L., Gerstein, M., 2004. Large-scale analysis of pseudogenes in the human genome. Cur Opin Genet Dev 14, 328-335.

Chapter 3

HIGH ACCURACY, HIGH-RESOLUTION PREVALENCE MEASUREMENT FOR THE MAJORITY OF LOCALLY EXPRESSED REGULATORY GENES IN EARLY SEA URCHIN DEVELOPMENT

Stefan C. Materna, Jongmin Nam, Eric H. Davidson

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

(Published in Gene Expression Patterns 2010, 10:177-184)

Abstract

Accurate measurements of transcript abundance are a prerequisite to understand gene activity in development. Using the NanoString nCounter, an RNA counting device, we measured the prevalence of 172 transcription factors and signaling molecules in early sea urchin development. These measurements show high fidelity over more than five orders of magnitude, down to a few transcripts per embryo. Most of the genes included are locally restricted in their spatial expression, and contribute to the divergent regulatory states of cells in the developing embryo. In order to obtain high-resolution expression profiles from fertilization to late gastrulation, samples were collected at hourly intervals. The measured timecourses agree well with, and substantially extend, prior relative abundance measurements obtained by quantitative PCR. High temporal resolution permits sequences of successively activated genes to be precisely delineated, providing an ancillary tool for assembling maps of gene regulatory networks. The data are available via an interactive website for quick plotting of selected timecourses.

[Keywords: Transcription factor, Gene expression timecourse, mRNA prevalence measurement, Embryogenesis]

Results and Discussion

Measurement of transcript prevalence is a direct method to assess gene activity. In conjunction with spatial expression data, it enables a targeted approach to identify genes that contribute to a developmental process. Prevalence data are of particular usefulness if sampling intervals are short enough to yield timecourses with high temporal resolution. When genes are activated in close succession, such data can reveal the order in which activation occurs. When regulatory

genes are profiled, these data can be used to establish a very preliminary flow of regulatory information (Materna and Oliveri, 2008). This limits the number of possible regulatory interactions and may directly point at previously unknown regulatory linkages. High-resolution expression data may also guide perturbation experiments that directly reveal the regulatory relationships between genes, as accurate knowledge of gene expression profiles allows optimal time points for sampling to be chosen, thus maximizing information gain and robustness of results (Materna and Oliveri, 2008; Revilla-i-Domingo et al., 2007; Smith and Davidson, 2008).

Previous expression profiling efforts in sea urchin were aimed at finding all regulatory genes that are active during early development (Howard-Ashby et al., 2006a, b; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). While these data indicate approximately when genes are being used, they do not provide the needed temporal resolution. To provide a set of measurements that could be used for the above purposes, we collected sea urchin embryos at hourly intervals and determined transcript prevalence for the majority of regulatory genes that are expressed in a localized manner up to mid-gastrulation (36 hours post fertilization). We counted embryos and added known quantities of external RNA standards to the embryo lysate to yield absolute numbers of transcripts per embryo.

For transcript quantification we made use of the NanoString nCounter (NanoString Technologies, Seattle, WA), a high-throughput RNA molecule counter. The instrument is based on hybridization of gene specific probes to total RNA eliminating the need for a reverse transcription step (Geiss et al., 2008). The instrument visualizes individual hybridized probes, and identifies the transcript species based on a fluorescent bar code. The counts for each transcript species are linearly proportional to its prevalence in the added RNA solution, achieving sensitivity comparable to QPCR (Geiss et al., 2008). Using the nCounter, we obtained an extensive collection of high-fidelity expression profiles with high temporal resolution and greatly increased depth.

Codeset Design

The NanoString nCounter identifies and counts RNA species based on a fluorescent barcode attached to a sequence specific hybridization probe (Geiss et al., 2008). Our probe set includes 172 genes covering the majority of active and spatially restricted regulatory genes in the early *Srongylocentrotus purpuratus* embryo up to mid-gastrulation (36 hours postfertilization) (Howard-Ashby et al., 2006a, b; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). Table 3.1 summarizes the gene families represented in the codeset; a full list of genes, accession numbers, and probe sequences is available as Supplementary Table 1. Briefly, the codeset contains probes for 142 mRNAs encoding transcription factors, 20 mRNAs encoding signaling ligands or receptors, and 15 mRNAs used as markers or standards. Of the 142 transcription factors, 115 are known to be localized for at least part of early development whereas only four are expressed ubiquitously. The spatial expression pattern of the remaining 23 is unknown. For two transcription factors (*otx*, *blimp/krox*) two splice forms were included by designing probes against unique exons. Of the signaling related genes 14 are known to be spatially restricted while the rest are undetermined. Included in the list of transcription factors are 30 C_2H_2 zinc finger proteins, only half of which have clear orthologs with known transcriptional regulatory activity (Materna et al., 2006). Probes were designed against confirmed cDNA sequences where available. Gene predictions only were available for 67 genes (obtained from SpBase; Cameron et al., 2009), but these genes had been previously shown to be expressed using QPCR (Howard-Ashby et al., 2006a, b; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). For these genes, probes were designed in the immediate vicinity of the QPCR amplicon. For all genes a suitable probe was picked by NanoString in-house probe design. A full list of probe sequences is available in Appendix D.

Table 3.1: Contents of the NanoString nCounter Codeset Listed by Gene Family. (A full list is given as supplementary material.)

Family	No. of Genes
bHLH	15
bZIP	4
C_2H_2 zinc finger	30
ETS	7
Forkhead	15
Homeobox	39
Nuclear Hormone Receptor	6
SOX/HMG	7
T-Box	4
Miscellaneous TFs	15
Signaling - Ligands and antagonists	16
Signaling - Receptors	3
Markers/Standards	11

Prevalence Measurements with the NanoString nCounter

To determine the dynamic range of nCounter quantification assays we spiked 100 ng of sea urchin total RNA with different amounts of in vitro transcribed GFP and RFP RNA spanning a range of several orders of magnitude in concentration. As has been reported previously, the NanoString counts show a linear proportional relationship to transcript abundance in the hybridization reaction over more than five orders of magnitude (Geiss et al., 2008). With RNA from 200 embryos added to the hybridization reactions for the timecourse measurements below, no gene in our codeset reached a transcript level high enough to exceed the linear range. We determined the technical error associated with our measurements and confirmed

that variation between technical replicates is minimal, even for low counts, as reported previously (Geiss et al., 2008).

The background resulting from non-specific interactions between capture probes and detection probes was determined in four runs without RNA and was found to be insignificant. Across the probe set the median of the probe-specific background amounts to just three counts. It reaches double digits for only 10% of genes. Using Poisson statistics with $\lambda = 5$, the probability of observing 11 or more counts is < 0.01. For comparison, 50 NanoString counts correspond to 25 transcripts per embryo, a transcript level that is exceeded by even very lowly expressed genes such as *scl* and *snail*. Only six probes have a background count of 17 or higher, with 77 counts being the highest background observed (*nfe2*). Even for these genes the background is negligible compared to their expression level that in all cases reaches several thousand transcripts per embryo. In conclusion, with 200 embryos per hybridization reaction, the detection limit lies at a few transcripts per embryo for virtually all genes included in the codeset.

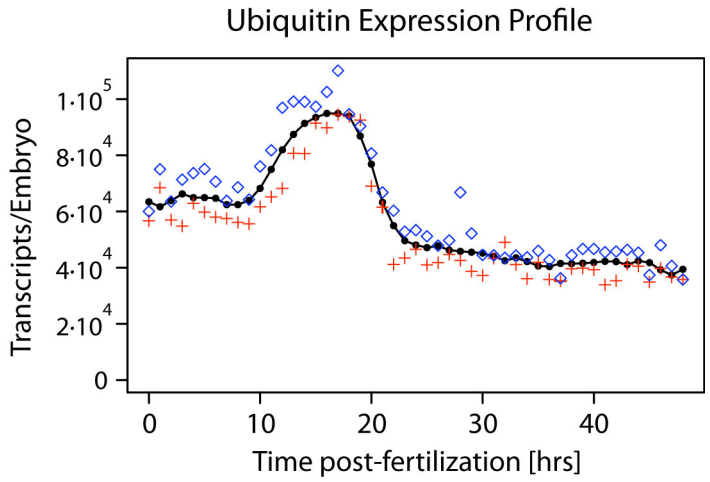


Figure 3.1: Expression profile for the ubiquitin domain. This domain has been used as an internal standard in previous expression profiling efforts (Howard-Ashby et al., 2006a, 2006b; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). The average prevalence is shown in black; data for each of the two embryo batches are shown in red and blue.

All quantification of nucleic acids based on hybridization of sequence-specific probes is susceptible to slight variation in hybridization efficiencies. According to the manufacturer NanoString probes may exhibit variation in hybridization efficiencies of up to twofold, but this may be exceeded by a few outliers. Although we have not tested the hybridization efficiency for each probe, the good agreement with prior QPCR data underlines the validity of our measurements. To illustrate this point: For the timecourse measurements described in the next section, 10^7 GFP and RFP molecules were added as an external standard. Based on the reported recovery efficiency of ~ 1% (Geiss et al., 2008), GFP and RFP were expected to produce about 100,000 counts. Indeed, GFP and RFP together produce a mean of almost exactly 100,000 counts after adjusting for hybridization efficiency. However, the counts for GFP are about 20% higher, while those for RFP are lower by roughly the same margin. Thus, even though very low abundance transcripts can be detected reproducibly with high fidelity, the absolute numbers of transcripts per embryo have to be viewed as an approximation.

NanoString nCounter Expression Profiling with High Temporal Resolution

The probe set described above was used to collect expression data from fertilization to late gastrulation (48 hours postfertilization). Samples were drawn at hourly intervals to capture in detail the dynamics of gene expression throughout early development. The relatively slow rate of transcription at 15°C in *Strongylocentrotus purpuratus* (Davidson, 1986) causes a considerable lag in subsequent gene activation events (Ben-Tabou de-Leon and Davidson, 2009; Bolouri and Davidson, 2003). Thus, this sampling interval is sufficient to reveal the order in which genes are activated, even where genes become transcribed in close succession. The counts obtained in the NanoString nCounter assay were normalized using the average counts for the external standards GFP and RFP to yield absolute numbers of transcripts per embryo. Examples of the measured timecourses are shown in Figs. 3.1–3; all timecourses are available in Appendix A.

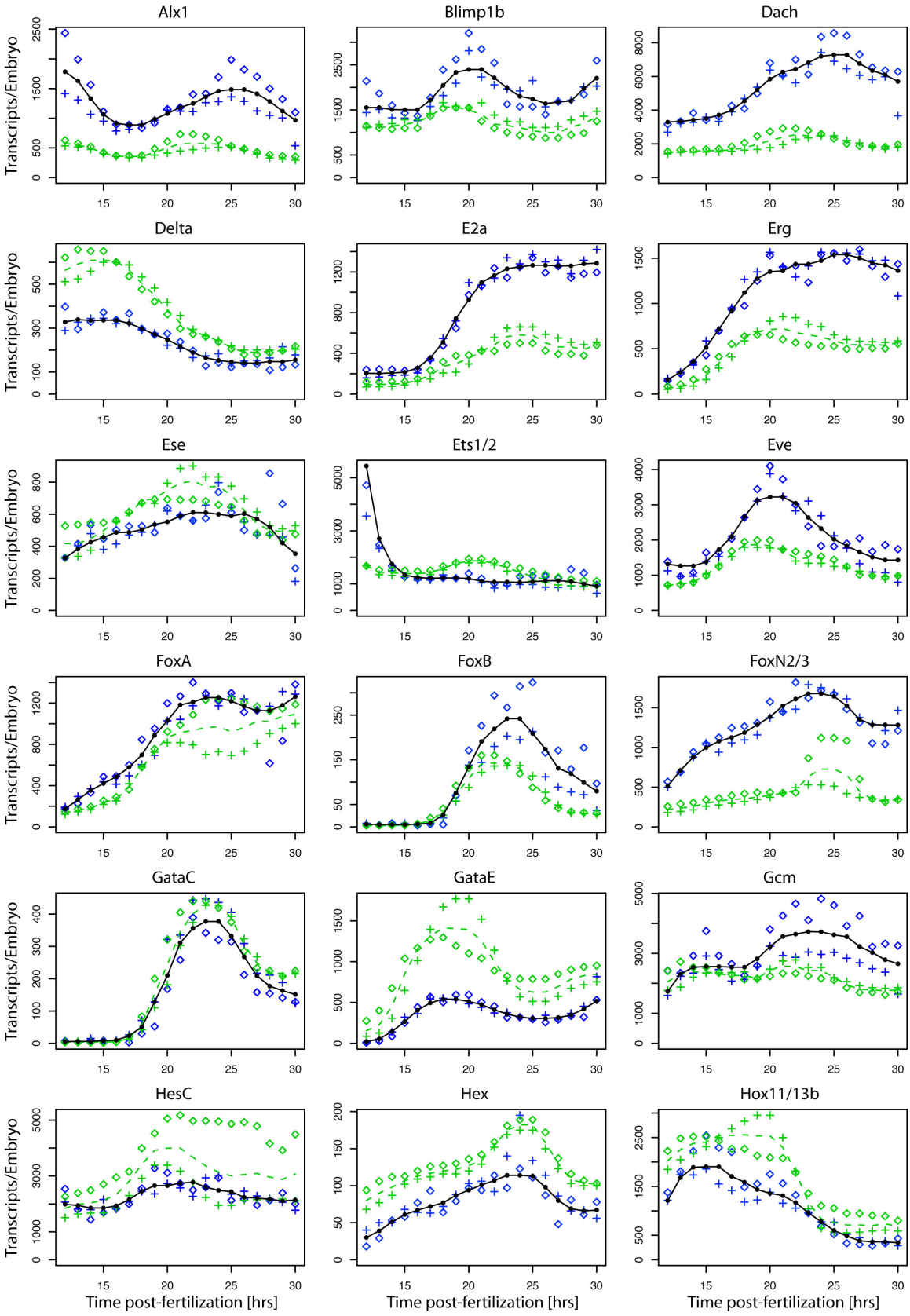
The poly-ubiquitin gene has been used in several earlier studies as an internal reference and has been assumed to be expressed at a constant level of about 88,000 ubiquitin domain copies per embryo, present in several copies in each transcript (Howard-Ashby et al., 2006a, b; Materna et al., 2006; Nemer et al., 1991). A graph showing the prevalence of ubiquitin domain copies as determined with the nCounter is presented in Fig. 3.1. The timecourse reveals a roughly twofold rise in prevalence between 10 hours and 22 hours postfertilization with a peak expression of about 105 domain copies per embryo. Thereafter it remains more or less constant throughout gastrulation at just under 50,000 ubiquitin domain copies per embryo, an altogether minor difference from the previous assumption.

The prevalence of the *spz12* message had been measured previously by RNA probe excess titration (Wang et al., 1995). Four time points of this earlier study fall into the period covered by our measurements and show the same overall expression profile. However, the absolute numbers we obtain are about 1.5 to twofold lower. RNA probe excess titration is highly sensitive and not susceptible to variation in hybridization efficiency, but the absolute values obtained depend directly on the accuracy of the probe-specific activity estimation (Davidson, 1986). Absolute prevalence as determined by this method may vary by a factor of two. The confidence intervals for the RNA probe excess titration measurements of *spz12* and our nCounter data overlap, confirming the fidelity of the present measurements.

Batch-to-batch variation of gene expression in the nCounter assays is apparent for about 40% of genes included in the codeset — a significant fraction. The remaining genes have essentially identical expression profiles in both batches. These differences indicate polymorphisms among the four genomes included in the comparison which affect synthesis or turnover rates for specific transcripts, not surprising given the very high level of genomic polymorphism characteristic of this species (Sodergren et al., 2006). Polymorphic differences are usually limited to given times, and only few genes show significant discrepancies throughout. The magnitude of variation generally does not exceed threefold. Many of the most noticeable differences are found very early, up to about ten hours post-fertilization. This is likely due to differences in maternal transcript accumulation (e.g., *ets1/2*, *z48*). In addition to polymorphisms that affect RNA synthesis or turnover, variation in maternal deposition may

contribute to variation of zygotically expressed genes downstream. Other common differences include expression peaks that reach different levels in different batches or the timing of a ramping step. For the latter, differences usually persist until a peak or plateau has been reached.

We compared a subset of the expression profiles collected in the nCounter assays to independently gathered high-resolution QPCR data (Fig. 3.2). Again two independent batches of fertilized eggs were assayed. In these QPCR assays, prevalence was measured relative to the ubiquitin domain transcript as the internal standard and converted to absolute transcript numbers per embryo using the prevalence data for the ubiquitin domain obtained with the nCounter (Fig. 3.1). In general, prevalence and overall shape of the expression profile are in excellent agreement. As might be expected from the fact that there are now eight genomes rather than four in the dataset, some genes that display no batch-to-batch variation in the nCounter assays display variation between the batches used for QPCR (e.g., *foxA*, *gataE*, *hesC*, Fig. 3,2). Vice versa, there are genes that show batch-to-batch variation in the QPCR but not the nCounter assays (e.g., *six1/2*, *z48*, Fig. 3.2).


Alx1
Blimp1b
Dach
Delta
E2a
Erg
Ese
Ets1/2
Eve
FoxA
FoxB
FoxN2/3
GataC
GataE
Gcm
HesC
Hex
Hox11/13b
Transcripts/Embryo
Time post-fertilization [hrs]

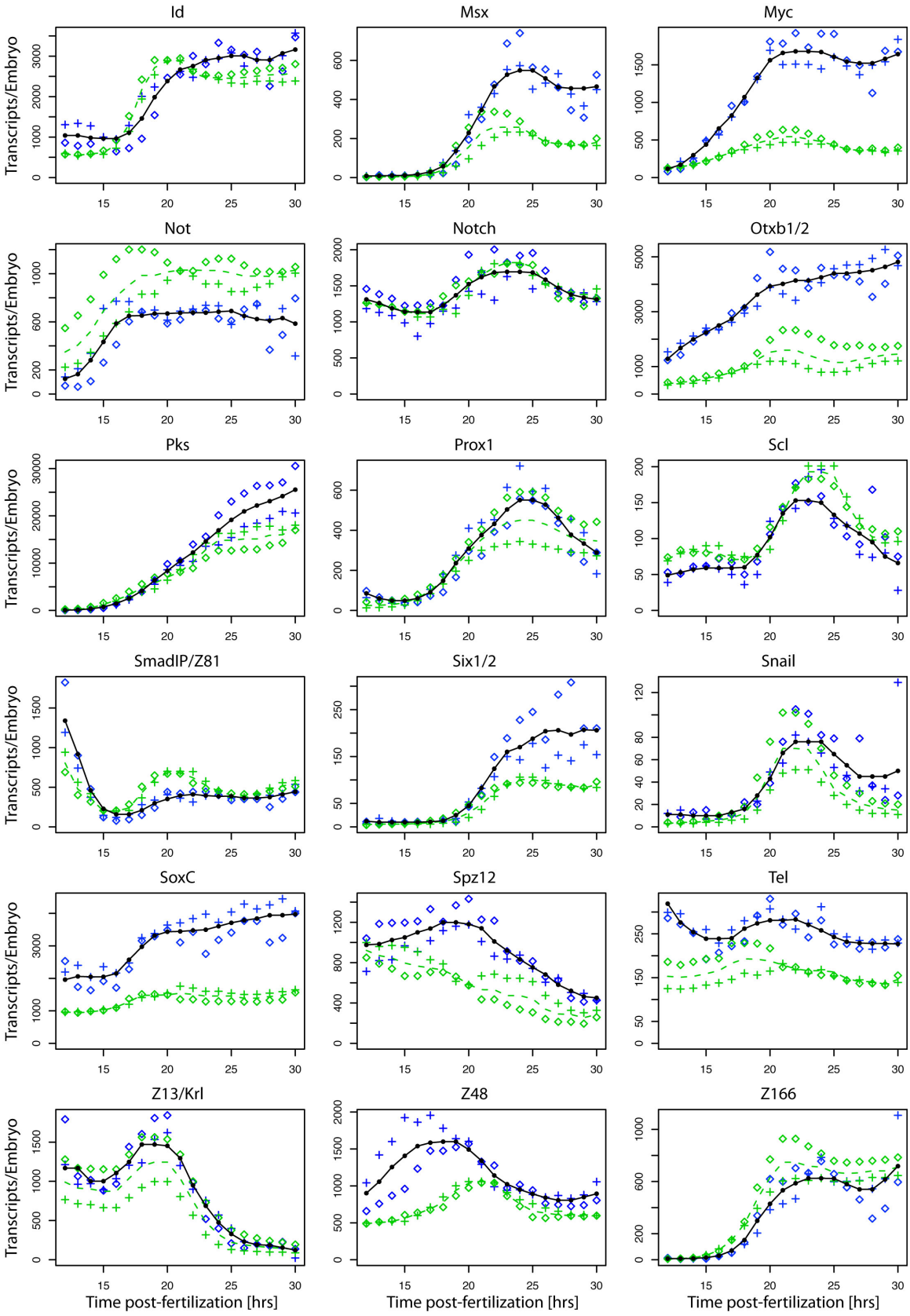

Id
Msx
Myc
Not
Notch
Otxb1/2
Pks
Prox1
Scl
SmadIP/Z81
Six1/2
Snail
SoxC
Spz12
Tel
Z13/Krl
Z48
Z166
Transcripts/Embryo
Time post-fertilization [hrs]

Figure 3.2: Comparison of NanoString prevalence measurements and QPCR data. The average transcript abundance as measured with the NanoString nCounter is shown in black, data for the two embryo batches as blue crosses and diamonds. Average prevalence data collected in QPCR assays are shown as dashed green lines, data for the two embryo batches used in these assays as green crosses or diamonds. Absolute transcript numbers were computed using the ubiquitin domain numbers obtained in the nCounter assays (Fig. 3.1).

Prominent differences between expression profiles collected in QPCR and nCounter assays were observed for eight out of 36 genes. In these cases the nCounter prevalence data are usually higher than those obtained with QPCR and they remain higher by a constant factor throughout their expression (Fig. 3.2: *alx1*, *dach*, *e2a*, *erg*, *foxn2/3*, *myc*, *otxb1/2*, *soxc*). Absolute values aside, the expression profiles are more or less identical, regardless of the quantification method. In addition, almost no batch to batch variation is observed for these genes between the embryo batches in either QPCR or nCounter assays. This indicates that the discrepancies are likely due to differences in hybridization efficiencies of NanoString probe and QPCR primers. Since QPCR values are lower than those obtained with the nCounter, it is most likely that the QPCR primers were suboptimal.

Only two transcripts showed significantly higher prevalence in QPCR compared to nCounter assays (Fig. 3.2; *delta*, *gataE*). But these genes, *gataE* in particular, display significant batch to batch differences in the QPCR assays, indicating that this is the likely cause for the discrepancy. Together these findings suggest that nCounter assays are less prone to the minor quantitative problems affecting absolute transcript determinations by QPCR, such as low amplification efficiency in the first few cycles due to secondary structures in the cDNA, and polymorphisms that result in poor primer hybridization (Geiss et al., 2008; Wong and Medrano, 2005).

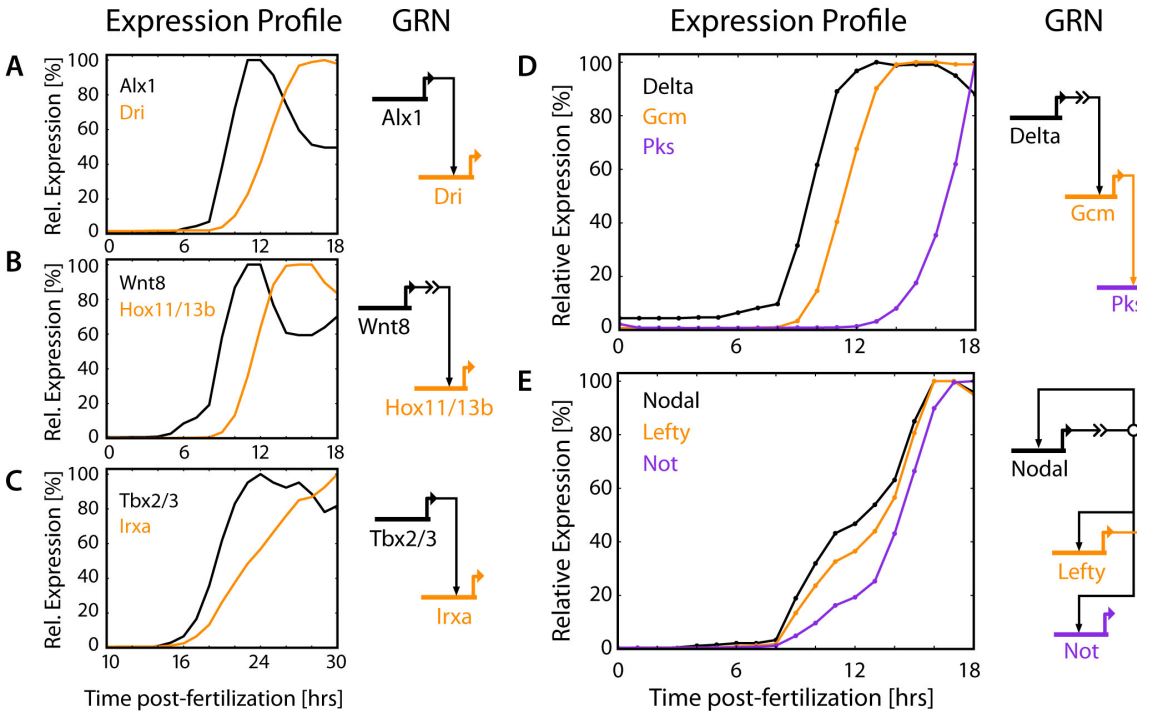


Figure 3.3: Expression dynamics of genes with close regulatory relationships. (A–C) Downstream targets often show a visible lag after their main activator first becomes transcribed. (A) Alx drives *dri* transcription in the skeletogenic cells of the sea urchin (Oliveri et al., 2008). (B) Wnt8 signaling leads to accumulation of Tcf1 in the nucleus of endomesodermal cells resulting in transcription of *hox11* (Peter and Davidson, 2010). (C) Tbx2/3 activates *irxa* transcription in the aboral ectoderm of the early sea urchin embryo (Su et al., 2009). (D) Delta/Notch signaling from the skeletogenic cells activates *gcm* in the neighboring endomesodermal cells. Gcm in turn is a direct activator of *pks* (Ransick and Davidson, 2006; Calestani and Rogers, 2010). (E) More complex regulatory relationships are reflected in the expression dynamics. Nodal activates its own transcription, *lefty*, and *not*. Lefty is an antagonist of Nodal signaling (Duboc et al., 2008; Nam et al., 2007; Range et al., 2007).

High-Resolution Expression Data as a Tool

Gene expression data are a useful ancillary tool for assembling maps of gene regulatory networks. High temporal resolution produces a detailed picture of the expression dynamics for he genes under study and can indicate which genes could be linked at the regulatory level.

However, they are no substitute for perturbation experiments, the effects of which directly reveal causal regulatory relationships among genes. High resolution expression data can, however, augment and guide perturbation experiments and expedite the discovery of unknown linkages (Materna and Oliveri, 2008).

Where a regulatory gene serves as the main driver of a downstream gene, the target is often expressed with a visible delay after the activator is first transcribed. This is due to the time needed to produce a sufficient amount of protein for activation to occur. Kinetic analyses using the transcription, turnover and protein synthesis rates for *S. purpuratus* have demonstrated a typical lag of about two hours between activation of an upstream gene and activation of its direct target (Ben-Tabou de-Leon and Davidson, 2009; Bolouri and Davidson, 2003). This is illustrated by the gene pairs in Fig. 3.3A–C. The activators in Figs. 3.3A and 3.3C are likely direct inputs into their downstream target genes, although some of the connections have yet to be confirmed at the cis-regulatory level (Nam et al., 2007; Oliveri et al., 2008; Ransick and Davidson, 2006; Su et al., 2009). The activation in Fig. 3.3B is indirect: Wnt8 is a signaling factor. Signal reception causes accumulation of a positively active form of the transcription factor Tcf1 in the nucleus, to which the target gene *hox11/13b* responds (Peter and Davidson, 2010, C. Theodoris, J. Smith, and E. Davidson, unpublished data; for kinetic analysis: Bolouri and Davidson, 2010). Figure 3.3D shows an example of three genes linked in a rather simple activation cascade. The signaling ligand Delta is presented on the surface of the skeletogenic cells, and via its effector, Suppressor of Hairless, activates *gcm* in the neighboring endomesodermal cells (Ransick and Davidson, 2006). Gcm in turn directly activates *pks* (Calestani and Rogers, 2010). However, activation of *pks* may require genes in addition to *gcm* as indicated by the relatively long period between *gcm* and *pks* activation.

Despite these clear-cut examples, the connection between expression dynamics and how genes are linked at the regulatory level might become clear only in hindsight. After all, gene activation is combinatorial and often requires several activating inputs acting in concert for a target gene to turn on. In addition, the flow of regulatory information is often not as unidirectional as in the examples above. Regulatory genes are frequently part of feedback loops and influence each other's expression, as in the example in Fig. 3.3E: the signaling ligand Nodal is initially

activated by bZIP factors but, once expressed, strongly activates itself (Nam et al., 2007; Range et al., 2007). Nodal also directly activates *lefty* transcription. Lefty, in turn, antagonizes the activity of Nodal thus effectively limiting its transcription (Duboc et al., 2008). *not*, a second downstream target of Nodal (S.C. Materna, E. Li and E.H. Davidson, unpublished data) has a similar expression profile. The two feedback loops between *nodal* and *lefty* make it impossible to infer the linkage pattern from timecourse data alone. However, perturbation experiments and cis-regulatory analysis have clearly established their relationship (Duboc et al., 2008; Nam et al., 2007; Range et al., 2007; Su et al., 2009). The kinetic aspescts of these interactions have been discussed elsewhere (Bolouri and Davidson, 2009).

In conclusion, we have acquired a set of high-quality prevalence data for the majority of spatially restricted regulatory genes in early sea urchin development. The data describe in detail the expression dynamics of the genes included. This information will be useful for unraveling the gene regulatory network underlying early sea urchin development.

Data availability

In order to allow quick access to the expression data for all genes we have created a simple visualization tool that is available via http://sugp.caltech.edu/endomes/. A screenshot is shown in Fig. 3.4. The website enables easy viewing of multiple timecourses covering the whole 48 hours of early development, or fractions thereof, as determined by the user. The data can be plotted either on a linear or logarithmic scale, to accommodate genes with different expression levels. Two options for plotting relative expression levels are available: The first relates the gene expression level to the maximum measured for this gene during the whole time period covered, regardless of the period chosen for plotting. The second option sets the maximum expression during the plotting period to 100%. These options are useful when the timing of gene expression matters more than the absolute transcript levels — as is often the case when examining regulatory linkages. The data for the two individuals sampled can be plotted if desired. With this tool all graphs displaying nCounter data in this paper can

essentially be recreated in print quality. A table with all expression data is available upon request.

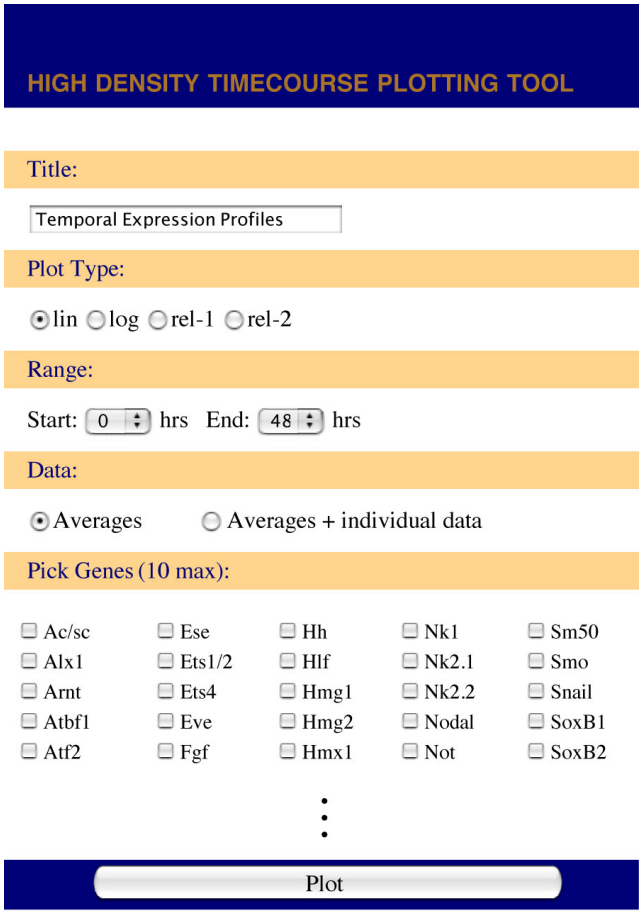


Figure 3.4: Screen shot of the online plotting tool. This website enables easy plotting of all expression profiles. What genes are to be plotted and the details of the presentation can be specified by the user. The website is available via http://sugp.caltech.edu/endomes/.

Materials and Methods

Embryo culture and RNA extraction

Sea urchin embryos were fertilized in filtered seawater and washed repeatedly to remove excess sperm. Embryos were cultured at low density at 15ºC and closely monitored for proper development. For the timecourse measurements 200 embryos were counted for each time point. Samples were collected hourly just prior to lysis in 350 µl RLT buffer from the Qiagen RNeasy Micro Kit (Qiagen, Hilden, Germany). Embryo lysates were stored at -70ºC until use. *gfp* and *rfp* genes were transcribed in vitro and $1x10^7$ transcripts of both GFP and RFP were added to each lysate after thawing. RNA was extracted according to manufacturers instructions but, to maximize recovery, RNA was eluted with 50 µl nuclease-free water. The samples were ethanol precipitated and resuspended in 5 µl nuclease-free water, all of which was used in the following NanoString hybridization.

NanoString nCounter assays

For each individual and timepoint, transcript prevalence was measured using the NanoString nCounter. Hybridization reactions were performed according to manufacturer's instructions with 5 µl RNA solution. Care was taken to minimize the time after addition of capture probeset in order to minimize background due to nonspecific interactions between detection probes and capture probes. All hybridization reactions were incubated at 65ºC for a minimum of 18 h. Hybridized probes were recovered with the NanoString Prep Station and immediately evaluated with the NanoString nCounter. For each reaction 600 fields of view were counted.

The resulting counts were normalized using the mean of GFP and RFP counts for each time point and converted to transcript numbers per embryo by using the known numbers of

GFP/RFP transcripts and embryos. For each gene and time point the arithmetic mean was calculated for the expression data from the two individuals sampled. To discard outliers in the average timecourse, a running average was calculated over five time points, discarding the minimum and maximum value and computing the arithmetic mean of the three remaining values.

QPCR assays

For quantitative PCR assays, RNA was collected from embryos of two different females. RNA was extracted with the Qiagen RNeasy Mini kit. 1 μg of total RNA was converted to cDNA using the BioRad iScript cDNA synthesis kit (BioRad, Carlsbad, CA). QPCR was performed with the BioRad SYBR Green reagent on an AB 7900 HT instrument (Applied Biosystems, Foster City, CA). Data were evaluated with the dCt method using poly-ubiquitin as the reference gene (Materna and Oliveri, 2008). dCt values were converted to transcripts-per-embryo by applying the transcript numbers for poly-ubiquitin obtained with the NanoString nCounter. Data were averaged as above.

All timecourses are available in Appendix A, probe sequences and accession numbers in Appendix D.

Acknowledgements

We would like to thank the NanoString team, and in particular Stephen Jackson and Sean Ferree, for their tireless support.Research was supported by NIH grant HD 37105. Most of the funding for the purchase of the nCounter was a gift from the Ahmanson Research & Equipment Fund.

References

Ben-Tabou de-Leon, S., Davidson, E.H., 2009. Modeling the dynamics of transcriptional gene regulatory networks for animal development. Dev Biol 325, 317–328.

Bolouri, H., Davidson, E.H., 2003. Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics. Proc Natl Acad Sci U S A 100, 9371–9376.

Bolouri, H., Davidson, E.H., 2009. The gene regulatory network basis of the "community effect," and analysis of a sea urchin embryo example. Dev Biol 340, 170–178.

Calestani, C., Rodgers, D. 2010. Cis-regulatory analysis of the sea urchin pigment cell gene polyketide synthase. Dev Biol 340, 249–255.

Duboc, V., Lapraz, F., Besnardeau, L., Lepage, T., 2008. Lefty acts as an essential modulator of Nodal activity during sea urchin oral-aboral axis formation. Dev Biol 320, 49–59.

Geiss, G., Bumgarner, R., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D., Fell, H., Ferree, S., George, R., Grogan, T., James, J., Maysuria, M., Mitton, J., Oliveri, P., Osborn, J., Peng, T., Ratcliffe, A., Webster, P., Davidson, E., Hood, L., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26, 317–325.

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

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

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

Materna, S.C., Oliveri, P., 2008. A protocol for unraveling gene regulatory networks. Nat Protoc 3, 1876–1887.

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

Nemer, M., Rondinelli, E., Infante, D., Infante, A.A., 1991. Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos. Dev Biol 145, 255–265.

Oliveri, P., Tu, Q., Davidson, E.H., 2008. Global regulatory logic for specification of an embryonic cell lineage. Proc Natl Acad Sci USA 105, 5955–5962.

Peter, I.S., Davidson, E.H., 2010. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. Dev Biol 340, 188–199.

Range, R., Lapraz, F., Quirin, M., Marro, S., Besnardeau, L., Lepage, T., 2007. Cis-regulatory analysis of nodal and maternal control of dorsal-ventral axis formation by Univin, a TGF-beta related to Vg1. Development 134, 3649–3664.

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

Revilla-i-Domingo, R., Oliveri, P., Davidson, E.H., 2007. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. Proc Natl Acad Sci U S A 104, 12383–12388.

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

Smith, J., Davidson, E.H., 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, 20089–20094.

Sea Urchin Genome Consortium. 2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. Science 314, 941–952.

Su, Y.-H., Li, E., Geiss, G.K., Longabaugh, W.J.R., Krämer, A., Davidson, E.H., 2009. A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. Dev Biol 329, 410–421.

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

Wang, D.G., Kirchhamer, C.V., Britten, R.J., Davidson, E.H., 1995. SpZ12-1, a negative regulator required for spatial control of the territory-specific CyIIIa gene in the sea urchin embryo. Development 121, 1111–1122.

Wong, M.L., Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. Biotechniques 39, 75–85.

Chapter 4

A COMPREHENSIVE ANALYSIS OF DELTA SIGNALING IN PRE-GASTRULA SEA URCHIN EMBRYOS.

Stefan C. Materna, Eric H. Davidson

Abstract

In sea urchin embryos Delta signaling specifies non-skeletogenic mesoderm. Despite the identification of some direct targets, several aspects of D/N signaling remain supported only by circumstancial evidence. To obtain a detailed picture of Delta function, we followed a systems-biology approach and evaluated the effects of D/N perturbation on expression levels of 205 genes up to gastrulation. The gene set includes virtually all transcription factors that are expressed in a localized fashion by mid-gastrulation and thus provide spatial information to the embryo. Also included are signaling factors and pigment -cell-differentiation genes. We show that the number of early, and likely direct, targets is small and almost exclusively restricted to non-skeletogenic mesoderm (NSM) genes. However, Delta signaling also activates *foxY* in the small micromeres that, like NSM, are in direct contact with *delta*-expressing skeletogenic mesoderm (SM). In contrast, endoderm genes are not activated by Delta signaling even when *delta* is expressed a second time, now in the NSM that is in direct contact with the endoderm. Instead Delta is an ongoing input and activates only *foxY* expression in small micromeres. Disruption of the second Delta phase specifically abolishes specification of late mesodermal derivatives such as the coelomic pouches to which the small micromeres contribute.

Introduction

The Delta signaling ligand is an important regulator of developmental processes across the animal kingdom. In contrast to many other signaling ligands it is bound to the cell surface of the delta expressing cell and not secreted. This limits its effective range to cells that are in direct contact with the source (Wang, 2011). In the receiving cell, Delta binds to the Notch receptor, causing cleavage of its intracellular domain (N_{ic}). N_{ic} then enters the nucleus where it binds to the co-activator Suppressor of Hairless (Su(H)) to activate transcription of target genes. In the absence of nuclear N_{ic}, Su(H) is bound to the co-repressor Groucho and

performs repressive function. Thus, D/N signaling operates as a toggle switch (Ransick and Davidson, 2006).

In sea urchins D/N signal specifies all non-skeletogenic mesoderm (NSM) cell types, such as pigment cells, blastocoelar cells, coelomic pouch cells and circumesophageal muscle (Sherwood and McClay, 1999; Sweet et al., 2002). *delta* is first expressed between 8 and 9 hpf in the skeletogenic mesoderm (SM) at the center of the vegetal plate. Initially it is received by the surrounding ring of veg2 endomesodermal cells. At 7th cleavage, the ring of veg2 cells divides into an inner ring that will develop into NSM, and an outer that is specified as endoderm (Peter and Davidson, 2010, 2011; Ruffins and Ettensohn, 1996). Continued reception of the Delta signal is essential for NSM specification. When ingression of SM gets under way expression of delta expression shifts to the NSM. Prior studies of D/N signaling have suggested that the first, or SM, Delta signal is responsible for specification of the earliest NSM cell types (i.e., pigment and blastocoelar cells), whereas NSM Delta function is to specify late mesoderm derivatives (such as coelomic pouch cells and muscles) (Sweet et al., 2002).

D/N signaling across the NSM/endoderm boundary has been discussed in previous studies. However, evidence for it remains largely circumstantial and is based on experiments with activated Notch (N_{act}) that expand endoderm at the expense of ectoderm and previous findings that some endodermal genes (e.g., *gataE*) are affected in D/N perturbations (Davidson et al., 2002; Sweet et al., 2002). Direct evidence for activation of endodermal genes has not been reported. Rather, it has been shown that D/N signaling serves to deactivate endodermal genes in the NSM precursors (Croce and McClay, 2010; Peter and Davidson, 2011).

In this study we follow a systems-biology approach to examine in greater detail the function of both SM and NSM Delta signaling. We analyze the effect of D/N perturbation on transcript levels of 205 genes. This gene set includes the majority of transcription factors that are specifically activated during early development up to mid-gastrulation, and all transcription factors that are know to be spatially restricted in their expression. In addition it contains signaling molecules with known function and pigment-cell-differentiation genes.

We find that the number of early, and thus presumably direct, targets is low. In fact, by the time SM *delta* expression comes to an end, only 6 NSM transcription factors have been activated, two of which are known direct D/N targets. Thus the large number of NSM genes that become activated in the NSM after the termination of SM *delta* expression must be indirect targets that are downstream of the early responders.

In contrast to NSM genes, no endoderm genes are found to be activated by D/N throughout the time period covered, or they are only transitory. Given that our gene set includes all known endoderm transcription factors, this finding thus proves that endoderm does not receive D/N at any time prior to gastrulation. Instead, the Delta signal is received in the small micromeres where it activates *foxY*. By specifically perturbing the function of second, or NSM, Delta, we show that it is required for maintenance of foxY expression into gastrulation. The phenotype of late Delta-perturbed embryos confirms earlier findings that its function is to specify late mesodermal derivatives, notably coelomic pouches, whereas it plays no role in pigment cell specification.

Materials and Methods

Delta/Notch perturbations

MASOs were obtained from Gene-tools LLC and injected at 300 μM in 0.12 M KCl. Injection volumes were about 5 pl. Sequences are as follows:

- Delta — CAAGAAGGCAGTGCGGCCGATCCGT
- Notch — CCTGGATGGGTAGTCCGCCTCATCT.

The dominant negative (DN) Su(H) contains a mutation in its DNA binding domain that prevents it from binding DNA, while leaving its interaction with N_{ic} and other proteins unaffected (Ransick and Davidson, 2006). DN-Su(H) mRNA was injected at 0.2 ng/μl in 0.12 M KCl. The γ-secretase inhibitor DAPT (N- [N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester) (Hughes et al., 2009) was dissolved in DMSO and added at 3 hpf or 17 hpf to a final concentration of 8 μM. Higher concentrations caused all embryos to exogastrulate, and a concentration of higher than 20 μM caused severe, and non-specific, defects. Lower inhibitor concentrations resulted in higher numbers of pigment cells.

Embryo culture and RNA extraction

Sea urchin embryos were cultured at 15°C and closely monitored for proper development. For lysis, sea water was removed before adding 350 μl RLT buffer from the Qiagen RNeasy Micro Kit (Qiagen, Hilden, Germany). Embryo lysates were immediately stored at -70°C until use. RNA was extracted according to manufacturer's instructions; to maximize recovery, RNA was eluted with 50 μl nuclease-free water. Samples were ethanol precipitated and resuspended in 11

μl nuclease-free water. Samples were split and 5 μl were used in Nanostring nCounter assays. The leftovers were reverse transcribed.

Transcriptional profiling

For each timepoint and condition transcript, prevalence was measured using the NanoString nCounter (probe sequences are given in Appendix D). Hybridization reactions were performed according to manufacturer's instructions with 5 μl RNA solution. Care was taken to minimize the time after addition of capture probe set in order to minimize background due to non-specific interactions between detection probes and 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. For each reaction 1150 fields of view were counted. The resulting counts were normalized using the sum of all counts for all genes in the codeset. Fold differences were calculated between experiment and control counts.

For quantitative PCR assays, leftover RNA was converted to cDNA using the BioRad iScript cDNA synthesis kit (BioRad, Carlsbad, CA). QPCR was performed with the BioRad SYBR Green reagent on an AB 7900 HT instrument (Applied Biosystems, Foster City, CA). Data were evaluated with the dCt method, using poly-ubiquitin as the reference gene (Materna and Oliveri, 2008). ddCt values were calculated between experiment and control embryos and converted to fold differences to be comparable with Nanostring data. QPCR primer sequences are available in Appendix B.

Whole-mount in situ hybridization

Probe templates were amplified from cDNA by PCR (primer sequences or source of probe template are list in Appendix C). DIG-labelled antisense probes were transcribed with Roche Sp6 or T7 RNA polymerase. Embryos were fixed in 2.5% glutaraldehyde, 32.5% sea water, 32.5 mM MOPS (pH7), and 162.5 mM NaCl on ice overnight. Embryos were treated with Proteinase K for 5 min at room temperature (25 ng/µl in TBST) followed by a 30 min fixation step in 4% paraformaldehyde, 32.5% sea water, 32.5 mM MOPS (pH7), and 162.5 mM NaCl at room temperature. Hybridizations were performed using a standard protocol. Probes were hybridized over night at 65°C using a concentration of 1 ng/µl hybridization buffer. Probes were detected using anti-DIG Fab fragments conjugated to alkaline phosphatase (1/1000 dilution) and NBT/BCIP.

Results

delta *expression in the early sea urchin embryo*

The sea urchin *delta* gene is first expressed between 8 and 9 hours postfertilization (hpf) in the cells of the SM lineage that lie at the center of the vegetal plate (Revilla-i-Domingo et al., 2007; Sweet et al., 2002). The expression reaches its peak at 12 hpf with only a few hundred transcripts per embryo (Materna et al., 2010). In addition to the SM, *delta* transcript can also be detected as early as 12 hpf in the apical domain, albeit weakly (Fig. 4.1A). *delta* expression ceases in the SM as these cells prepare for ingression (between 18 and 19 hpf). At the same time that SM cells terminate *delta* transcription, *delta* transcripts appear in the entire NSM. Initially, the NSM forms a ring around the SM (Figs. 4.1C, C'), but when ingression is complete it has replaced the SM at the center of the vegetal plate (Fig. 4.1D'). At this stage (24 hpf) strong *delta* expression is also visible in the apical domain. There, it is expressed in only a

few cells that align in a row or small cluster that appears to be off-center relative to the middle of the apical domain (Figs. 4.1E, E').

delta expression is activated by ubiquitously expressed transcription factors, most notably Runx (Fig. 4.1 K) (Robertson et al., 2008; Smith and Davidson, 2008), but its spatial expression is tightly regulated by the widely expressed repressor HesC (Revilla-i-Domingo et al., 2007; Smith and Davidson, 2008). In the SM *hesC* is repressed by Pmar, while in the NSM *hesC* is repressed by Blimp. We show here that *hesC* expression has already disappeared from the NSM even before ingression of SM cells is complete (Fig. 4.1 J). HesC does not clear from the neighboring endodermal cells and thus limits expression of *delta* to NSM.

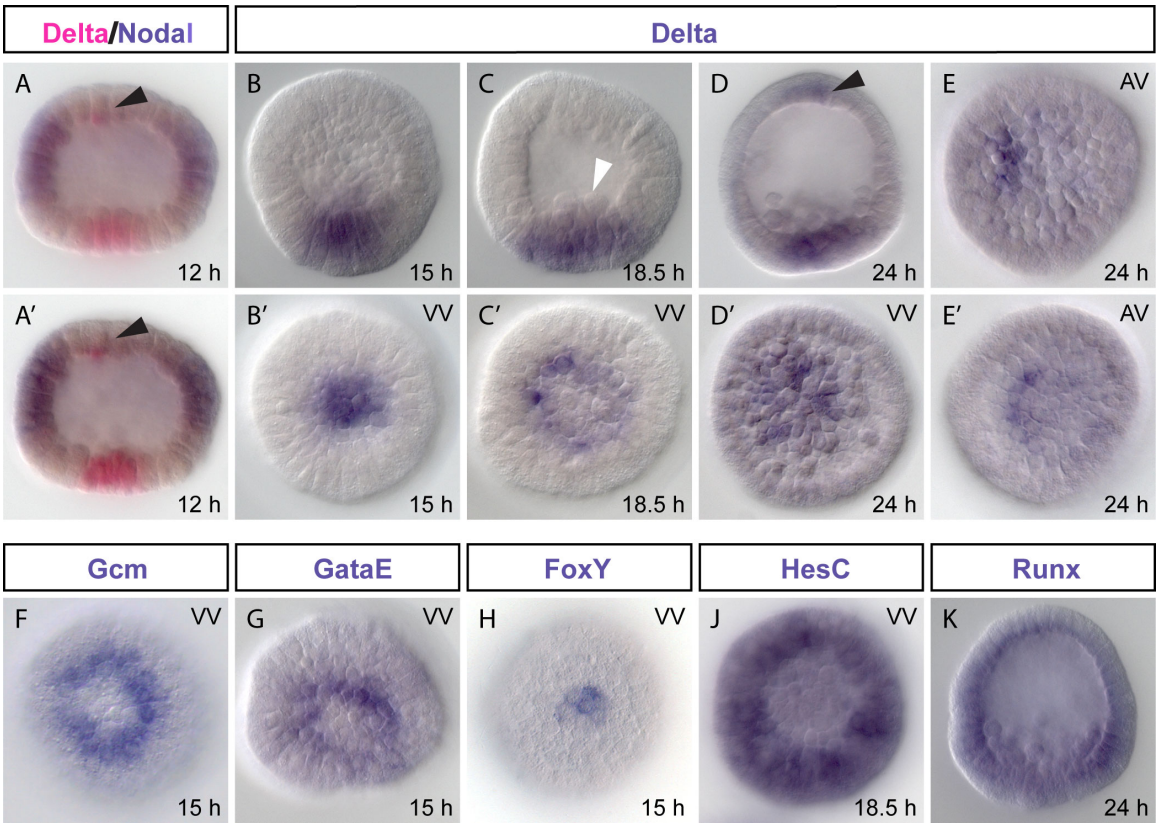


Figure 4.1: Expression patterns of the *delta* gene and its early targets. (A, B) *delta* transcripts are first detectable between 8 hpf and 9 hpf and are localized to the skeletogenic lineage. (C, C') Between 18 and 19 hpf the skeletogenic cells lose *delta* expression as they ingress into the blastocoel. At the same time the mesoderm that is adjacent to the skeletogenic cells starts to express delta. (D. D') After ingression of the skeletogenic cells is complete the mesoderm occupies the center of the vegetal plate and expresses delta throughout. In addition, *delta* is

expressed in the apical plate. With strong staining, *delta* expression can be observed at the apical plate as early as 12 hpf (black arrow head in A, A'), but is more easily detected at 24 hpf (black arrowhead in D, D'). (E, E') Apical *delta* expression is limited to a few cells that appear to be slightly off center of the apical domain. (F, G) *gcm* and *gataE*, the direct early targets of Delta, are expressed in a ring of mesodermal precursor cells that surround the Delta source. (H) *foxY* is expressed in the small micromeres, which are surrounded by *delta* expressing cells. (J) By the time *delta* transcription is activated in the mesoderm, *hesC*, a strong repressor of *delta*, has turned off there (compare to C'). (K) The *delta* activator Runx is expressed ubiquitously throughout the sea urchin embryo. All embryos are oriented with their oral side to the left. In lateral views apical is at the top. VV — vegetal view.

Perturbation of early (SM) Delta signaling

SM Delta functions to activate NSM genes.

Following binding of the Delta ligand, N_{ic} gets cleaved and nuclearizes where it binds to the co-activator Su(H) to activate target genes. To obtain a detailed picture of the function of Delta signaling in early sea urchin development, we employed several perturbation techniques: Injection of morpholino-substituted antisense oligonucleotides (MASO) to inhibit translation of the Delta ligand and the Notch receptor; expression of a dominant negative form of Su(H) that contains a mutation in its DNA binding domain (Ransick and Davidson, 2006); and treatment with the γ-secretase inhibitor DAPT that inhibits cleavage of the Notch intracellular domain (N_{ic}) following signal reception (Hughes et al., 2009). When these perturbations are performed, the resulting embryos fail to specify NSM, and consequently will lack pigment cells and blastocoelar cells (or develop only a few), and will fail to form coelomic pouches and circumesophageal muscle (Sherwood and McClay, 1999; Sweet et al., 2002). Usually, a significant fraction of embryos will exogastrulate when the perturbation agent is injected at fertilization.

We collected perturbed embryos at four timepoints in short succession (12, 15, 18, 24 hpf) to identify the genes that are activated by D/N signaling in a time-resolved manner. RNA was

extracted from these embryos and quantified using the Nanostring nCounter, an RNA counting device (Geiss et al., 2008). We used different probe sets that together contain 182 genes. This includes the majority of all transcription factors known to be expressed in a localized fashion by 36 hours (Howard-Ashby et al., 2006a, 2006b, 2006c; Materna et al., 2006; Rizzo et al., 2006; Tu et al., 2006). This data was supplemented with a number of pigment-cell-specific genes that were not covered by the codeset; transcript abundance of these genes was determined by quantitative PCR (QPCR) (probe and primer sequences in Appendix B, D).

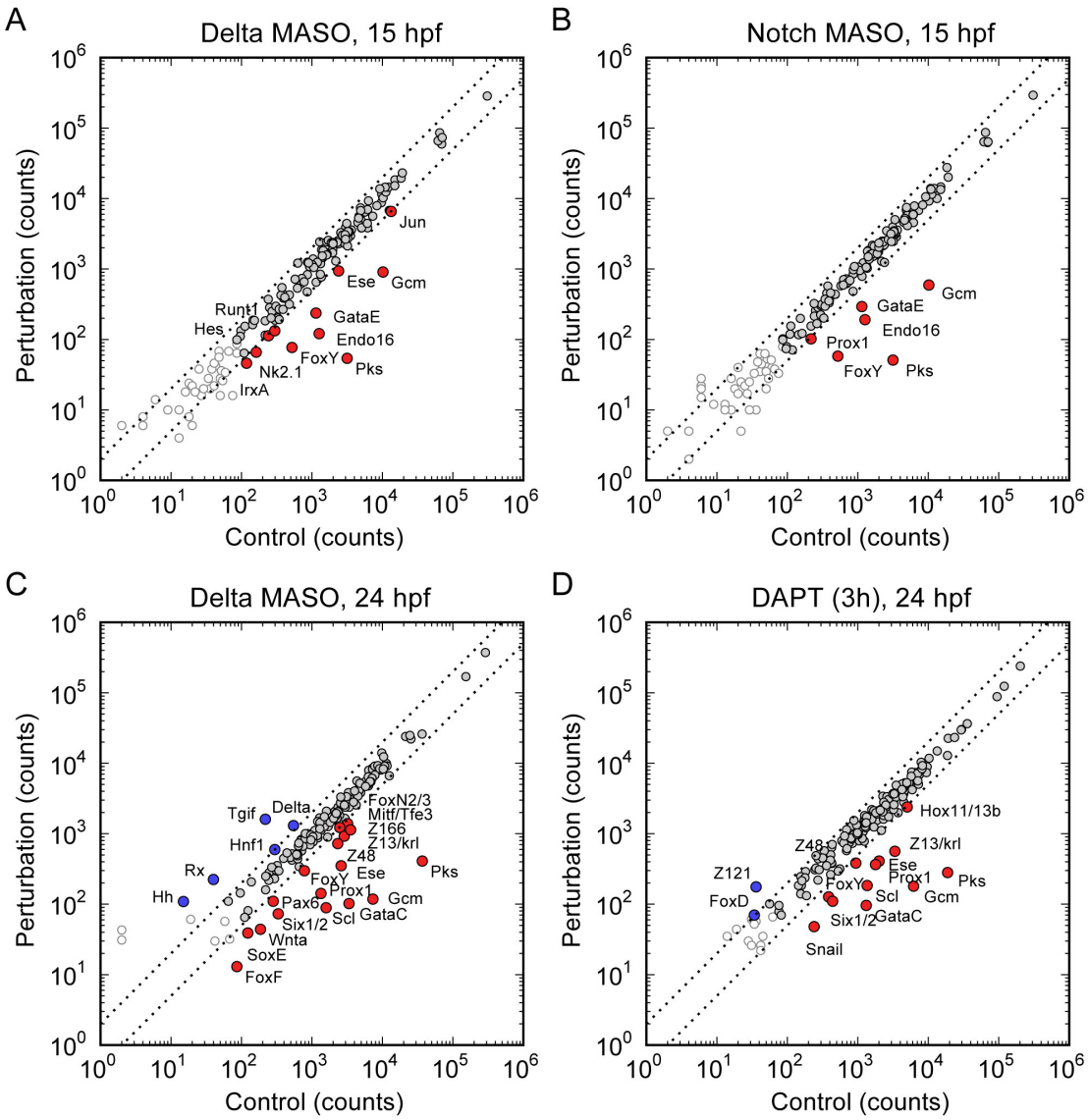


Figure 4.2: Quantitative evaluation of Delta/Notch perturbations. (A, B) RNA from Delta-MASO- or Notch-MASO-injected, and DAPT-treated embryos was extracted and quantified using the Nanostring nCounter. The counts obtained for each gene in the codeset in perturbed embryos are plotted against those of control embryos. Perturbation with Delta MASO or Notch MASO produces almost identical results at 15 hpf, except minor differences that are not substantiated in repeat experiments. Only five genes are reproducibly affected indicating that Delta/Notch signaling has a small number of direct targets. (C, D) Similarly, application of DAPT, a Notch inhibitor, at 3 hpf produces results equivalent to Delta MASO treatment. At 24 hpf essentially all mesodermal genes included in the Nanostring codeset are affected by both perturbations. The dotted lines indicate a threshold of twofold change. Transcription levels were estimated from previous quantification data; genes present with 25 transcripts or less per embryo are marked with an open, grey circle.

Plotting the RNA counts for perturbed embryos against those of control embryos revealed that the prevalence of most genes is not affected by the perturbations (Fig. 4.2). This confirms the phenotypic assessment that, aside from the specific defects caused by the perturbations, development proceeds normally. Differences between types of perturbations are negligible compared to biological variation: Genes that show a more than twofold change are robustly affected regardless of the kind of perturbation applied, thus demonstrating the equivalence of Delta and Notch MASO, as well as expression of DN-Su(H) and DAPT treatment (Fig 2). Nevertheless a few subtle differences do exist and are discussed below.

We find that after D/N signaling has been active for more than six hours (15 hpf) a small number of genes have a strongly reduced transcript level in the perturbed embryos. In our data set, only *gcm*, *gataE*, *foxA*, *endo16* and *FoxY* are significantly affected in their expression level in repeat experiments (Figs. 4.3 A, B; 4.4 A, B; 4.5 A,B). Given that we queried the regulome more or less comprehensively, at least with regard to spatially restricted factors, this strongly suggests that the number of direct targets of D/N signaling is low. *gcm* was previously shown to be a direct target of the N_{ic}/Su(H) complex (Ransick and Davidson, 2006) and is the first gene to become activated by D/N signaling. It is turned on in the veg2 tier of endomesodermal cells that surround the SM when Delta is first activated in these cells (Fig 4.1 F). Following the next cleavage that creates an inner and outer tier of veg2 cells, *gcm* expression is restricted to the inner tier. Only these cells are in touch with the Delta source on which *gcm*

expression is dependent. The spatial expression pattern of *gataE* is similar to *gcm* (Fig. 4.1 G) but it is activated only about three hours after *gcm* (Lee and Davidson, 2004; Materna et al., 2010). The regulatory region of *gataE* contains functional Su(H) sites thus proving that *gataE* is a direct target of D/N signaling (Lee, 2007). However, the delay between *gcm* expression and activation of *gataE* indicates that other inputs are necessary. In fact, Gcm itself is an activator that is required for *gataE* expression to occur in the mesoderm (see Chapter 5).

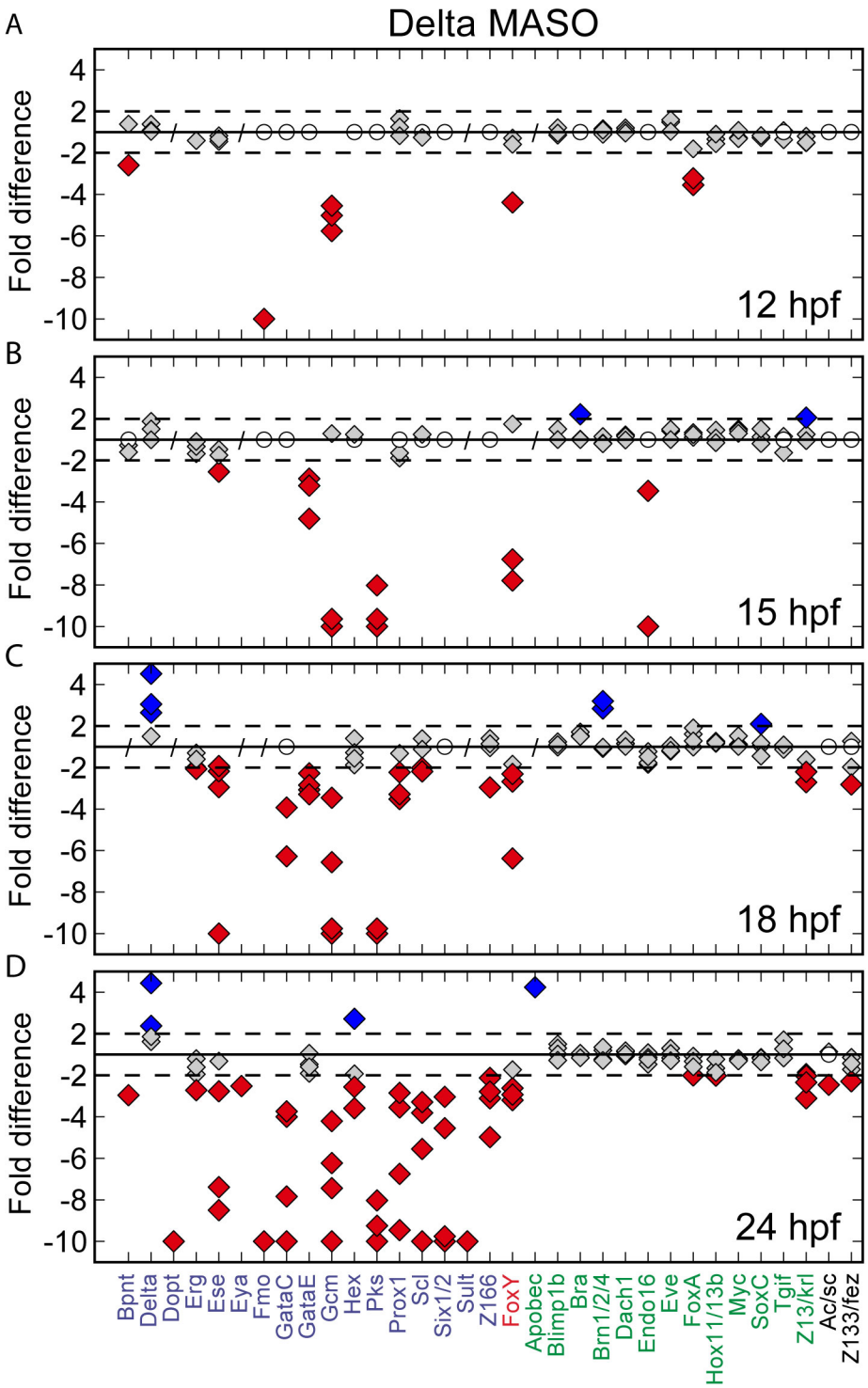


Figure 4.3: Effect of Delta MASO treatment on gene transcription. Fold differences were calculated using the quantitative data obtained with the Nanostring nCounter and supplemented with QPCR data for genes not included in the codeset. Each diamond

represents a single experiment. (A) The earliest gene affected by the perturbation is *gcm*, a known, direct target in the mesoderm. The expression level of *foxY*, a small micromere gene, is also affected at 12 hpf (see also Fig. 4.4A and 4.5A). (B) The transcription factor *gataE* and pigment-cell-differentiation gene *pks*, which are both confirmed direct targets of D/N, have reduced expression levels at 15 hpf. (C) At 18 hpf three (oral) mesodermal genes (*ese*, *gataC*, *prox1*) have reduced expression levels in perturbed embryos, but whether they are direct Delta/Notch targets is unknown. (D) At 24 hpf essentially all mesodermal genes (purple labels) have strongly reduced transcript levels. In contrast, expression of endodermal genes (green labels) is impacted only minimally if at all. The early Delta/Notch input into FoxA (12 h) is only transitory. (C, D) Apical genes (black) are only weakly perturbed by Delta MASO injection. Dashed lines indicate a significance threshold of twofold difference. Genes that are expressed at about 50 molecules or less per embryo are considered insignificant and marked with an open circle. Genes that were not evaluated are marked with a slash (/). For presentation purposes fold differences bigger than tenfold are shown as tenfold.

At 18 hfp, only three hours later, our perturbations reveal several additional mesodermal genes that are dependent on D/N signaling (Figs. 4.3C, 4.4C, 4.5C). These are the transcription factors *prox1*, *gataC*, *ese* and the zinc finger gene *z166*, which are activated at around 16 hpf (Materna et al., 2010). *prox1*, *gataC*, and *ese* are of particular interest as they are specifically expressed in the oral mesoderm (Fig. 4.6) (Poustka et al., 2007; Rizzo et al., 2006). Their strong reduction resulting from D/N perturbation raises the possibility that these genes are also direct targets of D/N signaling. Although similar to *gataE*, expression is first detected significantly later than the initial activation of D/N signaling, thus suggesting the existence of an additional regulatory layer. In fact, for these genes to be expressed, the mesoderm has to first be subdivided into oral and aboral segments. This is a function of Nodal/Tgf-β signaling, which establishes the oral aboral axis of the sea urchin embryo (Duboc et al., 2010; Su et al., 2009) (see Chapter 5).

In unperturbed embryos the NSM occupies the center of the vegetal plate at the mesenchyme blastula stage (24 hpf). However, when D/N signaling has been disrupted, all mesodermal genes are either entirely missing or exhibit a strongly reduced transcript levels (Figs. 4.3D, 4.4D, 4.5, D). This includes the transcription factors *scl* and *six1*/2, its cofactor *eya*, as well as the entire battery of pigment cell differentiation genes (*hpnt*, *dopt*, *fmo*, *papps*, *pks*, *sult*) (Calestani

et al., 2003; Rast et al., 2002). Genes that are expressed both in the NSM and elsewhere in the embryo, specifically lose expression in the NSM, but not in other territories. For example, *shr2* is a transcription factor expressed in both the aboral ectoderm and the oral NSM; while oral NSM expression is lost when D/N function is perturbed, expression in the aboral ectoderm is unchanged (Fig. 4.6 M, R). Similarly, the expression of Delta in the NSM is abolished while its apical expression is unaffected (Fig. 4.6 B, G). Thus, expression of the second, i.e. NSM phase of, *delta* is dependent on D/N signaling from the skeletogenic mesoderm, just like the expression of all other mesodermal genes. However, given the unusual control of *delta* expression through repression by HesC, activation of *delta* in the NSM is certainly not direct. Instead this result suggests that HesC does not clear from the NSM in SM D/N perturbed embryos.

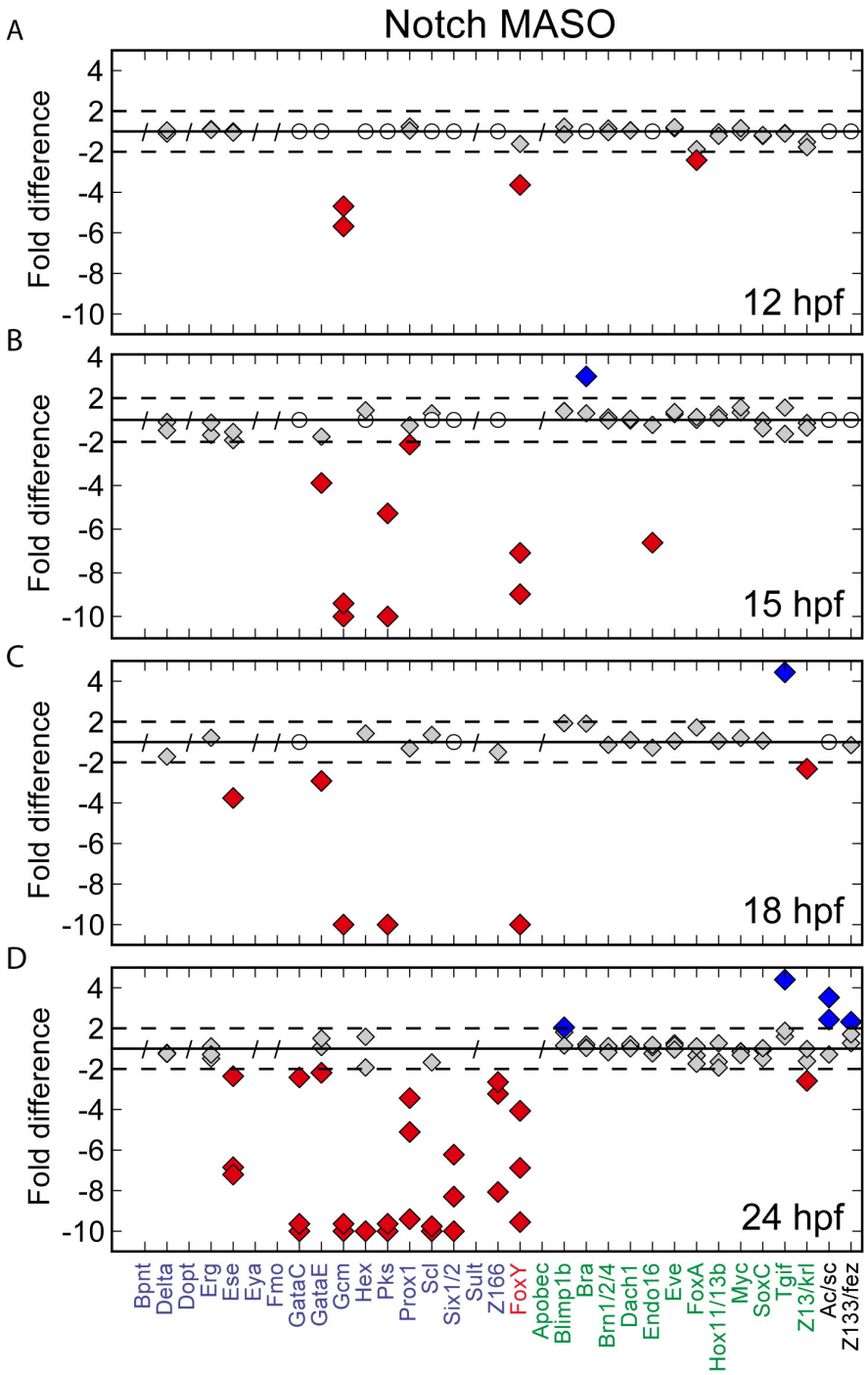


Figure 4.4: Effect of Notch MASO treatment on gene transcription. Data were acquired and analyzed as for Delta MASO (see Fig. 4.3). The results are essentially identical to Delta MASO treatment: Mesodermal genes, and *foxY* in the small micromeres, are strongly affected by Notch MASO injection while endodermal genes are not affected. (C, D) In contrast to Delta MASO injection, which causes an upregulation of *delta* transcripts at 18 hpf

and 24 hpf (Fig. 4.3 C, D), Notch MASO has no effect on *delta* expression. (D) At 24 hpf the apical genes *ac*/*sc* and *fez* exhibit increased abundance. This effect is stronger in Notch-MASO-injected embryos as compared to Delta-MASO-injected embryos, where these transcripts are down-regulated (Fig. 4.3D). Thresholds and symbols are as in Fig. 4.3.

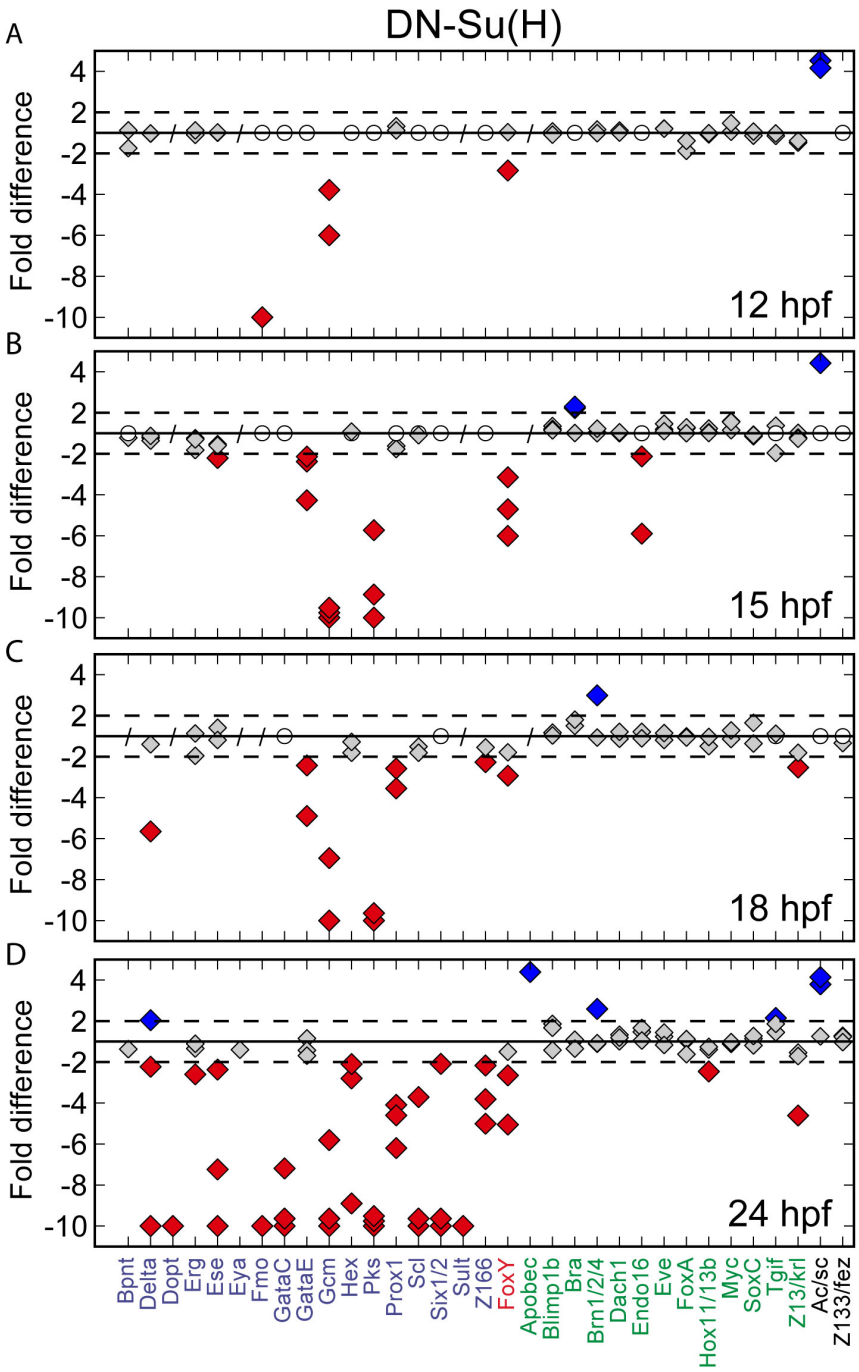


Figure 4.5: Effect of DN-Su(H) expression on gene transcription. Overall, the effects are essentially identical to Delta MASO treatment (Fig. 4.3). DN-Su(H) treatment causes an upregulation of the apical gene *ac/sc* starting at 12 hpf, which is earlier than the effects of Delta MASO and Notch MASO on this gene. Thresholds and symbols are as in Fig. 3 and 4.

Endodermal genes are not targets of NSM Delta.

In contrast to mesodermal genes, endodermal genes are affected only minimally or not at all by D/N perturbations. For example, the expression of *foxA* is reduced early but the effect is only transitory: At 12 hpf *foxA* is expressed in the veg2 tier of cells, directly adjacent to the Delta source, and is a direct target that contains functional Su(H) sites (Ben-Tabou de-Leon and Davidson, 2010). However, *foxA* is also expressed in the outer, or endodermal, tier of veg2 that is not in contact with the SM and cannot receive D/N input. *foxA* expression is predominantly due to activation by TCF, the main driver of the endoderm gene regulatory network (Ben-Tabou de-Leon and Davidson, 2010). After 16 hpf *foxA* and all other endodermal genes will clear from the mesodermal tier as a consequence of D/N signaling (Peter and Davidson, 2011). D/N perturbation interferes with this clearance, probably because β-catenin is not removed from the nuclei. This results in the expression of endodermal genes in cells that would normally become NSM. For example, the rings of *foxA* and *apobec* (an endoderm differentiation gene) expression are substantially smaller in DAPT treated embryos (Fig. 4.6 O, T, P, U). Overall, the absolute prevalence of endodermal transcripts is not significantly altered by D/N perturbation, aside from some minor effects (e.g., on h*ox11/13b*, *z13*, and *apobec*). But this is to be expected given the spatial rearrangements caused by the perturbation. Nevertheless, our data confirm that endoderm genes are not activated by D/N signaling.

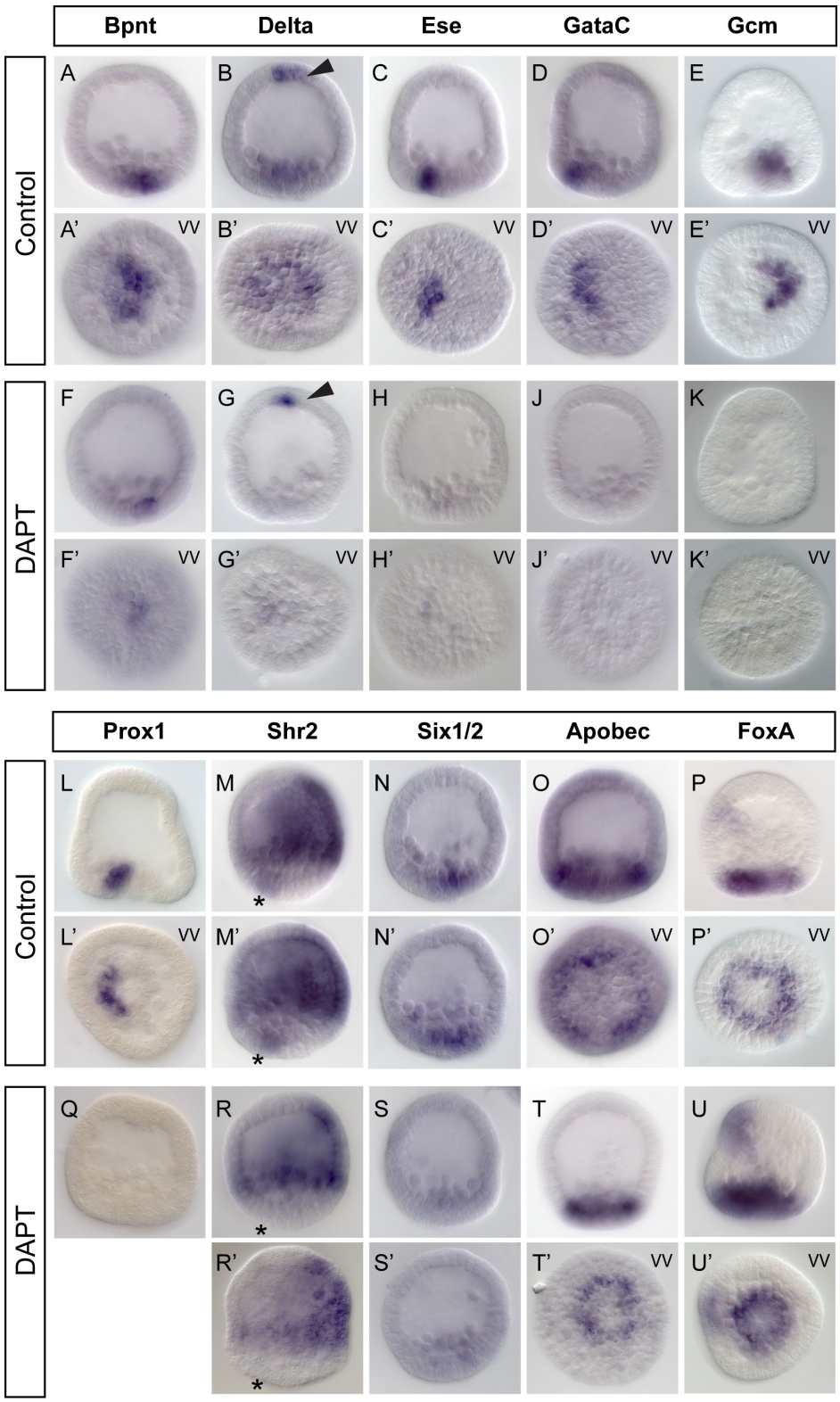

Bpnt
Delta
Ese
GataC
Gcm
Control
DAPT
A
B
C
D
E
A'
B'
C'
D'
E'
VV
F
G
H
J
K
F'
G'
H'
J'
K'
Prox1
Shr2
Six1/2
Apobec
FoxA
L
M
N
O
P
L'
M'
N'
O'
P'
Q
R
S
T
U
R'
S'
T'
U'
*

Figure 4.6. Spatial effects of D/N perturbation with DAPT at 24 hpf. WMISH confirms quantitative results: Following DAPT treatment, mesodermal genes show either severely reduced or no staining by WMISH (*bpnt* — A, F; *ese* — D, H; *gataC* — D, J; *gcm* — E, K; *prox1* — L, Q; *six1/2* — N, S) *.delta* and *shr2* are expressed in additional territories, but transcripts are specifically lost in the mesodermal domain (arrowheads in B, G indicate apical expression of *delta*; asterisks in M, R mark mesodermal expression of *shr2*). The endodermal genes *apobec* and *foxA* do not clear from the cells that would normally be specified as mesoderm (compare O', T' and P', U'), but the signal intensity remains the same. All embryos are oriented with their oral side to the left. In lateral views apical is at the top. VV — vegetal view.

The NSM Delta signal is received in the small micromeres

An unexpected finding in our Nanostring data was the observation that D/N is an activating input of *foxY*, which was significantly affected even at the earliest sampling time (12 hpf, Figs. 4.3A, 4.4A, 4.5A). This is surprising because *foxY* is expressed only in the small micromeres and not the NSM (Fig. 4.1H). The small micromeres are the product of the unequal fifth cleavage of the micromeres and are generally thought of as "set aside" in early development for later use in constructing the adult rudiment, which occurs in the coelomic pouches. They are unique in the early sea urchin embryo because they are generally quiescent and divide only once more before contributing to the coelomic pouches. The small micromeres express a battery of genes with conserved stem cell and germ line functions (e.g. *vasa*, *nanos*, *piwi*) (Juliano et al., 2010; Voronina et al., 2008), which suggests that they maintain a relatively undifferentiated state during embryogenesis. After their birth the small micromeres remain located on top of their sister cells (the skeletogenic mesoderm) in the middle of the vegetal plate. Once *delta* is activated in the SM, the small micromeres are encircled by cells that present the Delta ligand on their surface. Although the small micromeres are unique in many ways, the activation of *foxY* by D/N signaling reveals that they are not immune to signaling interactions.

Function of NSM Delta signaling.

When perturbing the early (SM) Delta expression phase, as with Delta MASO injection, Delta expression in the mesoderm is also abolished, thus making it difficult to delineate the effects of losing these two phases of D/N signaling. In other words, it is impossible to pick out the direct targets of late (NSM) Delta without specifically disrupting only its function. Furthermore, *delta*-expressing NSM cells are now in touch with the endoderm and could, in principle, receive the D/N signal as was previously proposed (Sherwood and McClay, 2001). To ask where the NSM Delta signal is received and what genes are activated by it, we added

DAPT at 17 hpf, a time just prior to the handoff of Delta expression from SM to NSM. Addition of DAPT at this time should not interfere with the function of SM Delta, which has basically run its course.

Evaluation of perturbation experiments with the Nanostring nCounter on embryos collected at 24 and 30 hpf revealed that out of all 205 genes tested, the only gene activated by mesodermal Delta is the small micromere gene *foxY* (Figs. 4.7C, D). No other gene, neither NSM nor endoderm, is affected by this perturbation. When the SM cells ingress, the small micromeres stay behind and come to lie on top of the NSM cells that occupy the spot previously taken up by the SM (Fig. 4.1H). This coincides with the start of delta expression in the NSM. As a result Delta ligand is continuously presented to the small micromeres where it is a required, activating input of foxY into gastrulation. Phenotypic evaluation of embryos in which only the NSM Delta function has been perturbed confirms these molecular findings. Late treatment with DAPT does not affect pigment cell formation (Figs. 8D – F); since no oral NSM genes are impacted, it can be assumed that blastocoelar cells are also specified normally (they are specified concurrently with pigment cells). The one clear defect we observe is the failure to develop coelomic pouches, the site where small micromeres are normally found (Figs. 4.8 D'–F').

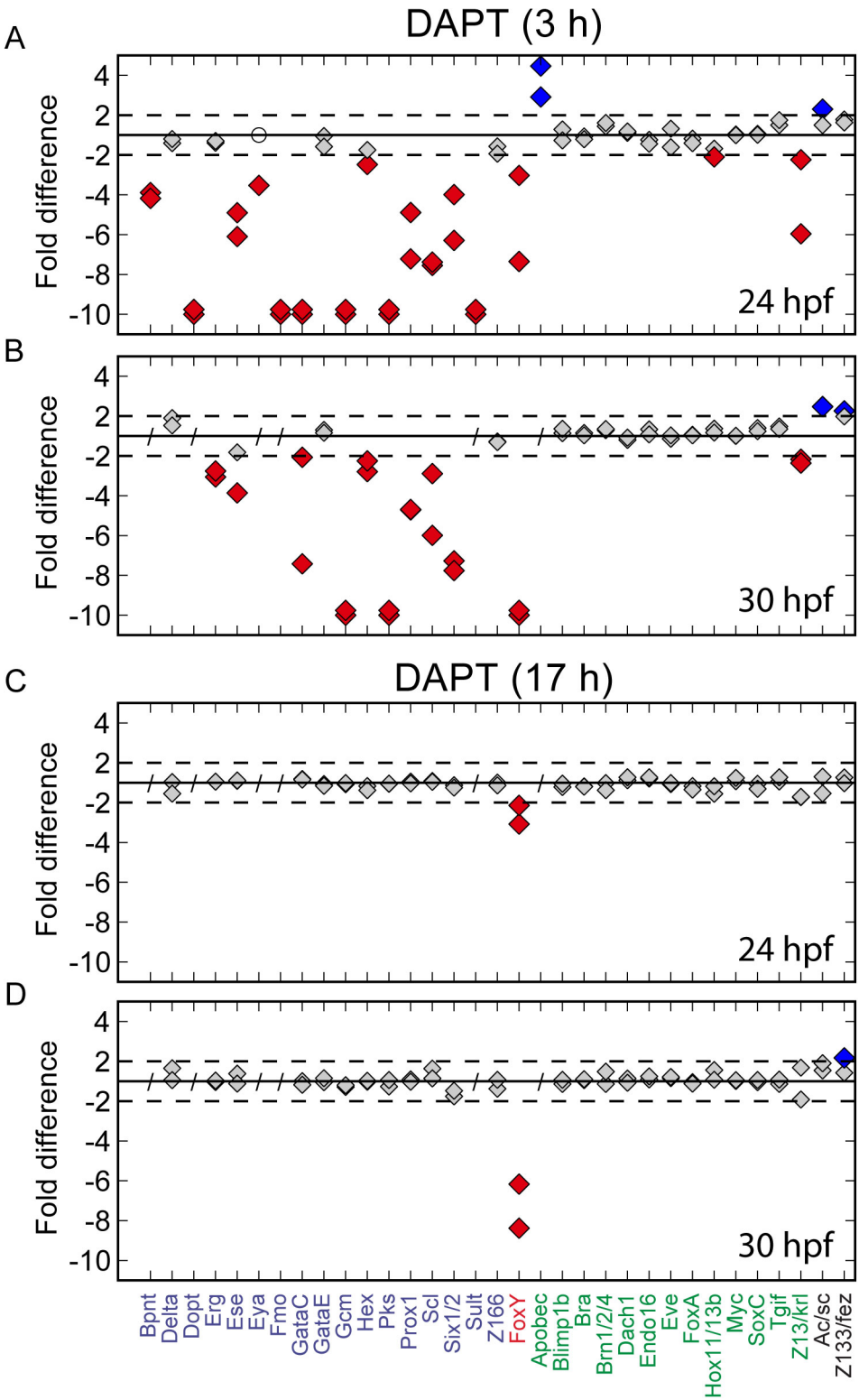

A
DAPT (3 h)
Fold difference
4
2
-2
-4
-6
-8
-10
24 hpf
B
Fold difference
30 hpf
C
DAPT (17 h)
Fold difference
24 hpf
D
Fold difference
30 hpf
Bpnt
Delta
Dopt
Erg
Ese
Eya
Fmo
GataC
GataE
Gcm
Hex
Pks
Prox1
Scl
Six1/2
Sult
Z166
FoxY
Apobec
Blimp1b
Bra
Brn1/2/4
Dach1
Endo16
Eve
FoxA
Hox11/13b
Myc
SoxC
Tgif
Z13/krl
Ac/sc
Z133/fez

Figure 4.7: Effect of perturbation of mesodermal Delta function on gene transcription. (A, B) Early addition of DAPT (at 3 hpf) prevents the activation of mesodermal genes while endodermal genes are not affected. (C, D) Addition of DAPT at 17 hpf perturbs the function of late, i.e. mesodermal, Delta. The only gene affected by this perturbation is *foxY* indicating that it requires a continuing activating input from Delta/Notch signaling for its expression. No mesodermal or endodermal genes are affected by loss of mesodermal Delta.

Differences between perturbation methods

As is immediately evident from the perturbation data presented in Figs. 4.2 – 4.5, the genes strongly affected by perturbation of D/N signaling are very similar regardless of the perturbation method used. However, a few subtle differences do exist and are worth noting. First, expression of delta itself is only affected in Delta-MASO injected embryos. Delta reaches a small plateau in its expression between 12 and 16 hpf, after which its abundance declines to about a third of its peak level by 24 hpf. In Delta-MASO-injected embryos the expression level of *delta* is significantly higher at 18 and 24 hpf (Fig. 3 C, D) while Notch MASO and DAPT treatment have no effect. Instead Delta protein itself must contribute to the regulation of *delta* transcript, presumably in SM cells, as these are the cells from which *delta* is cleared at this time.

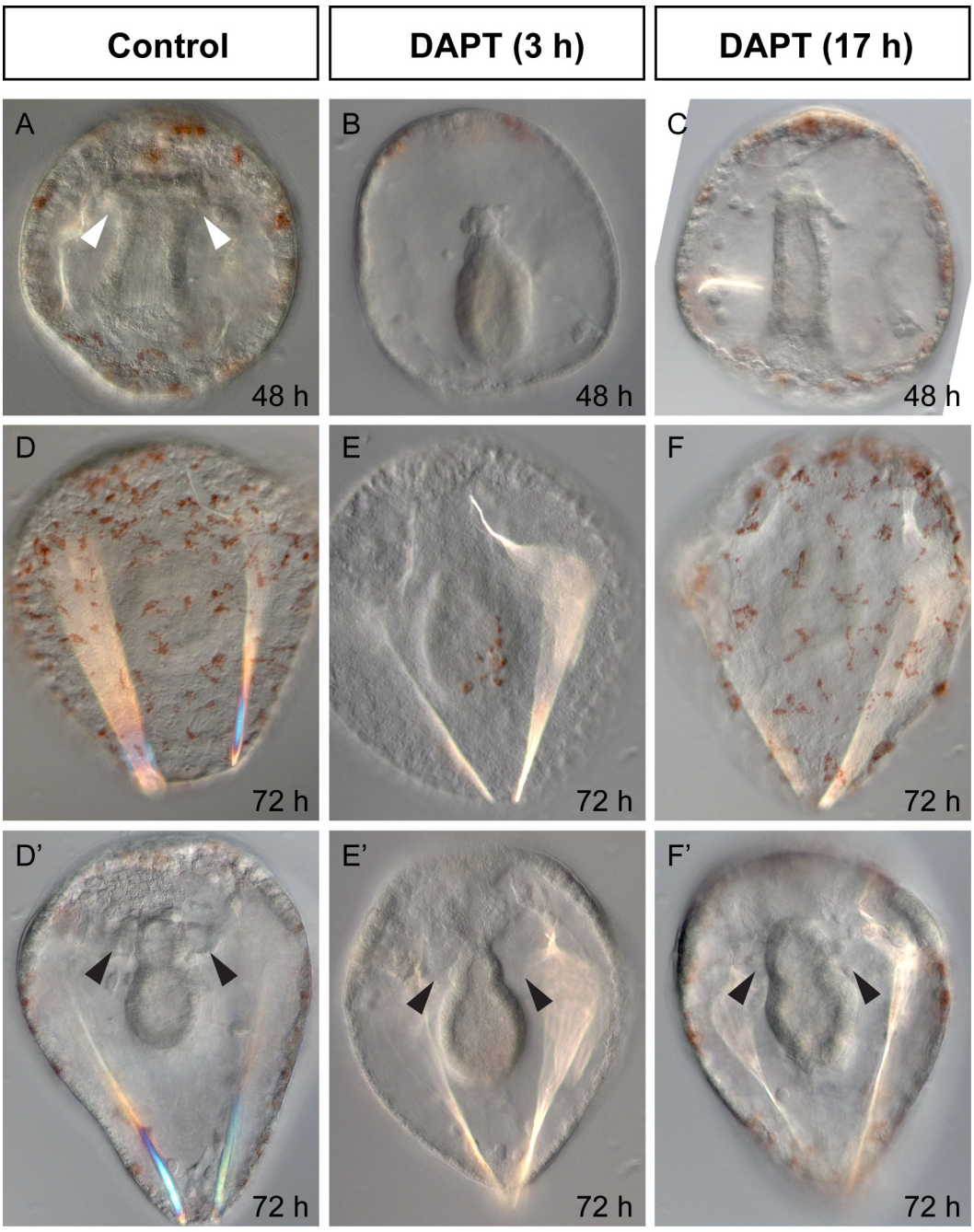


Figure 4.8: Phenotype of embryos with perturbed skeletogenic or mesodermal Delta function. (A, D) Control embryos form pigment cells and start developing coelomic pouches at about 48 hpf (white arrowheads A, black arrowheads in D'). (B, E) Perturbation of skeletogenic Delta (DAPT added at 3 hpf) causes loss of all mesoderm and produces embryos with significantly fewer, if any, pigment cells. At the DAPT concentration used here about half of the embryos exogastrulate, but those that proceed normally through gastrulation do not form coelomic pouches (black arrowheads in B, E'). (C, F) In contrast,

addition of DAPT at 17 hpf does not interfere with specification of pigment cells. However, these embryos do not form coelomic pouches (black arrowheads in C, F').

A second difference can be observed in the expression levels of *ac/sc*. Although we do not know its spatial expression, it is a proneural gene in other organisms and a direct target of D/N signaling (Younossi-Hartenstein et al., 1996). This may hint at apical expression in the sea urchin. *ac/sc* and a second gene known to be expressed apically (*z133/fez*), are upregulated in Notch-MASO-, DN-Su(H)-, and DAPT-treated embryos at 24 hpf (Figs. 4.4, 4.5D, 4.7A, B) but expression of DN-Su(H) has a much stronger effect than Notch MASO or DAPT treatment. In contrast, Delta MASO injection does not result in increased expression levels (Fig. 34.D). The spatial expression of *ac/sc* and *z133/fez* in relation to Delta are currently unknown, just as is their connection to other regulatory genes in the apical domain, thus preventing us from drawing any conclusion about causal relationships. However, *z133/fez* is required to determine the size of the apical plate by antagonizing the effects of Bmp signaling, indicating that its relationship to Delta signaling warrants further exploration (Yaguchi et al., 2011).

Discussion

Micromere-derived Delta is required for specification of all mesoderm derivatives. In this study we comprehensively analyzed the effects of D/N signaling on expression of transcriptional regulators active in early sea urchin development. Our large-scale analysis has high temporal resolution to unveil a complete picture as to the immediate effects of signal reception in the vegetal half of the embryo.

Possibly, the most surprising finding is that only very few genes are direct targets of D/N signaling. Effective Delta signaling is limited to the immediate neighbors of the Delta source.

For the vegetal half of the sea urchin embryo this means that early on only endomesoderm, and after 7th cleavage the NSM precursors, are in the position to receive the signal, as they surround the SM as a ring of cells. This also applies to the small micromeres that are the sister cells of the SM and that remain located on top of them after their birth. By the time the SM expression of *delta* terminates only *gcm*, *gataE*, *pks*, *prox1*, *ese*, and potentially *gataC* have been activated in the NSM, and *foxY* in the small micromeres (Fig. 4.9). Of these, *gcm*, *gataE* and *foxY* are confirmed targets (Lee, 2007; Ransick and Davidson, 2006; J. Smith, personal communication). *prox1*, *ese*, and *gataC* are potentially directly activated by D/N as well. However, since their expression depends on the specification of oral NSM, the onset of their expression is delayed by several hours (Duboc et al., 2010; Materna et al., 2010; Chapter 5). Expression of pigment-cell-specification genes, such as *pks* that is expressed early in the NSM, is a function of the activity of its upstream regulators, namely Gcm and GataE (Calestani et al., 2003). A number of genes are turned on specifically in the NSM after expression of SM *delta* has stopped. These genes are affected by Delta perturbation. However, since they do not overlap in time with active Delta signaling in the SM, these genes must be downstream of an intermediate.

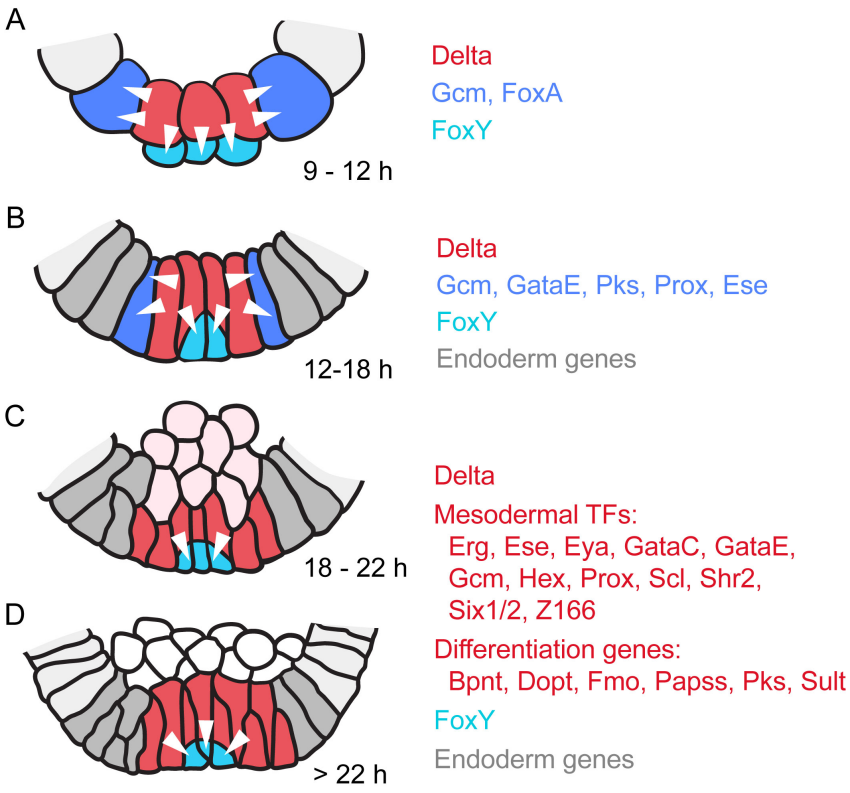


Figure 4.8: Schematic representation of effective Delta signaling in the pre-gastrula sea urchin. (A) The Delta ligand is present in the cells of the skeletogenic lineage starting at 9 hpf. It is received in the neighboring cells and turns on *gcm* in the endomesoderm. (B) After the veg2 tier of cells divides into an inner and outer tier, the D/N signal is only received in the inner tier adjacent to the skeletogenic cells, i.e. the mesodermal precursors. Here, D/N signaling activates *gataE* , and thereafter the transcription factors *prox1*, *ese* and *gataC* in the oral mesoderm. (C) Concurrent with ingression of skeletogenic cells, *delta* ceases to be expressed in the ingressing cells and instead turns on in the mesoderm. The genes expressed in the mesoderm now run autonomously and no longer require the D/N signal as an activating input. (D) As skeletogenic cells ingress, the small micromeres remain in the same position at the center of the vegetal plate and come into contact with the mesoderm. Mesodermal Delta is a continuing activating input into *foxY*, which is required to maintain its expression in the small micromeres.

Our study is limited to the genes in our gene set and we cannot exclude the possibility that we missed additional regulatory genes that are direct targets of Delta signaling before gastrulation. We consider this unlikely, but even if an additional gene were found, this does not change the overall conclusion that the number of targets is very small. Similarly, quantification of

perturbation effects may miss effects on genes with more complex expression patterns. Loss of NSM expression of genes like *shr2* can only be detected by spatial evaluation.

In contrast to the NSM, the endoderm is physically separated from the Delta source after 7th cleavage (Peter and Davidson, 2010). Although in principle endoderm genes could be activated by D/N, as the effect on *foxA* indicates, the spatial separation prevents the Delta signaling from reaching the endoderm. This is reflected in our perturbation data, which show a strong decrease in expression levels of NSM genes, while indicating that endoderm genes are not affected.

As *delta* is activated in the NSM, endoderm cells come in direct contact with the Delta source. Yet endoderm genes still show no change in expression level. Thus it is likely that the D/N signal is not properly received and/or processed by the endoderm. Endoderm cells express *notch* and several other genes that are essential for signal reception to occur. This includes *fringe*, at least up to gastrulation, and *numb*. Both genes have been shown to promote Notch signaling (Peterson and McClay, 2005; Range et al., 2008).

Although endoderm transcript levels are not, or only minimally, affected by D/N perturbation, their spatial expression is altered as they are expressed in cells that would otherwise become specified as NSM. This has been attributed to the clearance of β-catenin, the main driver of endoderm genes, in NSM cells as a consequence of D/N signaling (Peter and Davidson, 2011). In fact, it had earlier been noted that overexpression of an N_{act} construct expands the NSM and moves the endoderm towards the apical pole; this is accompanied by a shift in nuclearized β-catenin in support of this explanation (Sherwood and McClay, 2001).

The collapse of the endodermal ring into the NSM in D/N-perturbed embryos also affects the expression of delta in the NSM. NSM *delta* is clearly not a direct target of SM Delta signaling, due to the control of its expression by HesC (Revilla-i-Domingo et al., 2007; Smith and Davidson, 2008). However, endodermal cells stay *hesC*-positive and thus prevent delta expression. Since the NSM cells acquire the regulatory state of endoderm following D/N perturbation, they do not express *delta*. Thus, as earlier studies have described, the perturbation of SM Delta also perturbs NSM Delta and causes loss of all mesoderm derivatives.

Disruption of late (NSM) Delta by late addition of DAPT reveals *foxY*, in the small micromeres, to be a target. Like endodermal genes, no gene in the NSM is affected. Phenotypically these embryos exhibit defects in coelomic pouch formation, as has been reported for experiments with chimeric embryos (Sweet et al., 2002). It will be interesting to address the function of FoxY in coelomic pouch formation.

Our results raise the question as to why D/N targets turn on only in specific cell types. E.g., *foxY* expression is limited to small micromeres but not in NSM; vice versa gcm is activated only in NSM and not in small micromeres. Nothing is known about the effects of D/N signaling in the apical domain. The explanation will surely lie in the different regulatory states in these territories. Connecting these genes to additional activators (or repressors) will provide a causal explanation of our observations.

Our findings raise a point regarding the evolution of vegetal development in echinoderms. It has previously been reported that in starfish, endodermal genes are activated by D/N signaling from the mesoderm (Hinman and Davidson, 2007). However, here we clearly demonstrate that the sea urchin endoderm does not receive any activating input from the mesoderm. It is likely that development of sea urchin mesoderm, with its additional unequal cleavage, is derived, and that they may have lost the D/N input into endoderm genes. However, both animals have in common the formation of coelomic pouches in which the adult rudiment will form. It will be interesting to see if D/N signaling in starfish mesoderm has similar effects on formation of coelomic pouches. Mesodermal expression of *delta* may thus be a feature that is related to evolution of indirect development, an ancestral trait of echinoderms.

References

Ben-Tabou de-Leon, S., Davidson, E.H., 2010. Information processing at the foxa node of the sea urchin endomesoderm specification network, Proc Natl Acad Sci U S A 107, 10103–10108.

Calestani, C., Rast, J.P., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. Development 130, 4587–4596.

Croce, J.C., McClay, D.R., 2010. Dynamics of Delta/Notch signaling on endomesoderm segregation in the sea urchin embryo, Development 137, 83–91.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246, 162–190.

Duboc, V., Lapraz, F., Saudemont, A., Bessodes, N., Mekpoh, F., Haillot, E., Quirin, M., Lepage, T., 2010. Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. Development 137, 223-235.

Geiss, G., Bumgarner, R., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D., Fell, H., Ferree, S., George, R., Grogan, T., James, J., Maysuria, M., Mitton, J., Oliveri, P., Osborn, J., Peng, T., Ratcliffe, A., Webster, P., Davidson, E., Hood, L., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs, Nat Biotechnol 26, 317–325.

Hinman, V.F., Davidson, E.H., 2007. Evolutionary plasticity of developmental gene regulatory network architecture, Proc Natl Acad Sci USA 104, 19404–19409.

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

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

Howard-Ashby, M., Materna, S.C., Brown, C.T., Tu, Q., Oliveri, P., Cameron, R.A., Davidson, E.H., 2006c. High regulatory gene use in sea urchin embryogenesis: Implications for bilaterian development and evolution. Dev Biol 300, 27–34.

Hughes, J.N., Dodge, N., Rathjen, P.D., Rathjen, J., 2009. A novel role for gamma-secretase in the formation of primitive streak-like intermediates from ES cells in culture. Stem Cells 27, 2941–2951.

Juliano, C.E., Yajima, M., Wessel, G.M., 2010. Nanos functions to maintain the fate of the small micromere lineage in the sea urchin embryo. Dev Biol 337, 220–232.

Lee, P.Y., 2007. Function and regulation of the Strongylocentrotus GataE gene. PhD thesis Caltech.

Lee, P.Y., Davidson, E.H., 2004. Expression of Spgatae, the *Strongylocentrotus purpuratus* ortholog of vertebrate GATA4/5/6 factors. Gene Expr Patterns 5, 161–165.

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

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

Peter, I.S., Davidson, E.H., 2010. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. Dev Biol 340, 188–199.

Peter, I.S., Davidson, E.H., 2011. A gene regulatory network controlling the embryonic specification of endoderm. Nature 474, 635–639.

Peterson, R.E., McClay, D.R., 2005. A Fringe-modified Notch signal affects specification of mesoderm and endoderm in the sea urchin embryo. Dev Biol 282, 126–137.

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

Range, R.C., Glenn, T.D., Miranda, E., McClay, D.R., 2008. LvNumb works synergistically with Notch signaling to specify non-skeletal mesoderm cells in the sea urchin embryo. Development, 135, 2445–2454.

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

Rast, J.P., Cameron, R.A., Poustka, A.J., Davidson, E.H., 2002. brachyury Target genes in the early sea urchin embryo isolated by differential macroarray screening. Dev Biol 246, 191–208.

Revilla-i-Domingo, R., Oliveri, P., Davidson, E.H., 2007. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. Proc Natl Acad Sci USA 104, 12383–12388.

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

Robertson, A.J., Coluccio, A., Knowlton, P., Dickey-Sims, C., Coffman, J.A., 2008. Runx expression is mitogenic and mutually linked to Wnt activity in blastula-stage sea urchin embryos. PLoS One 3, e3770.

Ruffins, S.W., Ettensohn, C.A., 1996. A fate map of the vegetal plate of the sea urchin (*Lytechinus variegatus*) mesenchyme blastula. Development 122, 253–263.

Sherwood, D.R., McClay, D.R., 1999. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. Development 126, 1703–1713.

Sherwood, D.R., McClay, D.R., 2001. LvNotch signaling plays a dual role in regulating the position of the ectoderm-endoderm boundary in the sea urchin embryo. Development 128, 2221–2232.

Smith, J., Davidson, E.H., 2008. From the Cover: Gene Networks In Development and Evolution Special Feature Sackler Colloquium: Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo. Proc Natl Acad Sci USA 105, 20089–20094.

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

Sweet, H.C., Gehring, M., Ettensohn, C.A., 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. Development 129, 1945–1955.

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

Voronina, E., Lopez, M., Juliano, C.E., Gustafson, E., Song, J.L., Extavour, C., George, S., Oliveri, P., McClay, D., Wessel, G., 2008. Vasa protein expression is restricted to the small micromeres of the sea urchin, but is inducible in other lineages early in development. Dev Biol 314, 276–286.

Wang, M.M., 2011. Notch signaling and Notch signaling modifiers. Int. J. Biochem. Cell. Biol. 43, 1550-1562.

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

Younossi-Hartenstein, A., Nassif, C., Green, P., Hartenstein, V., 1996. Early neurogenesis of the *Drosophila* brain. J Comp Neurol 370, 313-329.

Chapter 5

THE REGULATORY ORIGIN OF ORAL AND ABORAL MESODERM IN SEA URCHIN EMBRYOS.

Stefan C. Materna, Enhu Li, Andrew Ransick, Jongmin Nam, Eric H. Davidson

Abstract

The non-skeletogenic mesoderm (NSM) of the sea urchin embryo is subdivided into oral and aboral segments. The oral NSM gives rise to blastocoelar cells, whereas the aboral side gives rise to pigment cells. We characterize the regulatory state of both kinds of NSM and make numerous additions to the list of genes specifically expressed in each segment. This analysis shows that after polarization is complete, the regulatory genes active orally are non-overlapping with those running aborally. The initiation of this division has been linked to Nodal signaling that establishes the oral/aboral axis in the sea urchin embryo. Our experiments show that Nodal signaling does not directly activate oral NSM genes. Instead Nodal activates the homeobox gene *not* that represses aboral cell fate in oral NSM cells. Perturbation analysis reveals that in a similar fashion aboral genes prevent expression of oral genes aborally. Oral NSM genes have the capacity to be expressed in the entire NSM when specification of aboral NSM is prevented, indicating that the activator of oral NSM genes is present throughout the NSM. We expand the gene regulatory network underlying specification of pigment cells and describe the linkages that enable lockdown of its regulatory state. We show that the mutual exclusion of oral and aboral NSM is a feature of the regulatory network underlying mesoderm formation in sea urchins.

Introduction

The vegetal half of the blastula-stage sea urchin embryo will give rise to endodermal and mesodermal cell types. The skeletogenic lineage that will give rise to skeletogenic mesoderm (SM) occupies the central part of the vegetal plate. It is surrounded by the veg2 ring of cells that are precursors of endoderm and mesoderm. During eighth cleavage this tier is subdivided radially into two rings, the outer of which will become the endoderm and form the gut. The inner ring will develop into the non-skeletogenic mesoderm (NSM). NSM will eventually give rise to four cell types (pigment cells, blastocoelar cells, coelomic pouch cells, and esophageal muscle). Specification of blastocoelar cells and pigment cells is dependent on Delta signaling from the SM (Sherwood and McClay, 1999; Sweet et al., 2002). The other cell types are dependent on a later phase of Delta signaling and are not specified until later in gastrulation (Sweet et al., 2002).

Initially, the NSM is uniform and the first gene to be activated (*gcm*) is expressed in the entire ring (Ransick et al., 2002). However, at the late blastula stage the NSM is divided into an oral and aboral segment. The oral NSM will specify blastocoelar cells that are thought to have immune function. They turn off *gcm* and instead express a number of transcription factors related to hematopoiesis, such as *gataC* and *scl* (Davidson et al., 2002; Duboc et al., 2010). Pigment cells derive from the aboral segment of the NSM; they may also have immune function and bear a certain resemblance to macrophages (J. Rast, personal communication). Although they usually intercalate in the blastocoelar wall, they can migrate through the blastocoel and are attracted to sites of injury and infection.

The subdivision of the NSM is dependent on establishment of the oral/aboral (O/A) axis in the sea urchin embryo. O/A axis formation is a function of Nodal/TGF-β and Bmp signaling (Duboc et al., 2004; Su et al., 2009). These signals have their origin in the oral ectoderm. It is thought that activation of target genes by Nodal is limited in its range by its antagonist Lefty, itself a downstream target of Nodal (Duboc et al., 2008). Wherever the Nodal signal is received, cells adopt oral fate. This applies to the entire oral side of the embryo, from the apical plate down to the SM. It has previously been shown that reception of the Nodal signal

in the vegetal half of the embryo is necessary for specification of oral NSM (Duboc et al., 2010). This has been interpreted as evidence that Nodal directly activates oral NSM genes. However, the huge lag in time between Nodal activation and the onset of expression of oral mesoderm genes makes this proposition at best doubtful.

Here we explore the regulatory origin of oral and aboral mesoderm fate. We show that Nodal is not a direct input into oral NSM genes and instead operates through the homeobox gene *not*. However, Not appears to function as a repressor in the NSM. We explored the relationship of oral and aboral NSM and in the process added several new genes to the gene regulatory network (GRN) underlying pigment cell formation. Interestingly, knockdown of aboral NSM genes causes an expansion of oral NSM, providing a second line of evidence that Nodal is not an input into oral NSM genes, as its activity is limited to the oral side. Instead, our results imply a driver that is active in the entire NSM, a requirement that is met by *delta* expression in the SM. This signal is received in the entire NSM and necessary for its specification. We characterize the regulatory state of oral and aboral NSM and find that they have no commonalities, and that instead all NSM genes are either oral or aboral. The temporary nature of the activating inputs into NSM genes indicates that the transcription factors expressed function to maintain their expression by locking down their regulatory state. We identify the linkages that provide a causal explanation for how this occurs in the aboral NSM.

Materials and Methods

MASO injection and inhibitor treatment

Morpholino antisense oligonucleotides (MASOs) were obtained from Gene-tools LLC and inject at 300 μM (100 μM) for Nodal, in 0.12 M KCl. Injection volumes were ca. 5 pl. Sequences are as follows:

- Delta-MASO: CAAGAAGGCAGTGCGGCCGATCCGT
- Ese-MASO: TTCCCTTCATGGCTGTAAAAACGAA,
- GataC-MASO: CATTAAAAGAAAATAACAAGTTCAC,
- GataE-MASO1: ACCACGCTTTGCTTCGTGTTTGGCC (translation block),
- GataE-MASO2: TCTCGTCTTGAGCCAGACTGCAATC (splice block),
- Nodal-MASO: TGCATGGTTAAAAGTCCTTAAAAAT,
- Not-MASO: GACATCAAGTTGGAACTCATCATAG ,
- Prox1-MASO1: TGCATCCTCGACCTTAGACATTGGC (translation block),
- Prox1-MASO2: ACACCAAAAAGGACTTACCGTGAAC (splice block).

Our GataE MASOs replace an earlier MASO (GACTTACACCGACCTGATGTGGCAT) that we find to have deleterious effects when injected at high concentrations. The γ-secretase inhibitor DAPT (Hughes et al., 2009) was dissolved in DMSO and added at 3 hpf to a final concentration of 8 μM (Materna and Davidson, 2012). Higher concentration cause all embryos to exogastrulate and a concentration of higher than 20 μM causes severe, and unspecific, defects. Lower inhibitor concentration results in higher numbers of pigment cells.

Embryo culture and RNA extraction

Sea urchin embryos were cultured at 15°C and closely monitored for proper development. For lysis, sea water was removed before adding 350 μl RLT buffer from the Qiagen RNeasy Micro Kit (Qiagen, Hilden, Germany). Embryo lysates were immediately stored at -70°C until use. RNA was extracted according to manufacturer's instructions; to maximize recovery, RNA was eluted with 50 μl nuclease-free water. Samples were ethanol precipitated and resuspended in 11 μl nuclease-free water. Samples were split and 5 μl were used in Nanostring nCounter assays. The leftovers were reverse transcribed.

RACE PCR and amplification of cDNA sequences

5' ends of transcripts were identified using the Clontech SMART RACE PCR kit. PCR fragments were cloned and sequenced to identify the approximate transcription start site. Full length transcripts were amplified from cDNA. The actual 5' end of the *six1/2* sequence differs from the gene prediction SPU_17397 (genbank accession numbers: *prox1*—JQ956375 , *six1/2*—JQ264781, *z166*—JQ945922).

Transcriptional profiling

Expression level was quantified with the Nanostring nCounter using customer-designed probe sets for 183 genes. The samples were processed according to manufacturers instructions and as described previously (Materna et al., 2010). Genes not included in the codeset were analyzed by QPCR profiling (Materna and Oliveri, 2008). All together 206 genes were included in this study. Gene names, accession numbers, probe and primer sequences are given in Appendix B and D.

Whole-mount in-situ hybridization

Whole-mount in-situ hybridization was conducted as described previously (Chapter 4). For two color in situs, probes labeled with digoxigenin and fluorescein were used. To deactivate alkaline phosphatase attached to the first antibody embryos were incubated in acidic glycine stop solution followed by an additional blocking step and incubation with the second antibody. First staining was performed with Fast-Red, second staining with NBT/BCIP (Chapter 4).

A collection of 129 barcoded reporter constructs was injected together with Delta MASO or Gcm MASO. RNA was extracted from embryos and quantified using a Nanostring codeset for detection of barcodes. Data were analyzed as in Nam and Davidson (Nam et al., 2010).

Results

Regulatory states of the NSM

Initially, the veg2 NSM precursors form a ring of single cells around the SM at the center of the vegetal plate. They receive the Delta signal from the SM, which causes the expression of *gcm* and *gataE* in the full ring of NSM precursors (Lee, 2007; Ransick and Davidson, 2006) (Fig. 5.2A; Fig. 5.4A; Fig. 5.7A, B). Both genes are confirmed direct targets of Delta/Notch signaling. Together, they immediately turn on pigment cell differentiation genes such as *pks*, *fmo*, and *sult*; these are also expressed throughout the entire ring of NSM precursors (Calestani et al., 2003). Initially, endodermal genes are expressed in the NSM precursors, but by 18 hpf their expression in NSM cells ceases as a consequence of D/N signaling (Peter and Davidson, 2011). Gene expression of endodermal and NSM genes is no longer overlapping and the endoderm encircles the entire NSM (Fig. 5.1A, B).

Prior to ingression of SM, the NSM develops asymmetries in gene expression. The expression of *gcm*, *gataE*, and downstream pigment cell (*pks*, *fmo*, *dopt*, *sult*) ceases on the oral side and becomes confined to the aboral side (Fig. 5.1, Fig. 5.4, Fig. 5.7) (Calestani et al., 2003; Ransick et al., 2002). At this time three additional regulatory genes (first *z166*, then *six1/2* and its cofactor *eya*) are activated specifically on the aboral side (Fig. 5.1D, E; Fig 5.2A) (Howard-

Ashby et al., 2006; Materna et al., 2006). Concurrent with the disappearance of *gcm* and *gataE*, several genes are specifically activated in close succession in the oral NSM, first *prox1*, then *gataC*, and *ese* (Fig. 5.1F, G) (Poustka et al., 2007). Expression of *ese* starts at 10 hpf and is initially ubiquitous, but is specific to the oral NSM by 18 hpf (Rizzo et al., 2006). The next oral NSM genes to turn on are *scl* (Fig. 5.1H, 5.2 B), and *shr2*. *shr2* is expressed in the aboral ectoderm in addition to the NSM (Fig. 5.1J). Once ingression of SM is complete, several genes first expressed in SM will also turn on in the oral NSM, including *erg*, *hex*, and *ets1/2* (Fig. 5.1K – L). The only gene that is expressed across the entire NSM is *delta* (Fig. 5.1 C; see Chapter 4). This is due to the fact that it is expressed everywhere the repressor HesC is absent, as is the case in the NSM (Revilla-i-Domingo et al., 2007; Smith and Davidson, 2008). All 13 transcription factors examined here (plus the six pigment cell differentiation genes) are confined to either oral or aboral NSM at 24 hpf. Thus, the NSM is firmly split into two non-overlapping subregions that display distinct regulatory states.

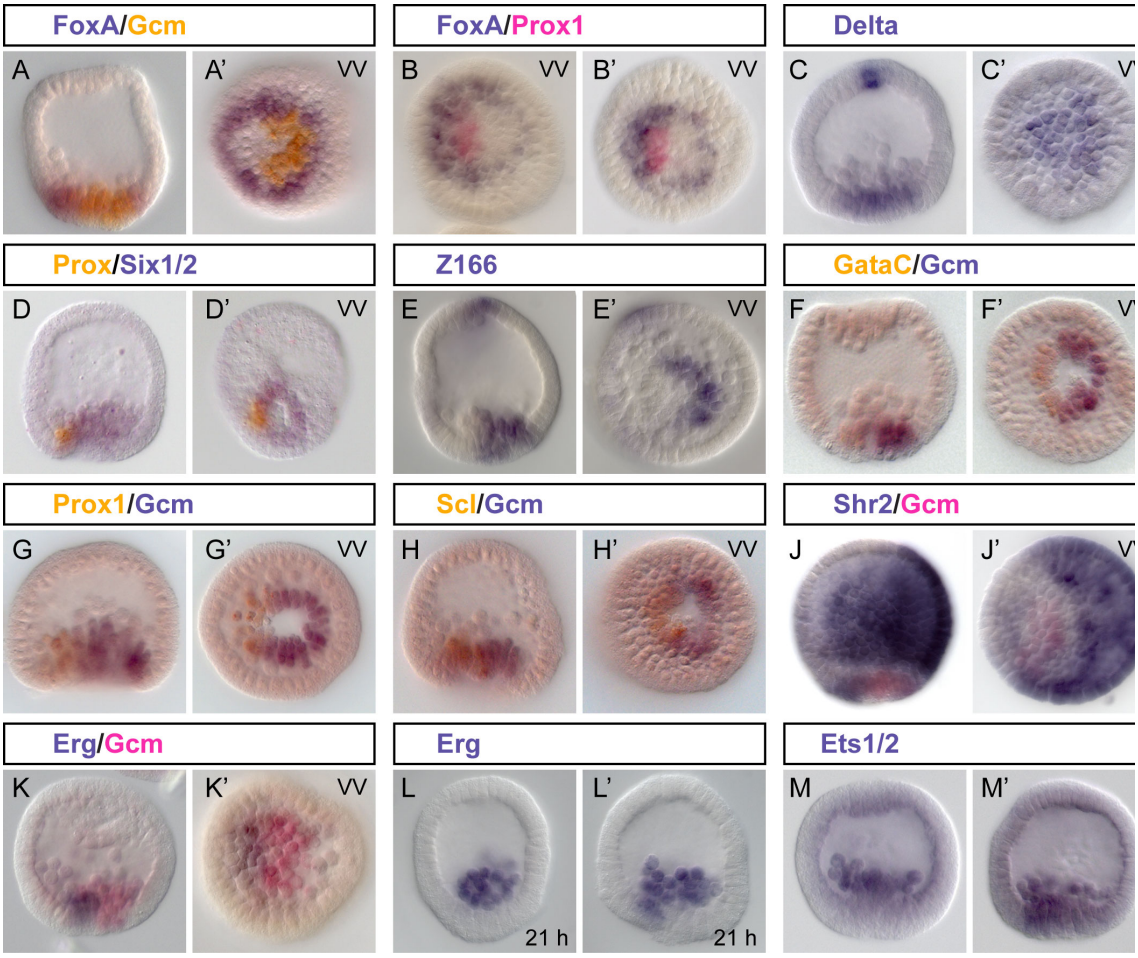


Figure 5.1: Spatial expression of regulatory genes in the non-skeletogenic mesoderm (NSM). A, B) At mesenchyme blastula stage a ring of *foxA*-expressing endodermal cells demarcates the NSM. C) The only gene expressed in the entire NSM is *delta* (it is also expressed in the apical domain. All other genes are expressed in either the oral or aboral NSM only. Aboral NSM genes shown here are *gcm* (A, F, G, H, J, K), *six1/2* (D), and *z166* (E; also expressed in the ciliary band). Genes expressed on the oral side include *prox1* (B, D, G), *gataC* (F), *scl* (H), *shr2* (J, also expressed in the aboral ectoderm), *erg* (K, L), and *ets1/2* (M). *erg* and *ets1/2* are first expressed in the skeletogenic mesoderm (SM) (L, M) but are turned on in NSM at mesenchyme blastula stage. At 21 hpf *erg* is still restricted to the SM in about half of the embryos examined (L) but at 24 hpf all embryos show expression in oral NSM. Similarly, *ets1/2* is primarily expressed in the SM at 24 hpf, as less than a quarter of embryos show expression in the oral NSM as well. Later, *ets1/2* is turned off in the SM and is expressed in the oral NSM only.

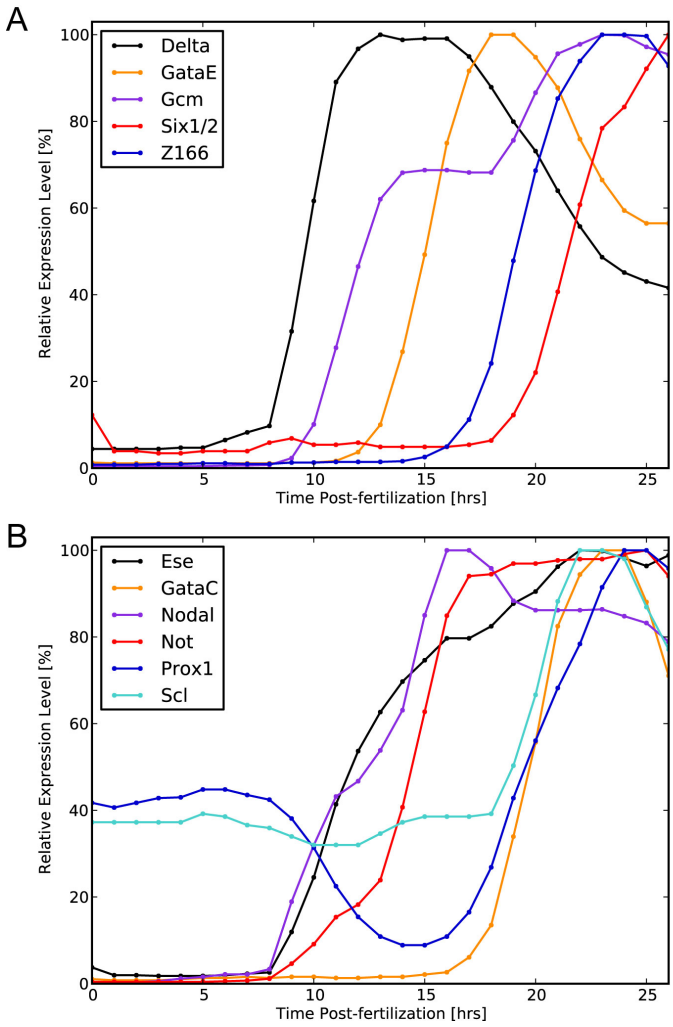


Figure 5.2: Temporal expression profiles of NSM genes. (A) The earliest genes expressed in the entire ring of NSM precursors are *gcm* and *gataE*. Both are direct targets of the Delta signal emanating from the neighboring skeletogenic mesoderm (SM). As *gcm* and *gataE* become restricted to the aboral side (see Fig. 5.4, 5.6), the zinc finger gene *z166* and the homeobox gene *six1/2* are activated on the aboral side. (B) Nodal signaling originating in the ectoderm entrains the oral/aboral axis and activates *not* expression on the entire oral side of the embryo (Fig. 5.4). By 16 hpf *prox1* has been activated in the oral NSM, followed by *gataC*. *ese* is transcribed starting at 9 hpf but expression is initially ubiquitous. By 18 hpf it is clearly expressed in the oral NSM (Fig. 5.8L). *scl* expression is activated specifically in the oral NSM by 19 hpf. Expression levels are given as the fraction of peak expression during the time range tested. For absolute transcript levels see Materna et al. (2010).

The subdivision of the NSM occurs along the oral/aboral axis, the establishment of which is dependent on Nodal/TGF-β signaling originating in the oral ectoderm (Duboc et al., 2004; Su et al., 2009). Consequently, perturbation of Nodal signaling disrupts expression of oral gene expression across the embryo, including in the NSM. To reveal the effects of Nodal signaling on gene expression, we perturbed Nodal signaling by injection of a Nodal MASO. We quantified the transcript levels of Nodal-MASO- and control-MASO-injected embryos using the Nanostring nCounter, an RNA counting device (Geiss et al., 2008). We made use of a custom-designed codeset that contains gene-specific probes for the majority of all transcription factors that are expressed in a spatially restricted manner in the early sea urchin embryo (Materna et al., 2010). Therefore, the perturbation effects were assessed in a comprehensive manner with regard to spatially restricted regulatory factors. We supplemented this analysis with QPCR analyses for pigment cell differentiation genes that were not included in the codeset. All together we examined transcript level of 206 genes.

Nodal signaling activates not *expression*

As expected, Nodal perturbation causes a significant reduction in oral NSM gene transcript levels at the mesenchyme blastula stage (24 hpf) (Fig. 5.3D), indicating that no oral NSM is specified. Similarly, the expression level of several aboral NSM genes shows a slight increase, confirming previous findings that these genes are expressed throughout the entire NSM as a result of Nodal perturbation (Duboc et al., 2010).

Given that *nodal* expression begins at about 8 hpf and the onset of oral NSM gene expression does not begin until 16 hpf, it is unlikely that Nodal directly activates the oral NSM genes. The earliest known Nodal targets include several signaling related genes such as *bmp2/4*, *lefty*, and *chordin* (Fig. 3 A). The Bmp2/4 signal primarily functions as an activator on the aboral side of

the embryo, whereas Lefty is an antagonist of Nodal and is thought to limit its effective range (Duboc et al., 2008; Su et al., 2009). The previously characterized transcriptional regulators activated by Nodal signaling are mostly restricted to the ectoderm (*foxG*, *gsc*, Fig. 5.3B, C) (Su et al., 2009).

The expression level of the homeobox transcription factor *not*, which has not been previously characterized, is significantly decreased in the Nodal perturbation (Fig. 5.3). *not* expression is similar to the expression of *nodal* in time and space, suggesting that *not* may be a direct target of Nodal signaling (Fig. 5.2B; Fig. 5.4). *nodal* expression is restricted to the ectoderm throughout early development. Initially its expression is relatively broad and may include veg1 ectoderm, however, *nodal* is clearly not expressed in veg2 cells and its descendants. In contrast, not is expressed on the oral side of the embryo, but reaches from the apical domain (as marked by foxQ2 expression) to the alx-expressing SM when it is first detectable by in situ staining (Fig. 5.4F–H). *not* expression thus includes the oral NSM where it overlaps with *gcm* expression, and it is the *not* expressing cells that will lose expression of the aboral genes, including *gcm*. By about 22 hpf *not* and *gcm* expression are mutually exclusive. At 24 hpf *not* expression is no longer detected in the NSM but continues to be expressed in the oral ectoderm.

Despite the strong correlation between *nodal* and *not* expression, it is possible that *not* in the NSM is activated by Delta signaling rather than Nodal signaling. All genes that are expressed specifically in the NSM are either directly or indirectly dependent on Delta (see Chapter 4), and no NSM is specified when Delta signaling is perturbed. However, when SM Delta signaling was perturbed by addition of DAPT (a γ-secretase inhibitor that prevents cleavage of the Notch intracellular domain) at 3 hpf, well before Delta signaling commenced, *not* expression is not affected; it still extended down to the vegetal pole at 21 hpf where it complements *gcm* expression in the NSM (Fig. 5.4M, N). Therefore, *not* expression is independent of Delta and its NSM targets, and is accounted for entirely by Nodal activation.

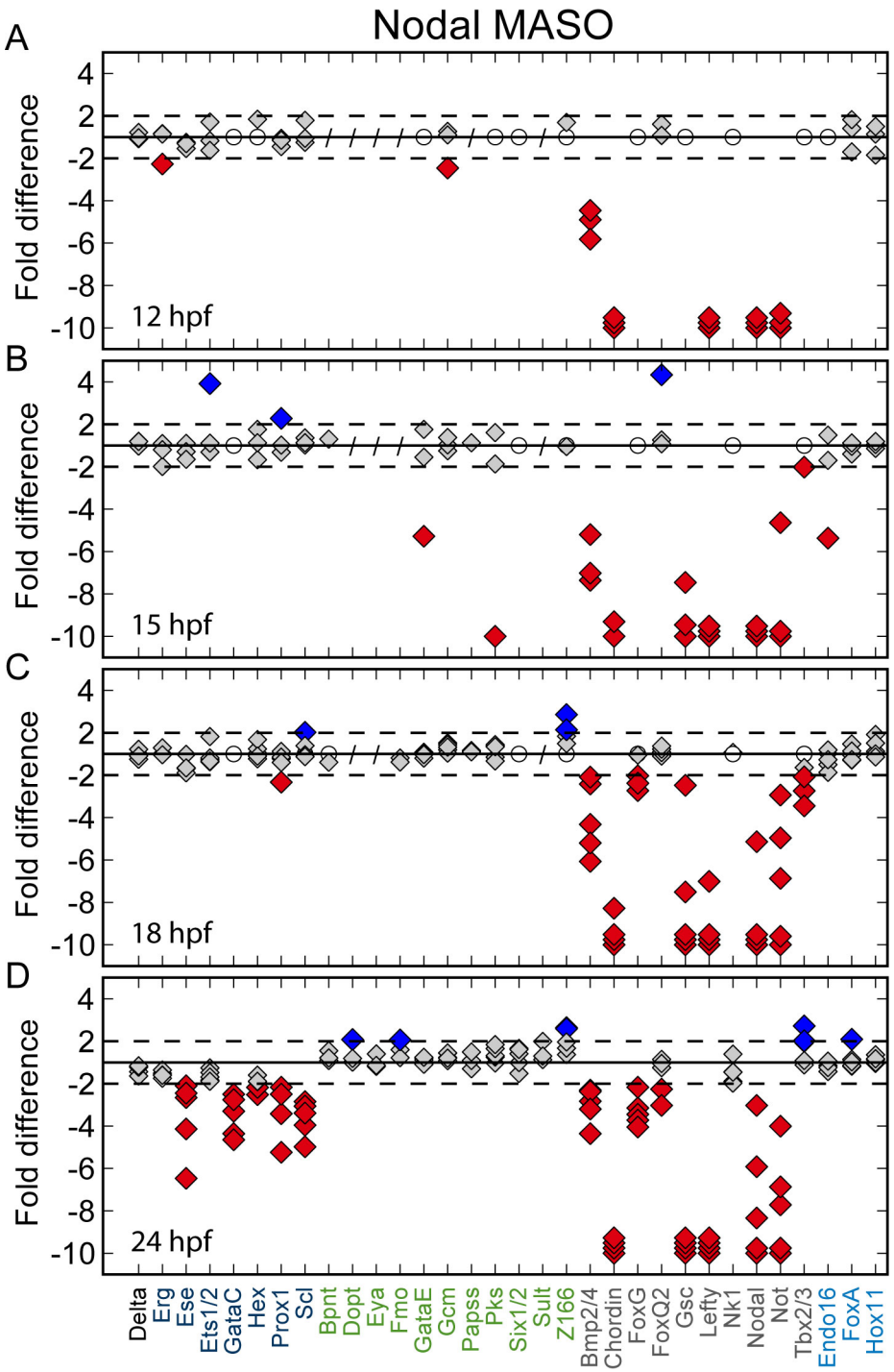


Figure 5.3: Effect of Nodal MASO treatment on gene transcription. RNA was extracted from embryos injected with Nodal MASO or control MASO, and transcript levels were determined with the Nanostring nCounter. The resulting counts were used to calculate fold differences. A, B) The first genes affected by the perturbation are the known Nodal targets *bmp2/4*, *chordin*, *lefty*, and *nodal,* that are expressed in the oral ectoderm and the previously undescribed *not* gene.

C) By 18 hpf a mild increase in the expression level of some aboral NSM genes can be seen (e.g., *hpnt*, *z166*). Although expressed, levels of oral NSM genes are still too low for significant changes to occur. D) At 24 hpf the transcript levels of oral NSM genes are strongly reduced (*ese, gataC, prox1, scl*). Aboral genes show slightly elevated transcript levels (e.g., *z166, fmo* in the NSM*, tbx2*/3 in the aboral ectoderm). For presentation purposes fold differences of more than tenfold are shown as tenfold. Each diamond represents a single experiment. Oral NSM genes are labeled dark blue, aboral NSM genes green, ectoderm genes grey, and endoderm genes light blue.

Not regulates mesoderm polarity

We determined the transcription start site of *not* by RACE PCR and subsequently amplified a full length version from cDNA. Sequencing revealed that a previously published sequence of not (AF109903) is incorrect. The newly determined sequence was confirmed independently by RNAseq (JQ945921; Qiang Tu and EHD) and we designed a translation-blocking MASO based on this sequence. Development of Not-MASO-injected embryos proceeds normally, except the embryos have an increased number of pigment cells that accumulate underneath the apical domain as compared to control embryos. Not MASO embryos also show defects in skeletogenesis and produce supernumerary spicules on the oral side.

In Not-MASO injected embryos, *gcm* expression expands to the entire NSM, as shown by *in situ* staining (Fig. 5.5A, B). Global evaluation of the effects of Not MASO injection on gene expression using the Nanostring nCounter reveals that only a small number of genes are affected. Therefore, this perturbation causes specific defects that do not interfere with the general progression of development. The expression levels of oral NSM genes is significantly lower in Not-MASO-injected embryos, as compared to controls (Fig. 5.5E). Since Gcm is the main driver of pigment cell specification (Davidson et al., 2002; and below), this explains the increase in pigment cell number. Thus, Not is required for repression of *gcm* expression in the oral NSM, a function that had previously been ascribed to Nodal (Duboc et al., 2010). These effects are almost identical to those observed in Nodal perturbations; this confirms that the loss of Not function causes an expansion of aboral NSM at the expense of oral NSM. Not perturbation thus recapitulates all Nodal perturbation effects as they pertain to the NSM, thus

demonstrating that Nodal signaling directs oral NSM specification through its activation of not expression.

Injection of Not MASO affects the expression level of only a few oral ectoderm genes (e.g., *gsc*, *nk1*; Fig 5.5). Notably, the expression of *nodal* and its known targets (*lefty*, *bmp2/4*) are not affected. This underscores the phenotypic observation that establishment of the O/A axis is not affected; embryos do not radialize, as is observed when Nodal function has been lost (Duboc et al., 2004). Nevertheless, the perturbation effects on oral ectoderm genes and the effects on spicule formation, a process that is controlled by the ectoderm, indicates that Not function is not limited to the NSM, but instead extends to its entire expression domain. The function of Not in the GRN covering ectoderm specification will be discussed elsewhere.

We injected *not* mRNA to determine whether Not functions as a repressor. Evaluation of gene expression at 15 hpf, i.e. before the NSM has been subdivided, shows a significant reduction of *gataE* and *pks* transcript levels. gcm levels are reduced by only 30%, but this effect is observed in two independent experiments (Fig. 5.5F). Two other lines of evidence support the notion that Not can function as a repressor: First, Not MASO injection causes an increase of its own transcript level (Fig. 5.5 E) indicating that Not may repress its own transcription. Second, the formation of supernumerary spicules is due to the expansion of *vegf3* transcription into the *not* expression domain. Not appears to repress *vegf3* and thus restrict its expression to two lateral patches where spicules normally form (EL and EHD, to be discussed elsewhere). Although overexpression of *not* does not prove that Not directly represses transcription of *gcm* and *gataE* in the oral NSM, the data nevertheless strongly supports the idea that Not acts as a repressor.

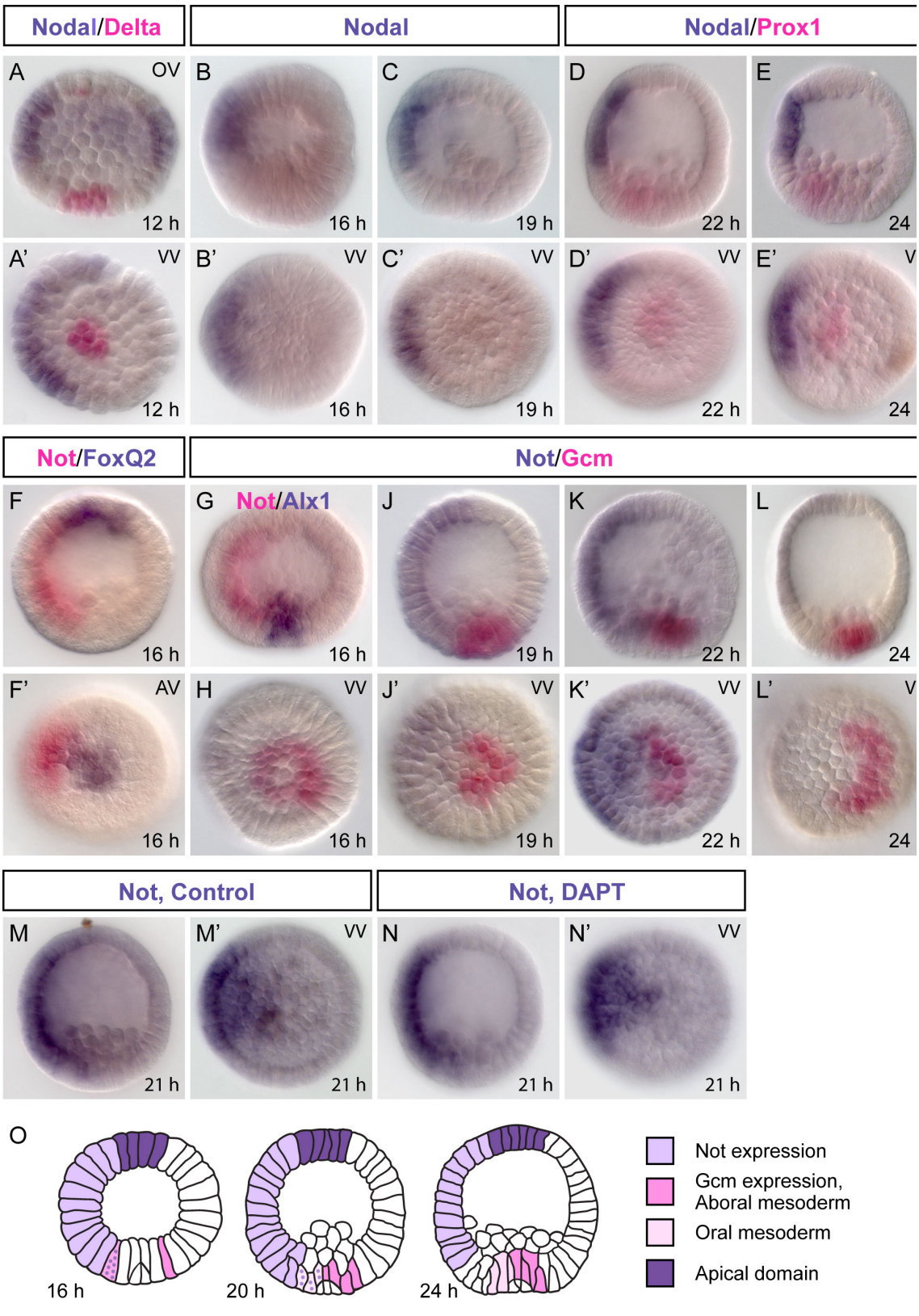

Nodal/Delta
Nodal
Nodal/Prox1
A
OV
12 h
B
16 h
C
19 h
D
22 h
E
24
A'
VV
12 h
B'
VV
16 h
C'
VV
19 h
D'
VV
22 h
E'
V
24
Not/FoxQ2
Not/Gcm
F
16 h
G
Not/Alx1
16 h
J
19 h
K
22 h
L
24
F'
AV
16 h
H
VV
16 h
J'
VV
19 h
K'
VV
22 h
L'
V
24
Not, Control
Not, DAPT
M
21 h
M'
VV
21 h
N
21 h
N'
VV
21 h
O
16 h
20 h
24 h
Not expression
Gcm expression,
Aboral mesoderm
Oral mesoderm
Apical domain

Figure 5.4: Spatial expression of *nodal* and *not*. A) At 12 hpf *nodal* is expressed on the oral side of the ectoderm and is excluded from the vegetal half of the embryo; two rows of cells separate the *nodal* expression domain from the *delta* expressing SM. B, C) Subsequently, the *nodal* expression domain gets restricted to a wedge that covers about a third of the embryo at the equator. *nodal* may extend into veg1 ectoderm but is clearly not expressed in the NSM. D, E) When transcripts of oral mesoderm genes can first be detected by in situ staining, the NSM cells are at least three cell diameters away from the Nodal source. Past 24 hpf *nodal* expression becomes confined to a subregion within the oral ectoderm. F, G) When *not* transcript can first be detected at 16 hpf, its expression domain spans the entire oral side of the embryo and borders *foxQ2*-expressing cells apically and *alx1*-expressing cells in the SM at the vegetal pole. H) Although the blue staining in panel H is weak, *not* is expressed in NSM cells and overlaps with *gcm* expression (see also F, G). K) While SM cells ingress, *not* expression remains strong in oral NSM. By 21 hpf *gcm* expression has stopped in the oral NSM so that it no longer overlaps with *not* expression which continues. L) Eventually *not* expression will fade in the NSM (and endoderm) while it continues to be expressed in the ectoderm. M, N) All genes specifically expressed in the NSM depend on Delta signaling. However, *not* expression in the NSM is independent of Delta signaling as DAPT treatment does not affect spatial expression of *not*. (O) A scheme summarizing the spatial expression of *not* and NSM genes. Unless otherwise noted embryos are presented in a lateral view with their oral side oriented to the left. VV—vegetal view, AV—apical view, OV—oral view.

Interactions of oral NSM genes

Expression of *not* in the NSM is temporary, but when it is turned off the regulatory state of the oral NSM has been firmly established, indicating that the repression of aboral NSM genes by Not is no longer necessary. This may be because the activators of aboral NSM genes are no longer expressed in the oral NSM. Alternatively, aboral NSM genes may be repressed in the oral NSM by the products of the regulatory genes now active in these cells. To test these possibilities, we knocked down expression of the three genes that are first activated specifically in the oral NSM: *prox1*, *ese* and *gataC*.

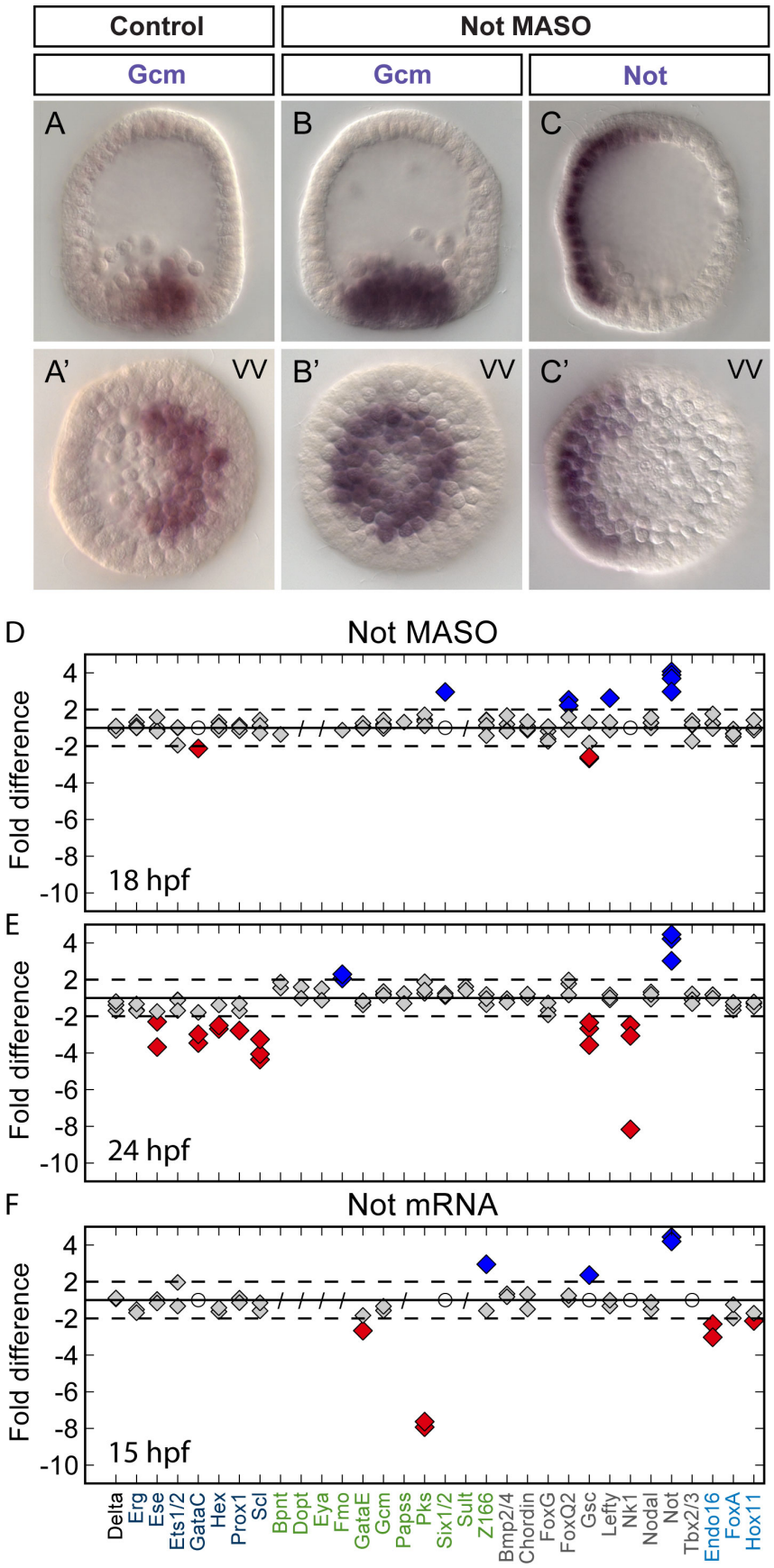

Control
Not MASO
Gcm
Gcm
Not
A
B
C
A'
VV
B'
VV
C'
VV
D
Not MASO
Fold difference
4
2
-2
-4
-6
-8
-10
18 hpf
E
Fold difference
24 hpf
F
Not mRNA
Fold difference
15 hpf
Delta
Erg
Ese
Ets1/2
GataC
Hex
Prox1
Scl
Bpnt
Dopt
Eya
Fmo
GataE
Gcm
Papss
Pks
Six1/2
Sult
Z166
Bmp2/4
Chordin
FoxG
FoxQ2
Gsc
Lefty
Nk1
Nodal
Not
Tbx2/3
Endo16
FoxA
Hox11

Figure 5.5: Effect of Not MASO treatment on gene transcription. Not MASO abolishes specification of oral NSM. A) In control embryos, *gcm* expression is restricted to the aboral NSM at 24 hpf. B) Not MASO expands *gcm* expression to the entire NSM. C) *not* expression is stronger in Not MASO injected embryos although its spatial expression is similar to controls (compare C to Fig. 4 L). D, E) Quantification of transcript abundance in Not MASO injected embryos reveals an increased transcript level for *not*. Only two oral ectoderm genes are significantly affected by the perturbation (*gsc, nk1*). At 24 hpf, oral NSM genes have strongly reduced transcript levels while some aboral NSM genes (e.g. *fmo*) show the opposite effect. F) F) Injection of *not* mRNA has a repressive effect and reduces the expression of *pks* and *gataE* at 15 hpf. Labels and cutoffs as in Fig. 5.3. Embryos in A–C are oriented with their oral side to the left. VV—vegetal view.

Treatment of embryos with GataC MASO has no notable effect on gene expression at 24 hpf (Fig. 5.7A). Members of the Rast lab have confirmed this independently. Instead GataC MASO treatment causes specific migration defects when blastocoelar cells ingress into the blastocoel (Cynthia Solek and Jonathan Rast, personal communication). Thus the function of GataC may be to control morphogenetic processes, rather than to lockdown the oral NSM regulatory state. Injection of Ese MASO causes only a relatively mild reduction of *prox1* transcript levels. This indicates that Ese is an activating input into *prox1* (Fig. 5.7 B). Injection of Prox1 MASO is lethal as embryos arrest at late cleavage stage, which is likely due to the loss of maternal transcript translation (Fig. 5.2B). Embryos injected with even a low concentration of Prox1 MASO often do not form a blastula, contain cells of varying sizes, and in many cases disintegrate by the time control embryos reach the mesenchyme blastula stage. We obtained the same results with two different MASOs (one translation blocking, one splice blocking) indicating that this is a real perturbation effect and not the result of non-specific MASO toxicity.

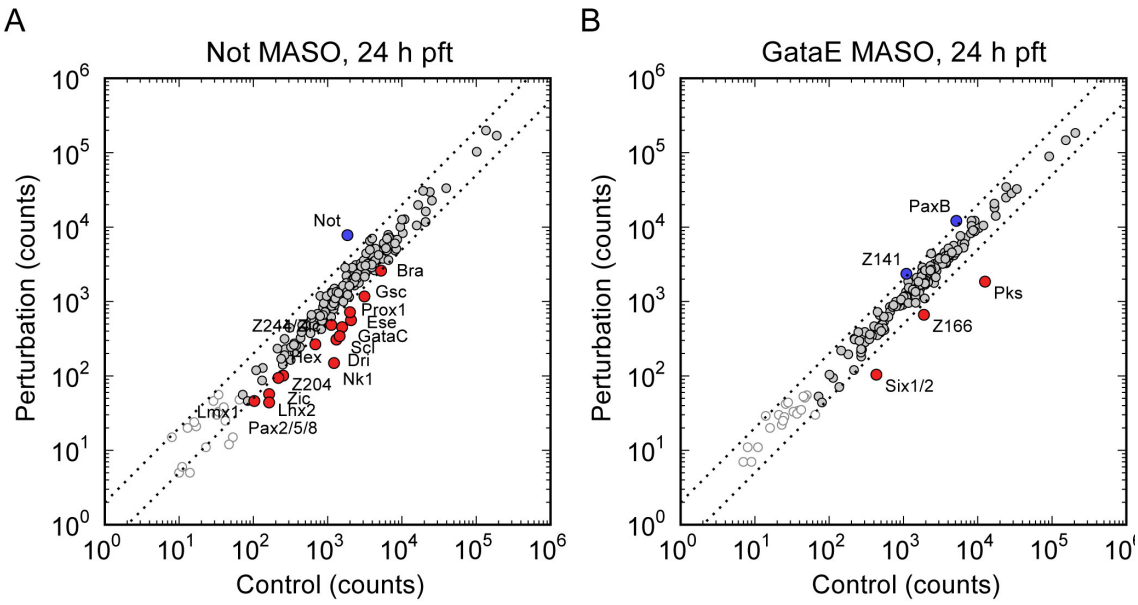


Figure 5.6: Evaluation of Not and GataE MASO treatments. RNA was extracted from injected embryos and quantified using the Nanostring nCounter. Counts for RNA from perturbed embryos are plotted against counts from unperturbed embryos. Genes that are not affected by the perturbation align on the diagonal. The small number of genes affected by either MASO treatment indicates that the MASOs are specific and do not cause any broad developmental defects. Dotted lines indicate a threshold of twofold change. Transcription levels were estimated from previous quantification data (Materna et al., 2010); genes present with 25 transcripts or less per embryo are marked with an open, grey circle.

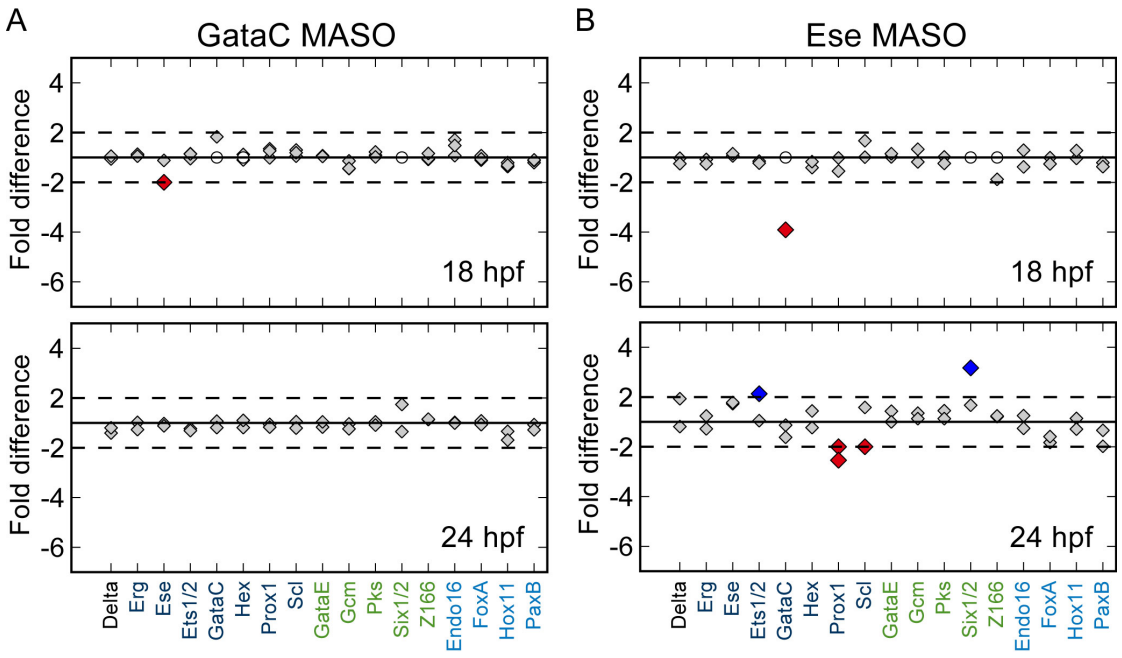


Figure 5.7: Effects of GataC and Ese MASO treatment on gene expression. GataC MASO injection does not appear to have any noticeable effect by 24 hpf. Ese-MASO-injected embryos show a slight reduction of *prox1* indicating that Ese may be an activating input. Conversely, in these embryos *six1/2* expression is slightly higher compared to controls, suggesting that Ese may function to repress aboral NSM genes on the oral side after *not* is no longer expressed in the NSM.

Interactions of the aboral mesoderm genes

The two earliest genes to be activated in the veg2 NSM precursors are *gcm* and *gataE* (Fig. 5.2B) and both are direct targets of Delta signaling (Lee, 2007; Ransick and Davidson, 2006). As the oral NSM is specified, both *gcm* and *gataE* are restricted to the aboral side (Figs. 5.4H–L, Figs. 5.8A–D). Previous work demonstrated that Gcm MASO injection abolishes development of the pigment cells that arise from aboral NSM (Davidson et al., 2002; Ransick and Davidson, 2006). Here we evaluated the effects of Gcm MASO injection on the transcription levels of early sea urchin regulatory genes with the Nanostring nCounter, and supplemented this dataset with QPCR evaluation of pigment cell genes. At 15 hpf Gcm knockdown decreases *pks* expression as expected, but decreased levels of *gataE* transcript were observed as well (Fig. 5.8E). Thus, it appears that Gcm is an input into *gataE* in the mesoderm.

At 18 hpf (not shown) and 24 hpf three additional regulatory factors exhibit strongly reduced transcript levels in Gcm perturbations. These genes are *z166*, a zinc finger gene (Materna et al., 2006), the homeobox gene *six1/2* (Howard-Ashby et al., 2006), and its co-factor *eya* (Fig. 5.2B, Fig. 5.8F). In addition, as previously described (Davidson et al., 2002), pigment cell differentiation gene expression is lost in Gcm-MASO-injected embryos.

gataE expression expands from the aboral NSM to the endoderm by 24 hpf, and is eventually expressed at the tip of the archenteron and in the mid and hindgut (Lee and Davidson, 2004). The role of GataE with regard to mesoderm development has not been examined before. Injection of a GataE MASO has very similar effects to Gcm MASO treatment, both phenotypically and with regard to effects on gene expression (Fig. 5.8G). GataE MASO also causes a severe reduction in the number of pigment cells specified (Fig. 5.9H). We repeated these experiments with a second (splice-blocking) GataE MASO, and the injected embryos displayed a similar reduction in pigment cell number; we thus consider these MASOs to be specific.

GataE MASO injection affects a very similar set of genes as Gcm MASO, including *z166*, *six1/2*, *eya* and several pigment-cell-differentiation genes. This result indicates that *gataE* may indeed be downstream of Gcm and the effects on these aboral NSM genes may, at least in part, be indirect. Although the Gcm And GataE perturbation effects are similar, they are not identical. The magnitude of change is usually smaller in GataE MASO treatment compared to Gcm MASO. One possible explanation for this result is that Gcm and GataE are required together to activate their downstream targets, as has been demonstrated for *pks* (Calestani and Rogers, 2010). Loss of GataE causes a reduction in levels of downstream targets but does not affect the Gcm function, thus resulting in a lower fold change in the expression of target genes. Furthermore, GataE MASO treatment affects two genes repeatedly that are not affected by Gcm MASO: *paxB* and *phb1*. *phb1* is expressed in the SM and GataE may serve to inhibit its expression in the NSM. paxB, however is expressed in the aboral ectoderm, and it is unclear what the relationship of GataE and PaxB is.

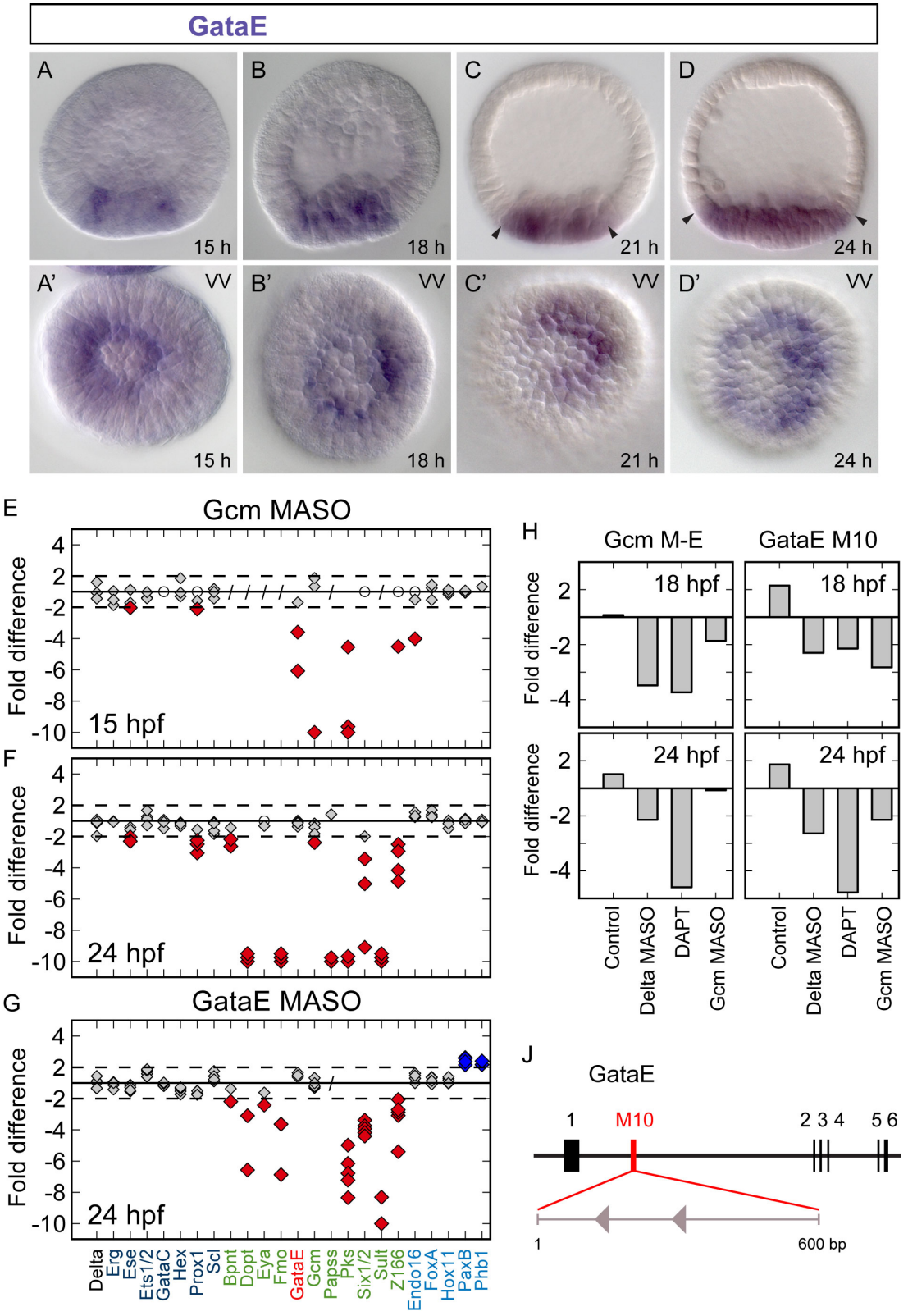

GataE
A
B
C
D
15 h
18 h
21 h
24 h
A'
B'
C'
D'
VV
E
Gcm MASO
Fold difference
15 hpf
F
24 hpf
G
GataE MASO
Delta
Erg
Ese
Ets1/2
GataC
Hex
Prox1
Scl
Bpnt
Dopt
Eya
Fmo
GataE
Gcm
Papss
Pks
Six1/2
Sult
Z166
Endo16
FoxA
Hox11
PaxB
Phb1
H
Gcm M-E
GataE M10
18 hpf
24 hpf
Control
Delta MASO
DAPT
Gcm MASO
J
GataE
1
M10
2 3 4
5 6
600 bp

Figure 5.8: Expression and function of aboral NSM transcription factors. A, B) *gataE* is first expressed in veg2 mesoderm precursors (compare to *gcm* expression, Fig 5.4H–L). C, D) It remains strictly mesodermal until after 21 hpf, when it becomes expressed in endoderm as well (arrowheads in C, D). C') At 21 hpf *gataE* transcripts, like *gcm* transcripts, disappear from the oral NSM. D) At 24 hpf the oral NSM cells do not express *gataE* but are surrounded by aboral NSM and endodermal cells expressing *gataE*. E, F) Gcm MASO injection causes a significant reduction in *gataE* levels at 15 hpf indicating that it may be an early activator of *gataE* transcription in the NSM. In addition, Gcm MASO causes a significant reduction of all aboral NSM genes, including pigment cell differentiation genes. G) Injection of GataE MASO also causes a reduction of transcript levels of aboral NSM while oral NSM genes *ets1/2* and *scl* show a slight increase. H) GataE module 10 is the cis-regulatory element that drives early expression of *gataE*. Injection of a reporter under the control of this element is sensitive to Delta perturbations confirming previous findings. However, this element is also affected by Gcm MASO injection. J) GataE module 10 lies in the intron between exon 1 and 2 and contains two putative Gcm binding sites (grey triangles) suggesting that Gcm is a direct input into *gataE* (not to scale). Gcm module E has previously been shown to be directly activated by Delta/Notch signaling and serves as a positive control. Color of labels in E–G as in Fig. 5.3.

The cis-regulatory module that drives early *gataE* expression has previously been identified (Lee et al., 2007). To query whether Gcm may directly activate *gataE* expression, we injected GataE cis-regulatory module 10 together with Delta or Gcm MASO, or the injected embryos were treated with DAPT to perturb Delta/Notch signaling (see Chapter 4). We injected Gcm module E that contains Su(H) sites as a positive control for D/N signaling perturbation (Ransick and Davidson, 2006). As reported previously, Gcm module E and GataE module 10 showed significant reduction in transcription in Delta perturbations. GataE module 10 also contains Su(H) sites that confirm that GataE is a direct target of Delta (Lee, 2007). Gcm MASO injection causes a strong reduction in the expression of GataE module 10, and this is consistent with the effect of the perturbation on endogenous gene expression (Fig. 5.8H). The organization of the GataE gene is shown in Fig. 5.8J. Module 10 lies downstream of the first exon and contains two potential Gcm binding sites that are similar to those found in the regulatory region of the *pks* gene (IUPAC sequence: ATRCGGGY). Therefore, our data suggests that Gcm is a direct input into *gataE* and operates through the identified binding sites, although this remains to be confirmed by mutation.

Aboral NSM excludes oral NSM gene expression

We confirmed the effects of Gcm and GataE MASO treatments on aboral gene expression by in situ staining (Figs. 5.9A–H). In agreement with the lack of pigment cells, *hpnt* and *pks* stain either weakly or not at all in MASO injected embryos. *gcm* expression can still be observed, but the number of *gcm*-positive cells appears to be lower (Figs. 5.9C–E), indicating that Gcm or its downstream targets are necessary for maintenance of *gcm* expression in the entire aboral NSM by feeding back into *gcm*. In sharp contrast, the expression of oral NSM genes is significantly expanded within the NSM in both perturbations. In control embryos, oral NSM genes are usually expressed in 10–12 cells. In Gcm- and GataE-MASO-injected embryos, this number significantly increases, often to the entire NSM (e.g., Figs. 5.9 M', P', Q', V', W'). Thus, Gcm and GataE, or their downstream targets, repress expression of oral NSM genes in the aboral

NSM. The fact that all oral NSM genes are now running in the entire NSM indicates that the driver(s) of their expression are not restricted to the oral NSM but active throughout the NSM.

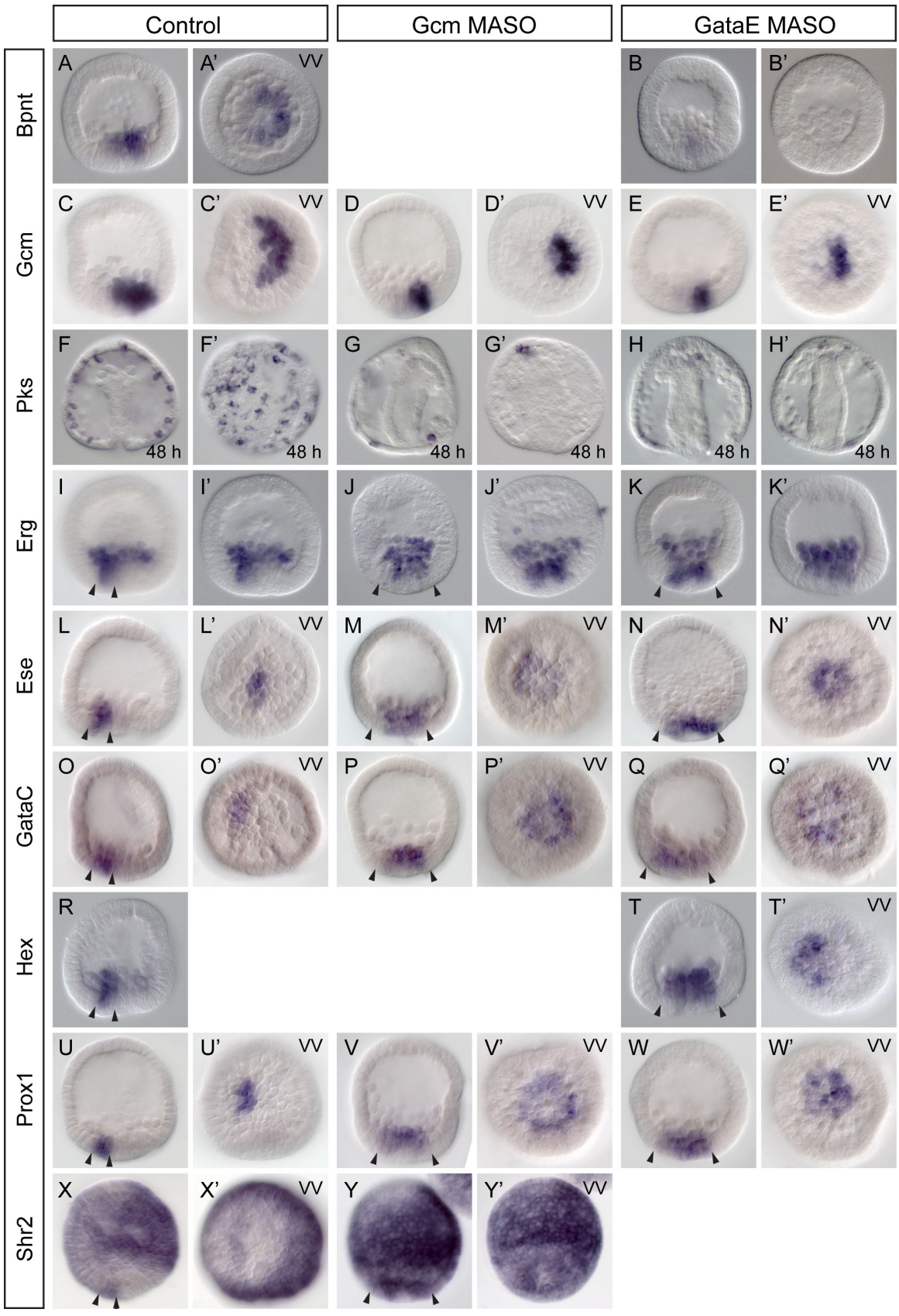

Control
Gcm MASO
GataE MASO
Bpnt
Gcm
Pks
Erg
Ese
GataC
Hex
Prox1
Shr2
A
A'
VV
B
B'
C
C'
D
D'
E
E'
F
F'
48 h
G
G'
H
H'
I
I'
J
J'
K
K'
L
L'
M
M'
N
N'
O
O'
P
P'
Q
Q'
R
T
T'
U
U'
V
V'
W
W'
X
X'
Y
Y'

Figure 5.9: Spatial effects of Gcm and GataE MASO treatment on NSM gene expression. Injection of either Gcm or GataE MASO causes a severe reduction in transcript levels of pigment cell specification genes (*bpnt*, A and B; *pks*, F–H). These embryos develop very few, if any, pigment cells, as demonstrated by *pks* expression at 48 hpf (F–H). *gcm* expression appears to be only slightly affected in Gcm- or GataE-MASO-injected embryos. The number of *gcm*-positive cells appears to be lower than in control embryos, although expression in these cells is strong (C–E). In contrast, Gcm or GataE MASO treatment causes an expansion of all oral NSM genes (I–Y, see distance between arrowheads in lateral views of control and MASO-treated embryos). Genes that are also expressed in non-NSM cells (e.g. *erg*, *shr2*) show an NSM specific expansion (I–K,X–Y). All embryos are 24 hpf unless otherwise noted. Embryos are oriented with their oral side to the left in lateral views. VV—vegetal view.

Discussion

In this study we explore the specification and subdivision of the sea urchin NSM. Specification of the entire NSM is dependent on activation of mesodermal genes by D/N signaling. The gene first and most strongly activated by D/N is *gcm*. It becomes expressed in the endomesoderm, and after the veg2 tier divides into an inner and outer ring, remains expressed only in the inner ring of NSM precursors that are neighboring the Delta source.

Gcm perturbation experiments reveal *gataE* to be a target of Gcm. To address whether this link is direct, we performed reporter assays for a regulatory module of *gataE* that is known to drive its early expression in the mesoderm (Lee et al., 2007). Co-injection of Delta MASO strongly affects the expression of this module. Previous work identified functional Su(H) sites in this module, but mutation of these sites were found to have a weaker effect on expression of the construct than the trans-perturbation (Delta MASO injection). Thus it had been predicted that a second gene, also downstream of D/N, also activates *gataE*. Injection of module 10 reporter together with Gcm MASO shows a strong reduction of expression from this construct. This experiment reveals that Gcm is the missing intermediate. Although we did not perform the cis-mutation, *gataE* module 10 contains two putative Gcm binding sites that are similar to those in the *pks* regulatory region to which Gcm binds to activate *pks* (Calestani and Rogers, 2010). It is likely that Gcm binds to these sites in *gataE* module 10 to activate *gataE* transcription.

We here show that *gataE* is expressed in a similar fashion as *gcm* up to 24 hpf, when it also turns on in the endoderm. Gcm and GataE are strong drivers of pigment-cell-differentiation-gene expression. Consequently these genes are expressed already in the full ring of NSM precursors (Calestani et al., 2003). GataE and Gcm are simultaneously active, direct inputs into at least *pks* (Calestani and Rogers, 2010). We show that other pigment cell differentiation genes are also strongly affected by perturbation of GataE function; both GataE and Gcm may be direct inputs into several pigment-cell-differentiation-genes.

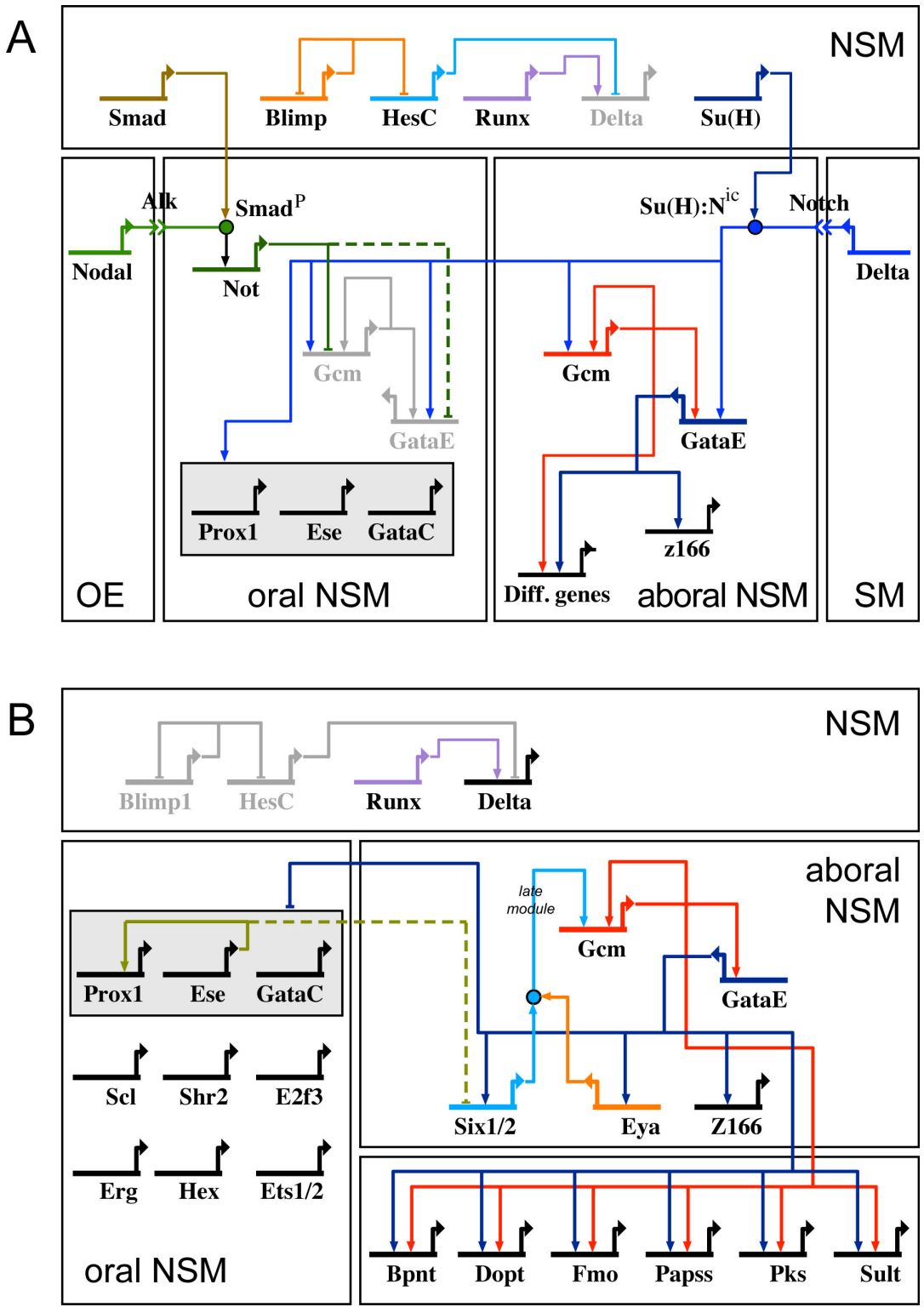


Figure 5.10: Gene regulatory network for NSM specification and partitioning. (A) Initially, Delta in neigboring skeletogenic mesoderm (SM) activates *gcm* and *gataE* in the entire NSM.

Gcm is a direct input into *gataE*. Together both turn on pigment cell differentiation genes that are first expressed throughout the entire NSM. Nodal signaling from the oral ectoderm (OE) activates *not* expression in the oral NSM. Not in turn represses *gcm* and *gataE*. Because *gataE* is downstream of Gcm, Not repression of *gcm* may be sufficient to explain loss of *gataE* in the oral NSM. Clearance of both genes allows activation of *prox1, ese*, and *gataC* in the oral NSM. Our results require the driver of their expression to be present throughout the entire NSM, a condition that is fulfilled by Delta signaling from the SM. Expression of *prox1, ese, gataC*, and other oral NSM genes has previously been shown to be dependent on SM Delta (see Chapter 4). After birth of the NSM precursors, endoderm genes such as *blimp* are coexpressed. Their expression depends on nuclear β-catenin/TCF and stops when β-catenin is removed from the nuclei as a consequence of Notch signaling (Peter and Davidson, 2011). *blimp* expression is included here as it is required for repression of HesC. (B) Following subdivision, the two NSM segments express distinct sets of genes. Only the signaling factor Delta, now active in the NSM, is expressed in the entire NSM as a consequence of the disappearance of its repressor (HesC). Early signaling inputs into oral and aboral NSM (Nodal, Delta) have stopped. Thus, the transcription factors present in each segment are sufficient to maintain the current regulatory state. In the aboral NSM, GataE activates the homeobox gene *six1/2* and its co-factor *eya*. Six1/2/Eya are a direct input into the late module of *gcm* (AR and EHD, to be discussed elsewhere) and provide the necessary feedback for the specification state of the aboral NSM to be locked down. GataE also assures repression of oral NSM genes on the aboral side. We added several new genes to the oral NSM; however, their function in locking down the oral NSM regulatory state has yet to be determined.

Perturbation analysis revealed two additional regulatory genes downstream of Gcm and GataE (*z166*, *six1/2*). However, these are activated at a time when the NSM already shows subdivision into oral and aboral segments and are activated only on the aboral side. *z166* becomes transcribed between 16 and 17 hpf. The reason for its relatively late expression may be that is relies primarily, or entirely, on activation by GataE rather than Gcm. *gataE* expression reaches significant levels (> 100 transcripts/embryo) by 14 hpf (Materna et al., 2010), only two hours before *z166* is activated (Fig. 5.2A). This gap is characteristic for progressive gene activation events due to the slow rate of transcription and translation in sea urchins at an ambient temperature of 15°C (Davidson, 1986; Materna et al., 2010). Furthermore, while GataE perturbation causes a smaller reduction of transcript level for pigment cell genes than Gcm perturbation (due to the auxiliary input of Gcm into *pks* etc.), the observed changes in *z166* and *six1/2* transcript levels are almost identical. Hence, Gcm activation of *z166* and *six1/2* may be indirect and run through GataE. The linkages that we draw based on these experiments are presented in the GRN in Fig. 5.10.

The origin of the oral NSM had previously been traced to Nodal signaling. It is undoubtedly true that Nodal signaling is a prerequisite for oral NSM specification, as it functions to establish the O/A axis across the embryo. However, our results indicate that Nodal is unlikely to be a direct input into oral NSM genes. Two lines of evidence support this.

First, we have identified the homeobox gene *not* as a target of Nodal. Our perturbation experiments of *not* reveal that it performs all the functions previously attributed to Nodal as they pertain to the oral NSM including the repression of *gcm* and *gataE*. This is demonstrated by Not knockdown that causes *gcm* expression to continue in the oral NSM beyond the time it is expressed there in controls. The expression level of oral NSM genes is severely lower in Not-MASO-treated embryos, just as is observed in Nodal-MASO-treated embryos. This result shows that the oral NSM is not specified. As the spatial expression of *gcm* demonstrates, the cells instead display the regulatory state of aboral NSM. Not is a homeobox factor that may be both activator and repressor, depending on its co-factors. While we do not have any insights into what these co-factors are, we have acquired circumstantial evidence that it can function as a repressor. This includes the observation that it represses itself, as revealed by an increase in

transcript level in Not MASO treatment. Furthermore it represses *vegf3* in veg1 ectoderm (EL and EHD, to be discussed elsewhere). Also, injection of Not mRNA causes a strong loss of *gataE* and *pks* expression, thus strongly suggesting that Not is a repressor in the NSM.

Second, knockdown of aboral NSM genes allows expression of oral NSM genes on the aboral side. Thus, expression of oral NSM genes requires an activator that is present throughout the NSM. Nodal signaling is thought to operate only over fairly short distances of up to a few cell diameters (Duboc et al., 2008). We have shown that *nodal* expression is confined to the oral ectoderm throughout early development, putting it out of range of the aboral NSM. If *not* expression is a proxy for where Nodal signaling is received in the vegetal plate, it proves that Nodal cannot be the sought-after pan-mesodermal activator. Not itself could function as an activator. However, after loss of *gcm* expression as it occurs in D/N perturbed embryos, *not* expression is not affected, thus indicating that its expression is entirely dependent on Nodal.

Aboral NSM genes have an antagonistic relationship with oral NSM and are unlikely candidates for activation of oral NSM genes. However, Delta signaling from the SM may be the activator of oral NSM genes. As has been discussed elsewhere, *delta* expression in the SM is necessary for specification of all NSM, including *prox1*, *ese*, *gataC* and other oral mesoderm genes. However, *prox1* and *gataC* are much delayed compared to when Delta is first received by the NSM. The cause of this delay may simply be that Gcm and GataE have to be excluded from the NSM before their expression can commence. *not* expression is under way well in time before SM *delta* expression expires once SM cells start to ingress.

All signaling related inputs into the NSM are temporary. This applies to Delta signaling as well as Nodal/Not. Delta signaling in the SM stops once these cells start to ingress between 18 and 19 hpf (see Chapter 4). As all NSM gene expression depends on Delta, this implies that at this point the transcription factors active in the NSM are sufficient for maintenance of their own expression. On the aboral side this is explained by the feedback of the GataE target Six1/2 back into *gcm*. The late module that drives *gcm* expression after 24 hpf contains a Six1/2 binding site that is critical for its correct expression (AR and EHD, to be discussed elsewhere). Thus, these three genes have locked down the regulatory state of the aboral NSM and are independent of Delta.

Within the oral NSM the situation is less clear. It is certain that neither Delta, nor continued *not* expression are required; *not* expression is becoming successively restricted to the oral ectoderm. Our attempt to unravel the connections between oral NSM genes was only partially successful. We reasoned that it must be (one of) the first genes activated in this territory that drives expression of the later ones. The first genes to become active specifically in the NSM are *prox1*, *gataC*, and *ese*. GataC appears to be a regulator of morphogenetic processes instead of specification. Ese has an activating input into *prox1*, and may repress the aboral NSM gene *six1/2* but effects are relatively weak. Prox1 knockdown failed because it causes abortive development already at the blastula stage before it is expressed in the oral NSM probably because the MASO blocks translation of the maternally deposited *prox1* transcript. In an unfortunate way this experiment is a complete success, as it demonstrates that Prox1 has important functions for the cleavage-stage embryo. It may thus be similarly important for specification of oral mesoderm once it is activated there. As Prox1 is the first gene specifically expressed in the oral NSM, it is essential to study its function to gain a complete understanding of oral NSM specification.

By mesenchyme blastula stage, the two regions of the NSM have been specified and are non-overlapping and mutually exclusive. Our survey of NSM genes revealed that there is no pan-mesodermal regulatory state: All regulatory genes expressed in the NSM are either oral or aboral. Only one gene appears to be to be spread out across the entire NSM: *delta*. As mentioned above and discussed elsewhere, this is a function of its unusual regulation: it is activated by ubiquitously expressed Runx1 and repressed by HesC everywhere except in the NSM and part of the apical domain (Revilla-i-Domingo et al., 2007; Robertson et al., 2008; Smith and Davidson, 2008).

In pigment cell precursors we observe specific expression of four transcription factors (*gcm*, *gataE*, *z166*, *six1/2* [plus its co-factor *eya*]). However, it is not certain that the zinc finger gene *z166* actually has transcriptional regulatory capacity and its function is unknown (Materna et al., 2006). In contrast, the oral NSM expresses at least nine transcription factors (*prox1*, *ese*, *gatac*, *scl*, *shr2*, *e2f3*, *erg*, *hex*, *ets1/2*) (*e2f3*, EL and EHD, to be discussed elsewhere). We did not examine two genes that have expression patterns similar to *erg*, *hex*, and *ets1/2*: *z48* and *foxN2/3*. These genes are also first expressed in SM, and in NSM later; it is thus reasonable to

expect them to follow the pattern and be restricted to oral NSM. The regulatory gene *lmo2*, although not DNA binding, was reported elsewhere as specifically expressed in the oral NSM (Duboc et al., 2010). This could bring the number of regulatory genes in the oral NSM up to 12 and may mean that the regulatory landscape of oral NSM cells is more complex than that of pigment cells.

It has been noted before that the regulatory network underlying pigment cell specification is rather shallow (Calestani et al., 2003; Davidson, 2010). Here we incorporate new genes in the pigment cell GRN that can explain the lockdown of its regulatory state. However, these genes do not add any new depth to the pigment cell regulatory network: Gcm and GataE, the two most upstream genes, are direct inputs into genes of the differentiation gene battery. In contrast, there is a striking absence of any differentiation genes specifically expressed in the oral NSM. Other than the regulatory genes discussed in this work, no markers of blastocoelar cells, the derivatives of the oral NSM, have been identified that are expressed before ingression of the blastocoelar cell into the blastocoel (Shoguchi et al., 2002), although such genes may await discovery.

The shallowness of the regulatory architecture has been used in the past to support the idea that pigment cells are a relatively recent invention (Davidson and Erwin, 2006; Ransick and Davidson, 2006). The high number of regulatory genes expressed in this cell type, plus the apparent lack of expression of differentiation genes indicates a depth that is clearly different than found in the pigment cell subnetwork. Thus the argument can be made in reverse: The regulatory program underlying oral NSM specification is the more ancestral part in sea urchins. In fact, at least as far as expression patterns are concerned, the oral NSM bears resemblance to starfish mesoderm, a distantly related echinoderm. The starfish embryo does not specify skeletogenic mesoderm, as it makes no larval skeleton. It also does not show any subdivision into oral and aboral segments and does not develop pigment cells. Conspicuously absent from starfish mesoderm is expression of *gcm* and *gataE* (Hinman and Davidson, 2007). Instead it expresses many of the same transcription factors as the oral NSM in sea urchins (e.g., *gataC*, *erg*, *ets1/2*) and gives rise to cells with blastocoelar features (Hinman and Davidson, 2007). It will be interesting to unravel the connections among oral NSM to better understand the evolutionary origin of the different kinds of mesodermal cell types in echinoderms.

References

Calestani, C., Rast, J.P., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. Development 130, 4587–4596.

Calestani, C., Rogers, D.J., 2010. Cis-regulatory analysis of the sea urchin pigment cell gene polyketide synthase. Dev Biol 340, 249–255.

Davidson, E., 1986. Gene Activity in Early Development, third edition. Academic Press, New York.

Davidson, E.H., 2010. Emerging properties of animal gene regulatory networks. Nature 468, 911–920.

Davidson, E.H., Erwin, D.H., 2006. Gene regulatory networks and the evolution of animal body plans. Science 311, 796–800.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J.C., Rust, A.G., Pan, Z., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo, Dev Biol 246, 162–190.

Duboc, V., Lapraz, F., Besnardeau, L., Lepage, T., 2008. Lefty acts as an essential modulator of Nodal activity during sea urchin oral-aboral axis formation. Dev Biol 320, 49–59.

Duboc, V., Lapraz, F., Saudemont, A., Bessodes, N., Mekpoh, F., Haillot, E., Quirin, M., Lepage, T., 2010. Nodal and BMP2/4 pattern the mesoderm and endoderm during development of the sea urchin embryo. Development 137, 223–235.

Duboc, V., Röttinger, E., Besnardeau, L., Lepage, T., 2004. Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo. Dev Cell 6,397–410.

Geiss, G., Bumgarner, R., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D., Fell, H., Ferree, S., George, R., Grogan, T., James, J., Maysuria, M., Mitton, J., Oliveri, P., Osborn, J., Peng, T.,

Ratcliffe, A., Webster, P., Davidson, E., Hood, L., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26, 317–325.

Hinman, V.F., Davidson, E.H., 2007. Evolutionary plasticity of developmental gene regulatory network architecture. Proc Natl Acad Sci USA 104, 19404–19409.

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

Hughes, J.N., Dodge, N., Rathjen, P.D., Rathjen, J., 2009. A novel role for gamma-secretase in the formation of primitive streak-like intermediates from ES cells in culture. Stem Cells 27, 2941–2951.

Lee, P.Y., 2007. Function and regulation of the *Strongylocentrotus* GataE gene. PhD thesis Caltech.

Lee, P.Y., Davidson, E.H., 2004. Expression of Spgatae, the *Strongylocentrotus purpuratus* ortholog of vertebrate GATA4/5/6 factors. Gene Expr Patterns 5, 161–165.

Lee, P.Y., Nam, J., Davidson, E.H., 2007. Exclusive developmental functions of gatae cis-regulatory modules in the *Strongylocentrorus purpuratus* embryo. Dev Biol 307, 434–445.

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

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

Materna, S.C., Oliveri, P., 2008. A protocol for unraveling gene regulatory networks. Nat Protoc 3, 1876–1887.

Nam, J., Dong, P., Tarpine, R., Istrail, S., Davidson, E.H., 2010. Functional cis-regulatory genomics for systems biology. Proc Natl Acad Sci U S A 107, 3930–3935.

Peter, I.S., Davidson, E.H., 2011. A gene regulatory network controlling the embryonic specification of endoderm. Nature 474, 635–639.

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

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

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization, Dev Biol 246, 132–147.

Revilla-i-Domingo, R., Oliveri, P., Davidson, E.H., 2007. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. Proc Natl Acad Sci USA 104, 12383–12388.

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

Robertson, A.J., Coluccio, A., Knowlton, P., Dickey-Sims, C., Coffman, J.A., 2008. Runx expression is mitogenic and mutually linked to Wnt activity in blastula-stage sea urchin embryos. PLoS One 3, e3770.

Sherwood, D.R., McClay, D.R., 1999. LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. Development 126, 1703–1713.

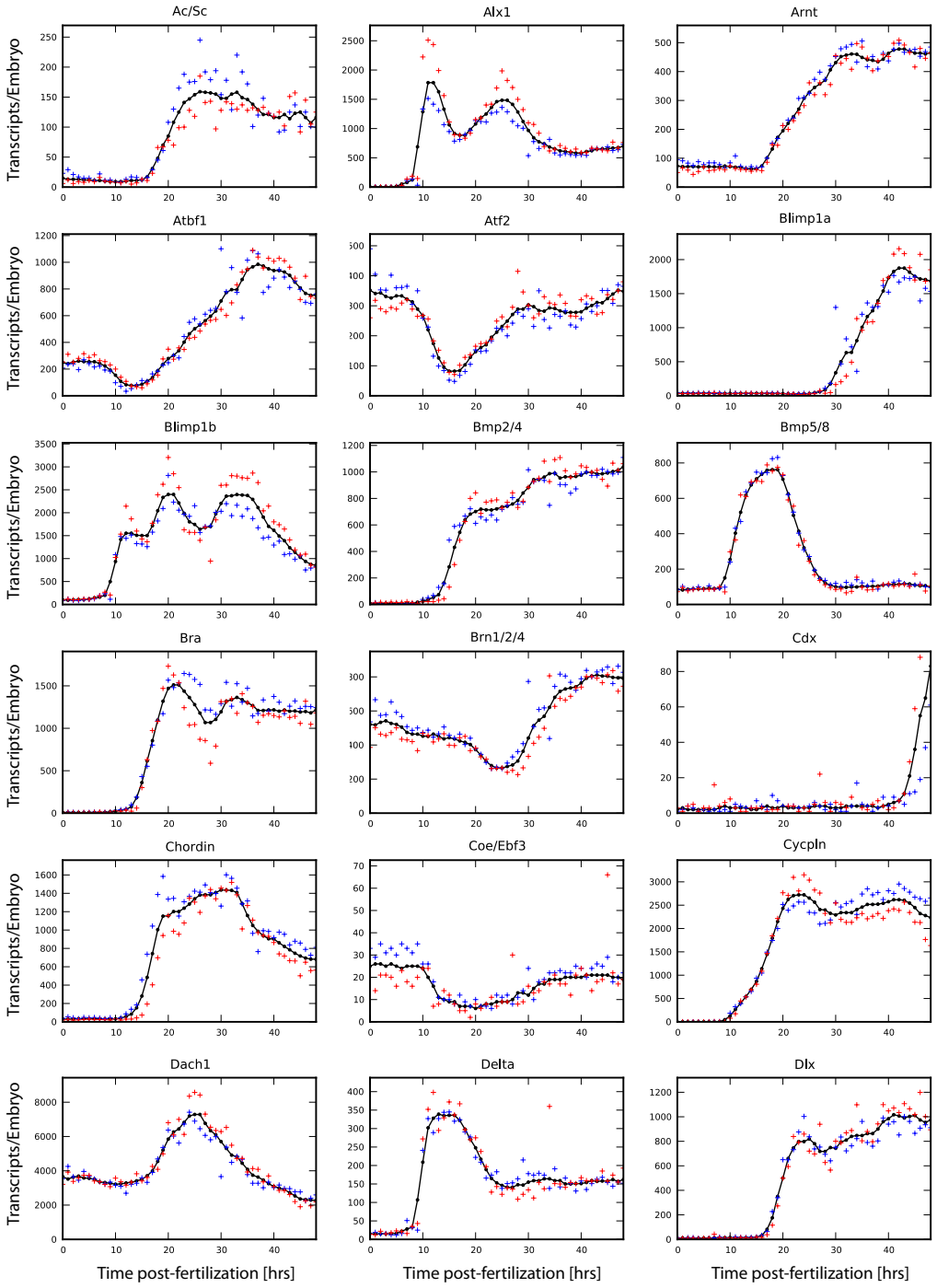
Shoguchi, E., Tokuoka, M., Kominami, T., 2002. In situ screening for genes expressed preferentially in secondary mesenchyme cells of sea urchin embryos. Dev Genes Evol 212, 407–418.

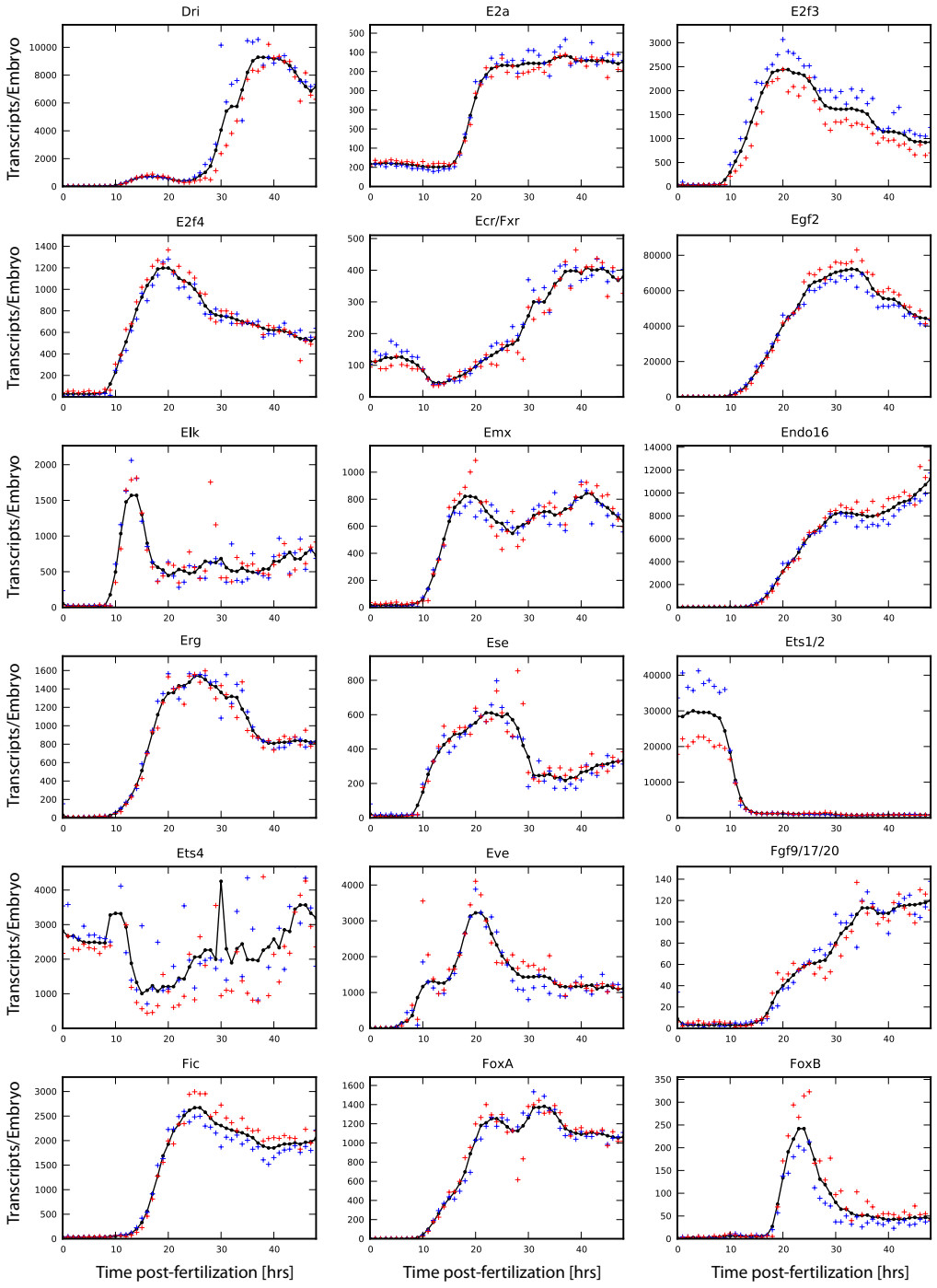
Smith, J., Davidson, E.H., 2008. From the Cover: Gene Networks In Development and Evolution Special Feature Sackler Colloquium: Gene regulatory network subcircuit controlling a dynamic spatial pattern of signaling in the sea urchin embryo. Proc Natl Acad Sci USA 105, 20089–20094.

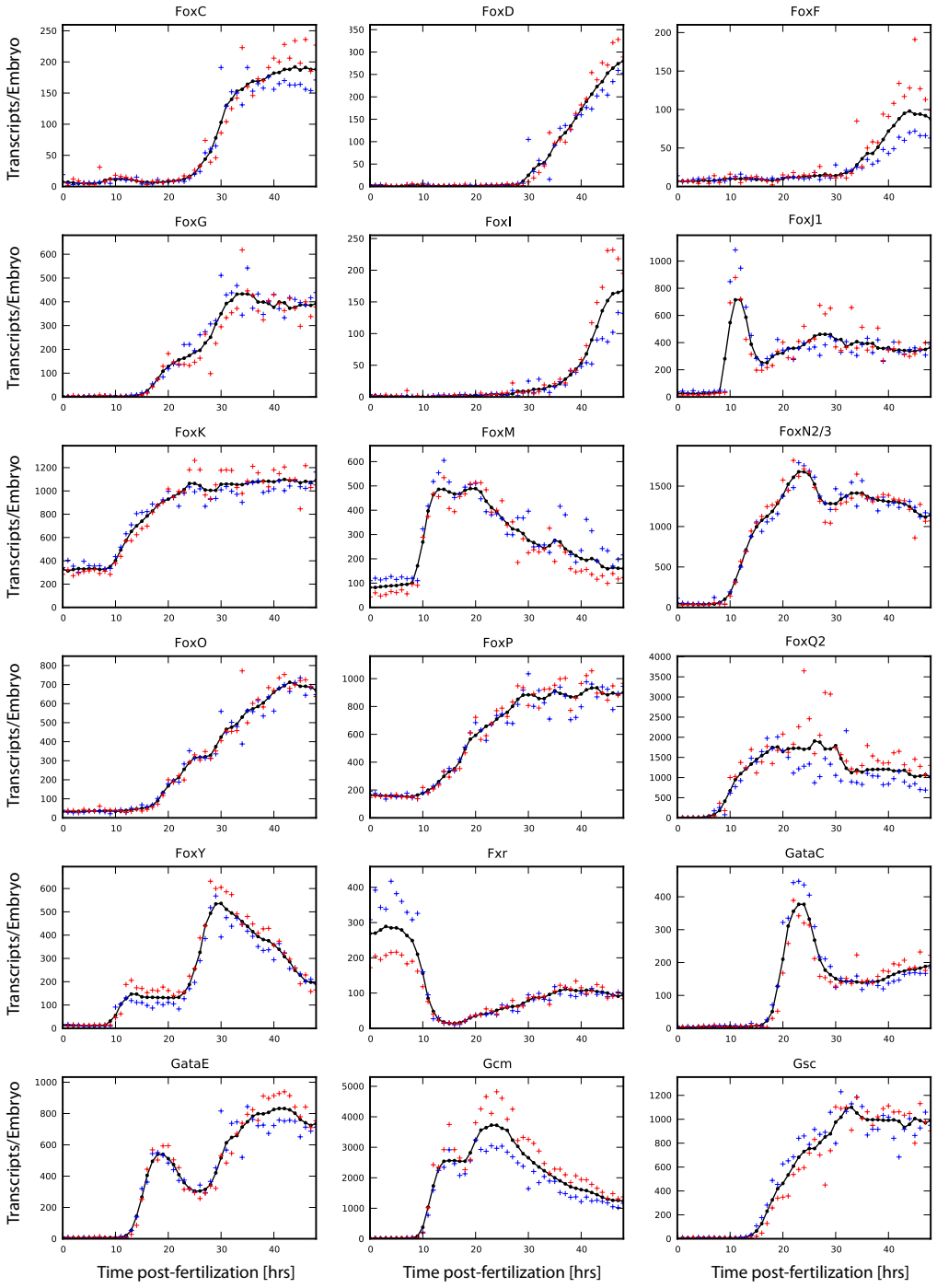
Su, Y.H., Li, E., Geiss, G.K., Longabaugh, W.J., Kramer, A., Davidson, E.H., 2009. A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo. Dev Biol 329, 410–421.

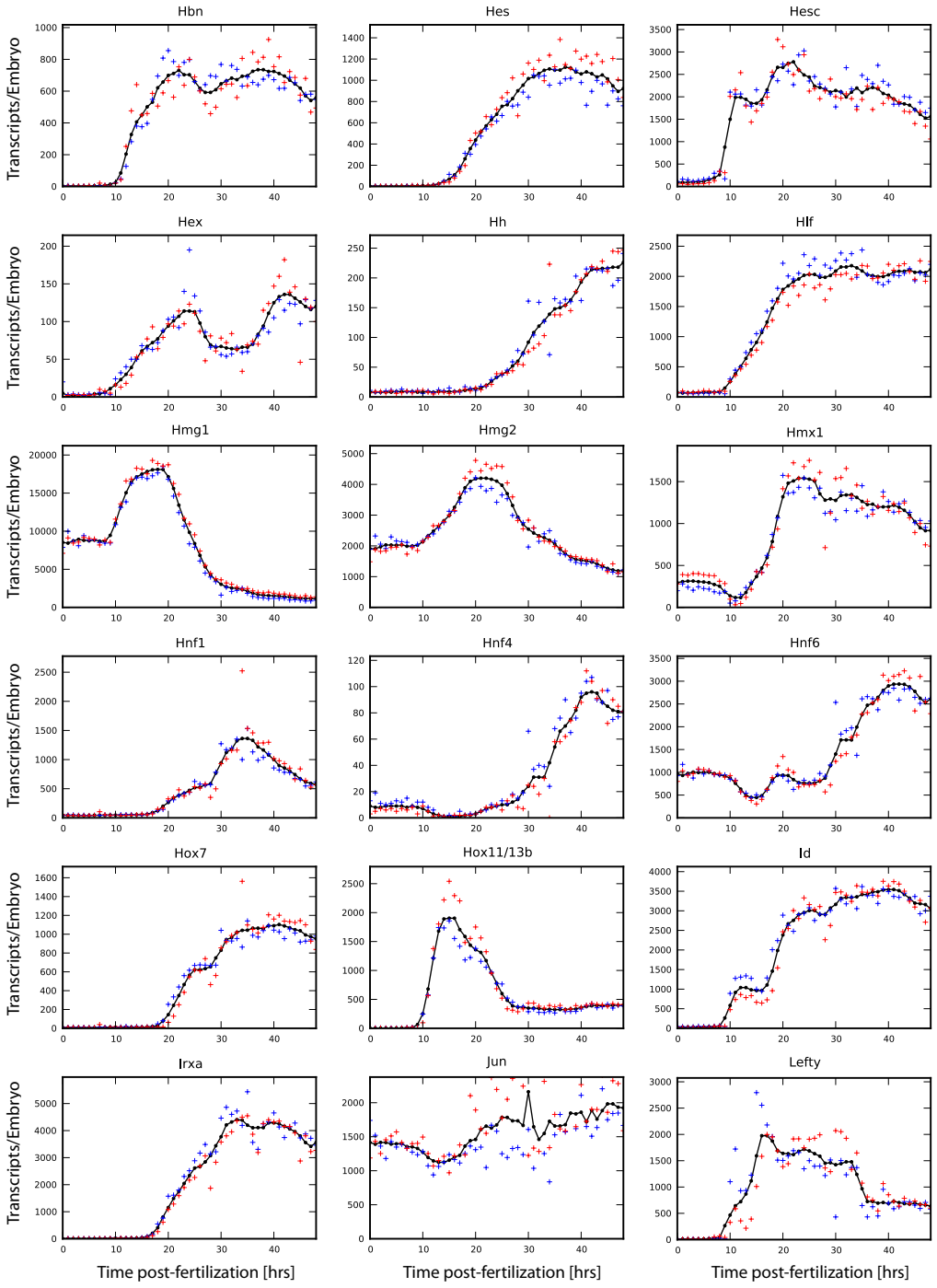
Sweet, H.C., Gehring, M., Ettensohn, C.A., 2002. LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties, Development 129, 1945-1955.

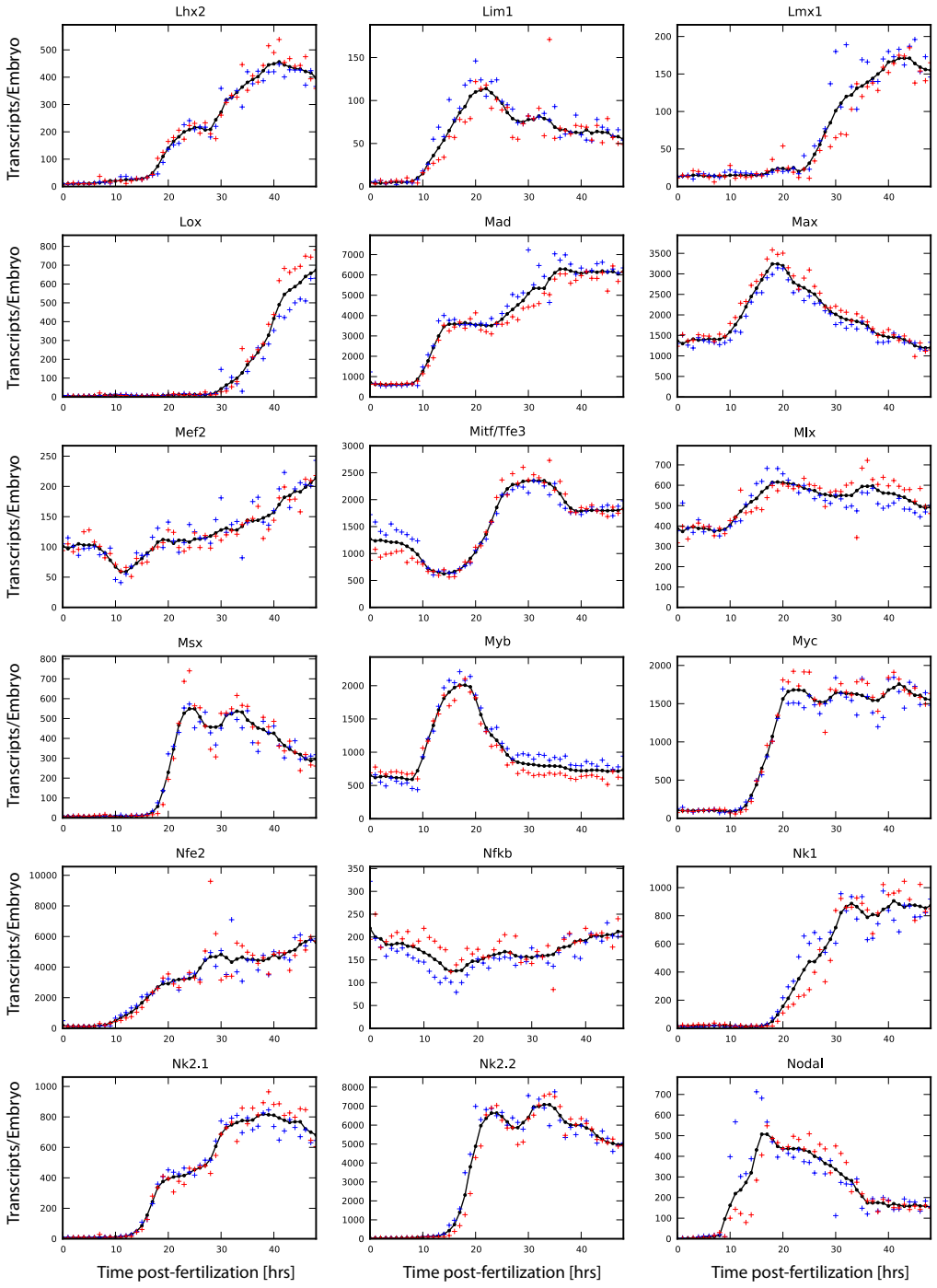
APPENDIX A — HIGH DENSITY TIMECOURSE DATA

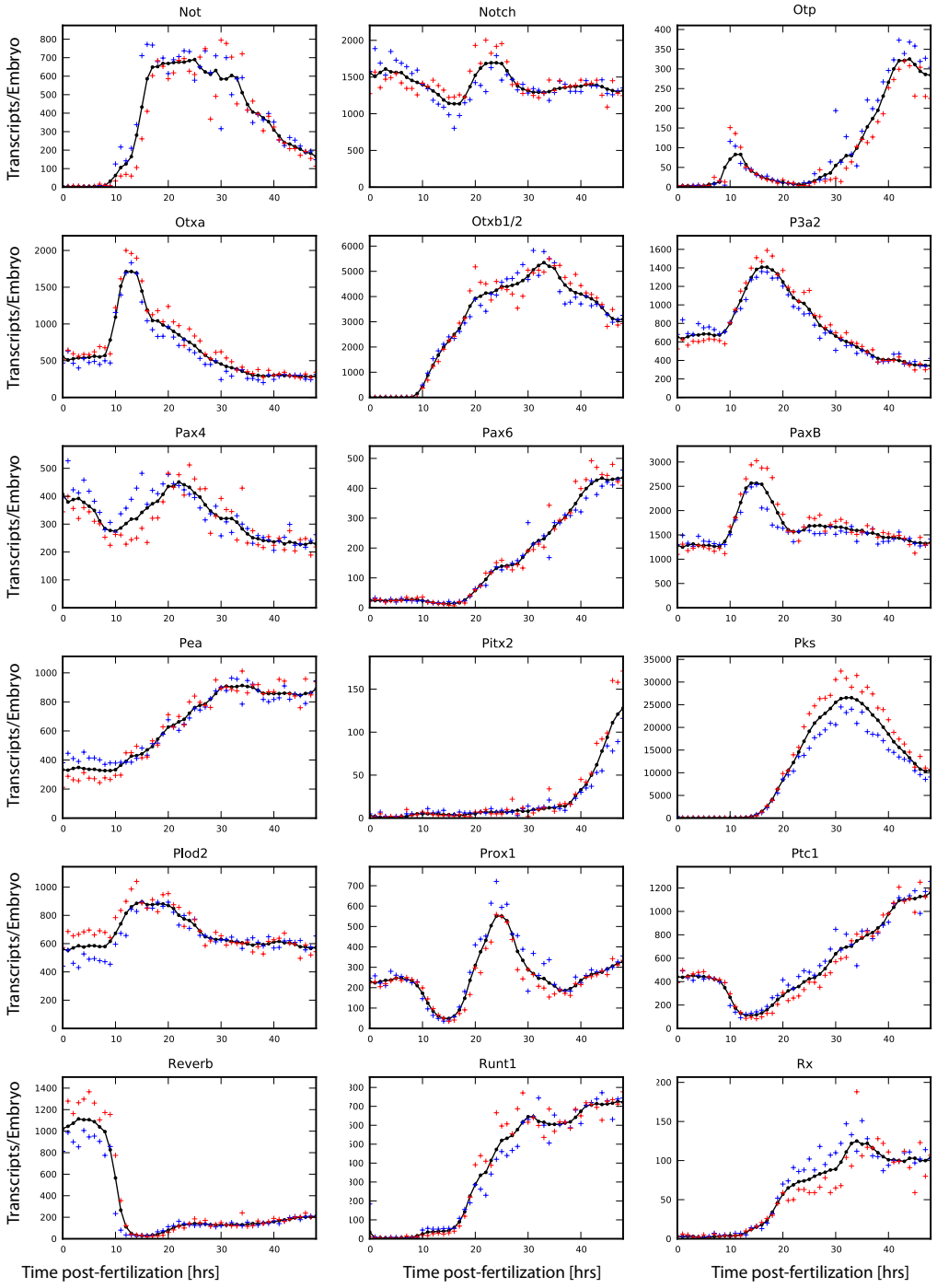


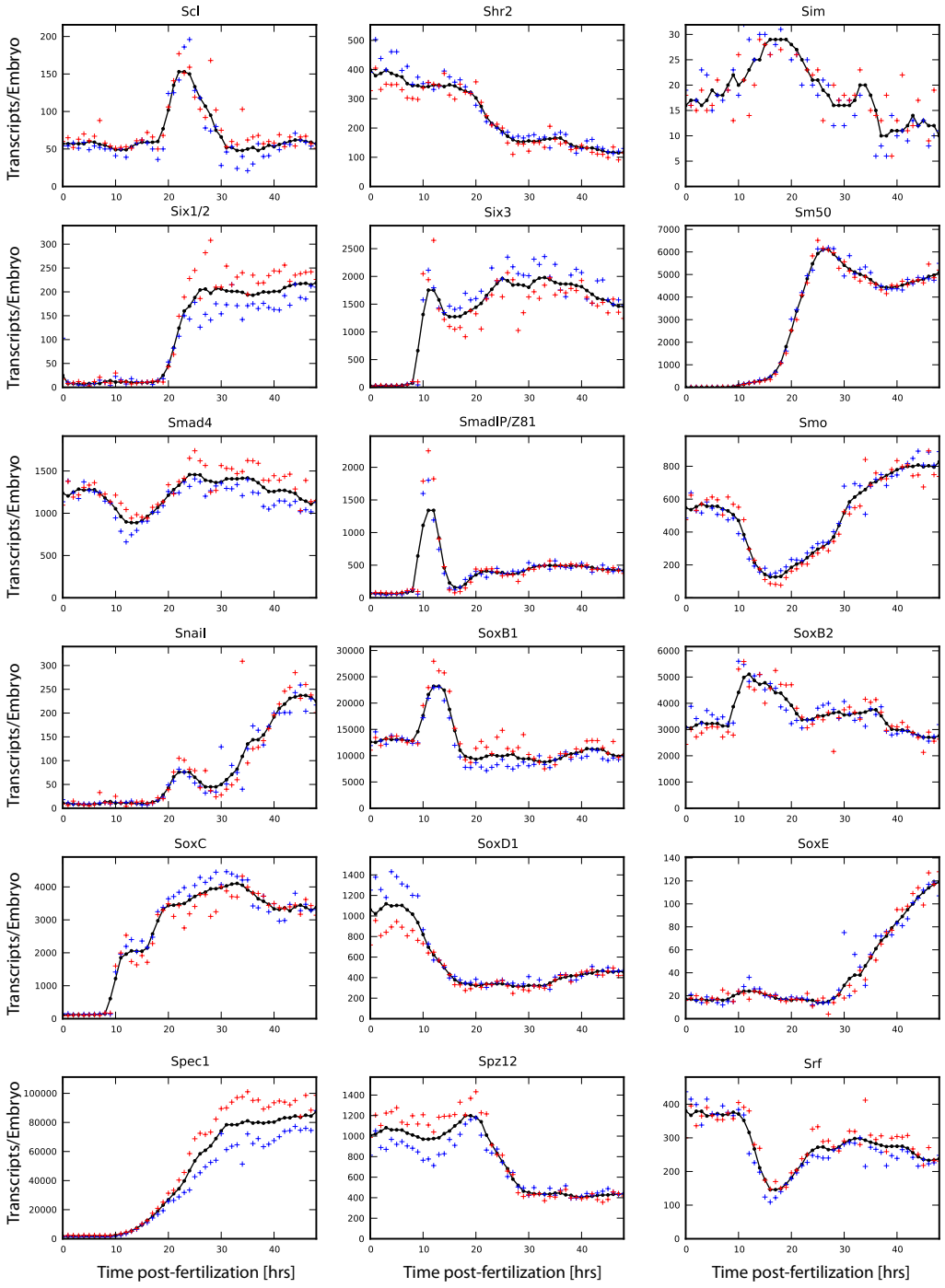

Dri
E2a
E2f3
E2f4
Ecr/Fxr
Egf2
Elk
Emx
Endo16
Erg
Ese
Ets1/2
Ets4
Eve
Fgf9/17/20
Fic
FoxA
FoxB
Transcripts/Embryo
Time post-fertilization [hrs]

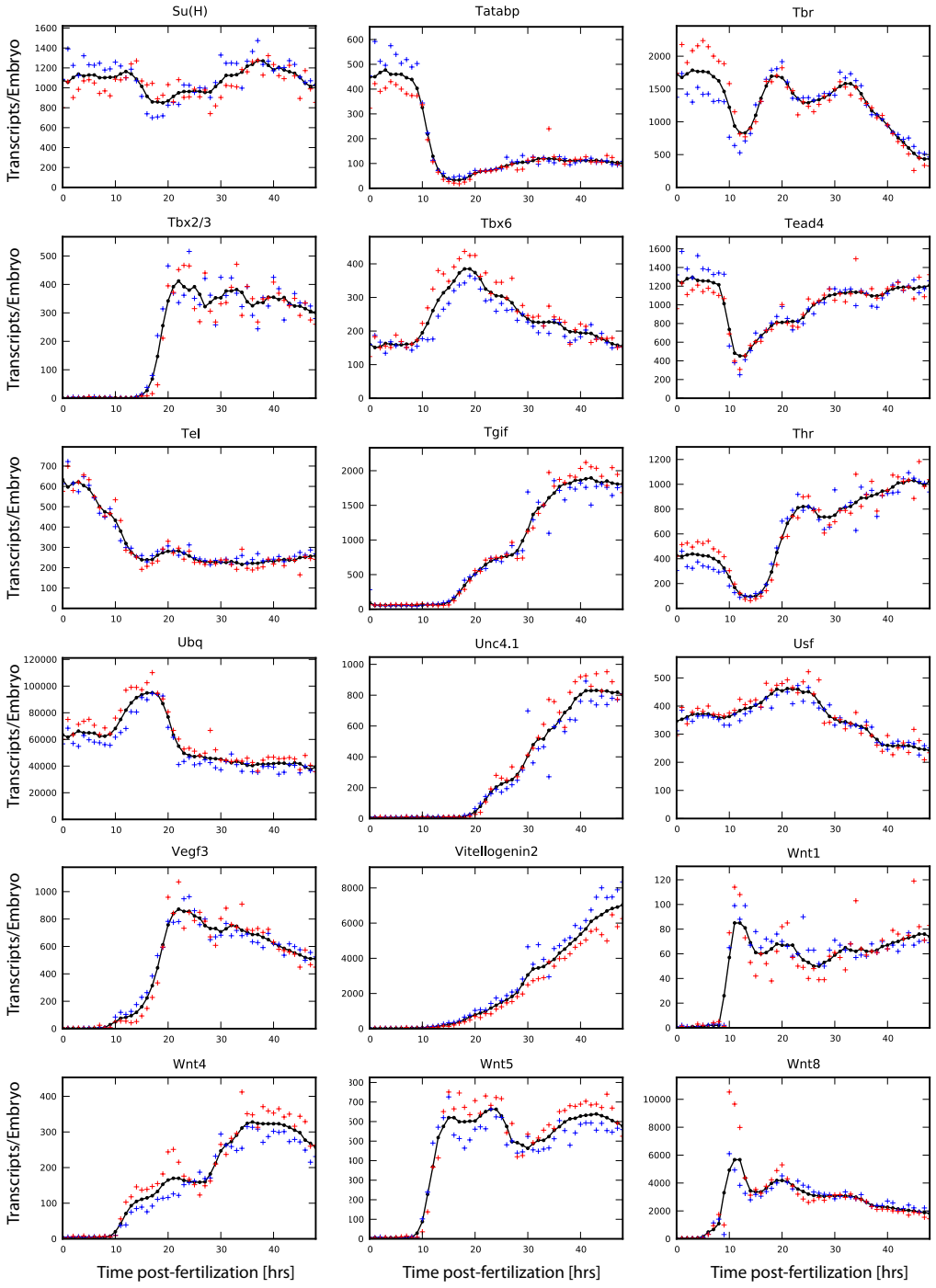

FoxC
FoxD
FoxF
FoxG
FoxI
FoxJ1
FoxK
FoxM
FoxN2/3
FoxO
FoxP
FoxQ2
FoxY
Fxr
GataC
GataE
Gcm
Gsc
Transcripts/Embryo
Time post-fertilization [hrs]

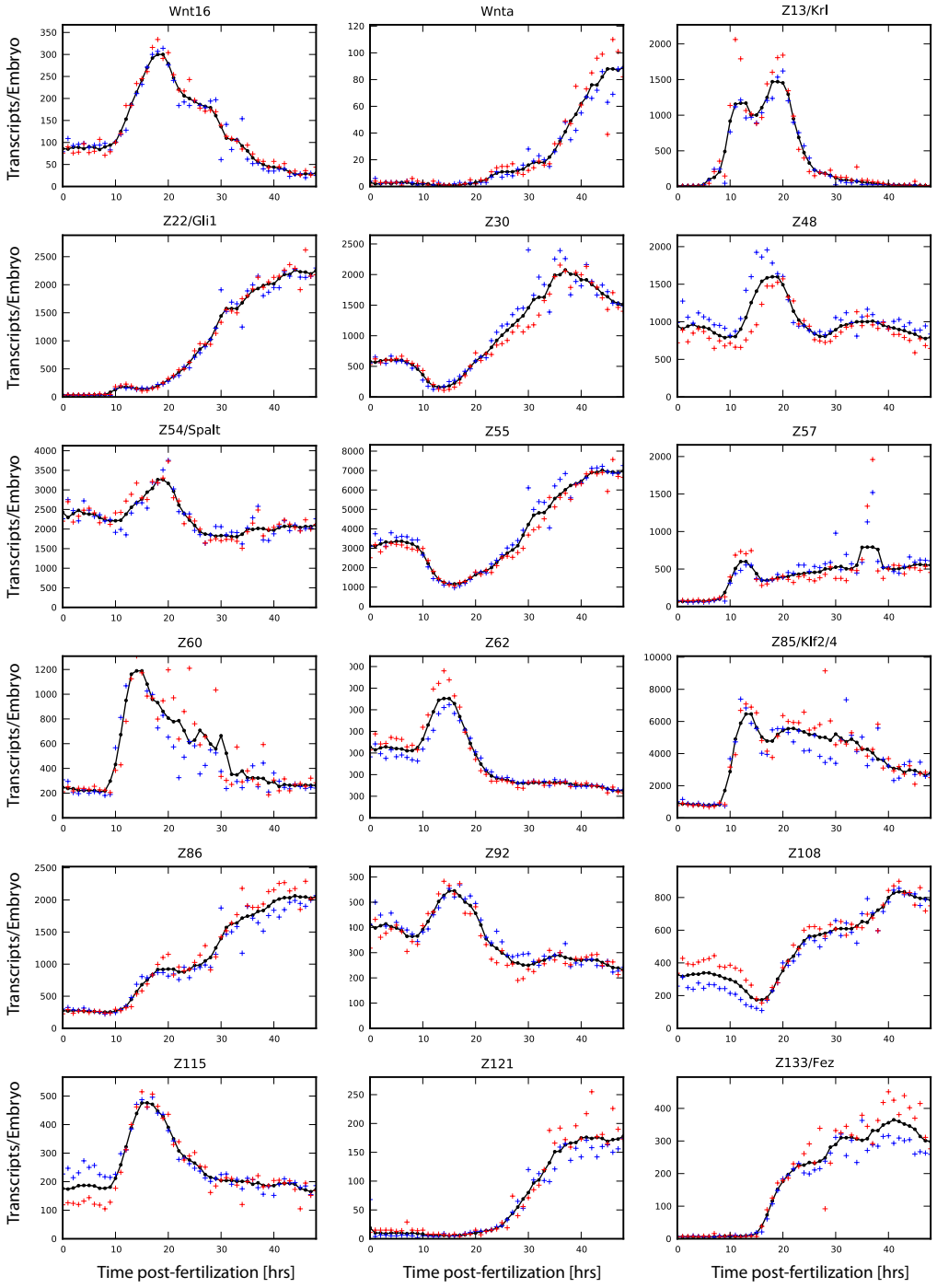

Hbn
Hes
Hesc
Hex
Hh
Hlf
Hmg1
Hmg2
Hmx1
Hnf1
Hnf4
Hnf6
Hox7
Hox11/13b
Id
Irxa
Jun
Lefty
Transcripts/Embryo
Time post-fertilization [hrs]

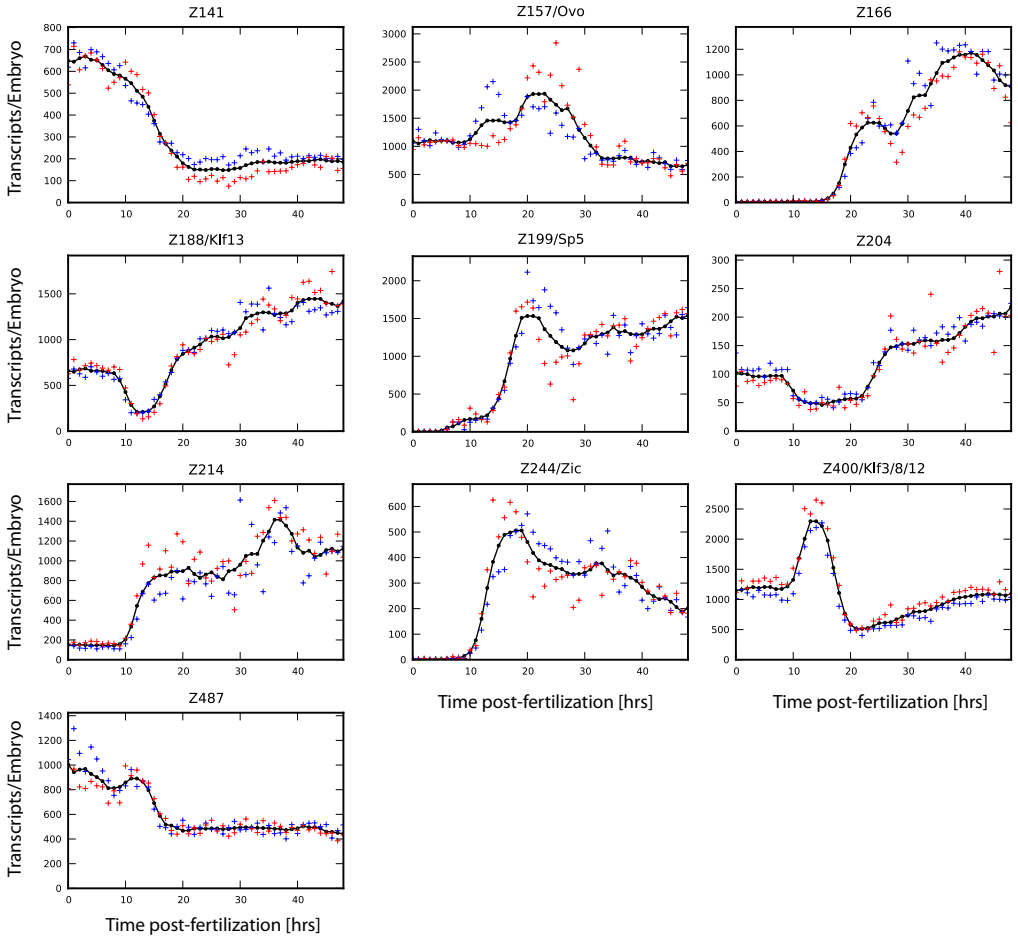

Lhx2
Lim1
Lmx1
Lox
Mad
Max
Mef2
Mitf/Tfe3
MIx
Msx
Myb
Myc
Nfe2
Nfkb
Nk1
Nk2.1
Nk2.2
Nodal
Transcripts/Embryo
Time post-fertilization [hrs]


Not
Notch
Otp
Otxa
Otxb1/2
P3a2
Pax4
Pax6
PaxB
Pea
Pitx2
Pks
Plod2
Prox1
Ptc1
Reverb
Runt1
Rx
Transcripts/Embryo
Time post-fertilization [hrs]


Scl
Shr2
Sim
Six1/2
Six3
Sm50
Smad4
SmadIP/Z81
Smo
Snail
SoxB1
SoxB2
SoxC
SoxD1
SoxE
Spec1
Spz12
Srf
Transcripts/Embryo
Time post-fertilization [hrs]


Su(H)
Tatabp
Tbr
Tbx2/3
Tbx6
Tead4
Tel
Tgif
Thr
Ubq
Unc4.1
Usf
Vegf3
Vitellogenin2
Wnt1
Wnt4
Wnt5
Wnt8
Transcripts/Embryo
Time post-fertilization [hrs]


Wnt16
Wnta
Z13/Krl
Z22/Gli1
Z30
Z48
Z54/Spalt
Z55
Z57
Z60
Z62
Z85/Klf2/4
Z86
Z92
Z108
Z115
Z121
Z133/Fez
Transcripts/Embryo
Time post-fertilization [hrs]


Z141
Z157/Ovo
Z166
Z188/Klf13
Z199/Sp5
Z204
Z214
Z244/Zic
Z400/Klf3/8/12
Z487
Transcripts/Embryo
Time post-fertilization [hrs]

APPENDIX B — QPCR PRIMER SEQUENCES

Gene	Accession	Q-fwd	Q-rev
Ap4	SPU_003179	TGCTCAATTTCTCTCCTTCGT	CGAGAGATAGCCAATAGCAATG
Apobec	SPU_011837	ACCCAGTTTCACCCTCCTCT	AGGCACTCAGCTGCAAAGTT
Atf6	SPU_007749	GGCAGCACACTTTCTTCACTA	CTTTGGAGCCAGGGGTAACT
Awh	SPU_018954	GATCCGCACCACTTTCACC	CCGTGAGTTTTGAAACCAGA
Bpnt	SPU_021171	TGAATGGGGAGTCAGTAGGG	GATCACCCTTCTGGGACTCTG
Capk	SPU_022038	CCAAGTACGCAGGAGGAAGA	GAGAGCATCGGCTATTGTCA
Dopt	SPU_001152	CGAGTTCGCGTACAGCATAG	GAATCCTTCGGGAAACTGCT
Eya	SPU_013869	AGCATCCGATGACAATGGAC	TCCTCCTATACCGGTACGC
Fmo2	SPU_014947	GTCGGTGGACAATCACCTCT	ACGAGGACATTCTTGCCATT
Gelsolin	SPU_014715	CTCCATCGACGAGAGGAGAA	CCTTCTGCTACGACCGAAAC
Hmg1	SPU_027981	GGACCGAGACAGTTCAAAGC	CTTCTCCTCCAGAGCCTTCC
Kakapo	SPU_003256	GTGGCATTTATGAGCGGTCT	CGGCCCAGTACTTCAAGGAGA
Nlk	SPU_010846	GATATGAAGTACGCCTGTGACG	AGTGGACAGCTTCATGAGTGG
Numb	SPU_015466	ACAAGCTGTGTGGAGTCAAGG	TCTTCTTGGGATAGCATGTGG
Papss	SPU_016157	AGTTGGAGGCGTTCACACTC	CCACCAACTAGCCAATCACC
Phb1	SPU_008112	GCAGCCGTACCATTTACACC	AACCTGGACTCTGGCTTCAG
Sin3a	SPU_015425	ACAACACGCTGTTGGATCTG	TATCCAGAGTGAAGGCCATGT
Smad6	SPU_001998	AAAAATTCGCCAGAAGATCG	CTGTGAACGTCCTGGAGTGA
Sult	SPU_006187	AATTCATGCCAGAGCCATTG	CCGAGAACTCGACCTTCAAC
Ubq	SPU_021496	CACAGGCAAGACCATCACAC	GAGAGAGTGCGACCATCCTC
Xbp1	SPU_008703	TCAGTGGTCGTTTTGGATCA	TCGTCAGACTCCACATCAGC
Z220	SPU_015751	CTGAACTTCTTGCCGCAAAT	CAACAGACAACACAAATGCCA
Z258	SPU_011687	TGACAGTGCCTCATCAGACC	CTTCCATGTCTGCCACCTG
Z265	SPU_016149	CAGCCTCAACAAAAAGCCTC	AAAAGCCTTGTCACACTGGG
Z316	SPU_009053	TTCTCATATGGCAACCCACA	ATAGGGTTTCTCGCCCGTAT
Z400	SPU_017470	AAAAGCCGTACAAATGCCAA	GATCTTGAGAACGCCCTGTC
Z472	SPU_020726	GCACAGTGGAGAAAAGCCTC	ACTGGGAACACTTGAAAGGC
Z475	SPU_002140	CAAGGTGTTTCCCACAACAG	AGTGGAGATGGGGATGACTG

APPENDIX C — WHOLE MOUNT IN SITU PRIMERS / PROBE REFERENCE

Gene	Forward Primer	Reverse Primer
Alx1	Damle and Davidson, 2012	
Apobec	GACTTTGATGGAGCGTACCG	GCACAGGACTGTCTGATTG
Bpnt	ATGGAGGAGATGGAGAGAGG	GTTCATGCGCTTGATGTCC
Delta	Oliveri et al., 2002	
Erg	CGTGCACATTTCCTTTTCAA	CCCTTCCAACAAGGCTATCA
Ese	CCATCATGCACACCATCAATC	ATAGGTCATCTGCGGGTTGTT
Ets1/2	CTCGGAATCCCAAAAGATCC	TGTATCCCAGAAGGCTCTGC
FoxA	Oliveri et al., 2006	
FoxQ2	Tu et al., 2006	
FoxY	ACTGGATGATCAACCCAAGC	AGAAACCTTTGGTGGTGTCG
GataC	GCTGGAGTAGCAAGCAGTCAAG	TGTTACCCATTACCCCAGGATG
GataE	TCTTACCCATCAGCCAGGAC	TCGAGTGGAACAATGGAACA
Gcm	Ransick et al., 2002	
HesC	Revilla-i-Domingo et al., 2007	
Hex	CAGGGAAACCTTTCTTGTGG	CCTGGCTGTGACACCTAAGC
Nodal	CACAAAGTGTGTTTGTGCAAG	GTCGATGAAATTGAAAATATCATGA
Not	ACGATTTGGAGGGTTTTAAGC	GCAACTTACCCGTCATCACC
Pks	Calestani et al., 2003	
Prox1	GGAGACATTGACGTTGTTTGG	GGCTTCGGTCAGTCAGAAAGG
Runx	GCCCAGATCATACCCTCTCA	TTTAAGGAGCCAGGATGGTG
Scl	TGCTGGTGATGAGTACCACGGC	GCTTCACACCAGTCACTTCGTTG
Shr2	AGAAGTGCCTCGACATGG	ACCTCAAGACCATTGTAC
Six1/2	TGAAACACCGTCAAAACAAGG	AACTTGTGTCGTGGTCAGTCC
Z166	CATCTCCAGCAGCGGTTATC	GGTGTGTTTCTTCATGTCCG

Calestani, C., Rast, J., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. Development 130, 4587–4596.

Damle, S., Davidson, E.H., 2011. Precise cis-regulatory control of spatial and temporal expression of the alx-1 gene in the skeletogenic lineage of S. purpuratus, Dev. Biol.357, 505–517.

Oliveri, P., Carrick D.M., Davidson, E.H., 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev. Biol. 246, 209–228.

Oliveri, P., Walton, K.D., Davidson, E.H., McClay, D.R., 2006. Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo. Development 133, 4173–4181.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246, 132–147.

Revilla-i-Domingo, R., Oliveri, P., Davidson, E.H., 2007. A missing link in the sea urchin embryo gene regulatory network: hesC and the double-negative specification of micromeres. Proc. Natl. Acad. Sci. USA 104, 12383–12388.

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

APPENDIX D — NANOSTRING PROBE INFORMATION

The genes listed in this table were part of at least one of three codeset version used in the preceding chapters.

Gene	Accession	Probe Sequence
Ac/Sc	SPU_028148	GTCGCACGCCGTAACGAACGGGAGCGAAACCGCGTCAAACTCGTCAACCA TGGATTCGCCAATTTACGCCAGCAGCTCCCTAACGGCGCCAATAATAAGA
Activinb	EU526314.1	AGCTCTGGTTGTACTGGAACGGTGAAAATTCATCCGGCTCTTCCACAAAT CAAACCTTAACCATCTCACAGATATCAGGACATGGGTTGAGAACTAGTAG
Alx1	NM_214644.1	AAGAAGAGGAGAAATAGGACGACGTTCACCAGCTATCAACTTGAGGAGAT GGAGAAGGTATTTCAAAGAACGCACTATCCCGATGTGTACTGCAGAGAAC
Arnt	SPU_000129	CAAGCCAAAAGTCGAGACTGGGTGTGGCTTCGAACCAGCTGCTTCAGCTT TCAGAATCCTTACACCGATGAAGTGGAGTACATCGTCTGCACCAATACAG
Atbf1/z30	SPU_017348	GAGGAGAGGAGATTTGATGAAAGAAATGAAGTGAAGAGTTGGGAGGAGCG AAATGAGAGAGAGATGCAGTCACGGATATCTAGCCATAAAGGTAGTGAGA
Atf2	SPU_026905	GATATATGTTCATAAGAGGCAGCGTGTTTCGGAGGAGGACGATAACGAGA ATAAACGACAAAAGTTCCTGGAGAGGAATCGAGCTGCTGCTAGTCGGTGT
Blimp1a	DQ225099	TACCACCACAGAAAAGAGAAAAAGCGAAAAAAAGAAGTCTAAAGAAATTC CCCGATTGAAACCGCAAGCTATTTTTTCTACGCTTCAACGAGGTCTTCGA
Blimp1b	DQ177152	GTTTGTTGTGATTTTGTACCGCGGTGTTTTTCAAGCGAAAGGGAGAAATG GTTATTGTTGAACACCCTCGCCCGCTCGCTATGCGGGATCTCTACTACCT
Bmp2/4	AF119713.1	AGGAGGAAGTAATGACGACGGCCGAACTGAGATTGTTCCGCAAAGACCTC GATGAGCACCACATCGTAAAACGACACGCACTTCACGACAGAGAAAGTCT
Bmp5/8	NM_214655.1	AAGGAGATAAAAGAAGAGTGACAAGTGCCTTACGAATAGGGGAGAATCCT TTGAAATGTTCCTGGTTGCGTGAACAGGATAGGTGAGGAGTTTAGCCTCA
Bra	SPU_013015	CCCTCAGTCACCGCTCCCAACCGGCTTGTTCCGTAACCCTCATCCAACCT CGTCACACCAGCACAACCTCGCATCAACGGCGCACGGAATGGCTCCGGTA
Brn1/2/4	SPU_016443	GGGTGTTCTGGGACCAGGTGGTGGTCAGTTACCTAGTCATAATGGTTCCG AAACGGTGATCGAAGACGATGCACCGTCATCGGATGATCTCGAGCAGTTT
Cdx	SPU_024715	ACAGCATTCATCAAGCCCTCATCATACCACAGTCGACACCAAACCCAAAC TCCCGATCATAACAACGATGACCACGGTCTCCGGCGATGTCATCAAGACG
Chordin	NM_001110245.1	GATTGGAGTCCCTAACCGCCTGATTGTGAGATCACCAAAGATGGTGTGAA GATGATGCCGACATCCTCAATGTCAGAGGAATGGAAGGGGTCAGCTGTTG
Coe/ebf3	SPU_004702	GGCCGGAGAGCGAGAAGACTTGACCCATCTGAAGGAGCTACGCCGTGCAT CAAGGCAATCAGCCCTAGCGAAGGATGGACAACAGGGGGCGCCACTGTCA
Cycpln	NM_001033647.1	CAAGACCAAGTGGCTGAACGGTGCCCATGTTGTCTACGGTAAAGTCCTGG ATGGCCTTGATGTTCTGGCCACCATTGAGAACTCTGCTACAGATGAGAAC
Dach1	SPU_018581	GCGGACTACACACGGTTTACACAAAACTGAAGAGACTGAGCATCACTCCG GTGGTTTGTAACGTAGAACAGGTTCGCATCCTGCGGGGACTTGGGGCGAT
Delta	NM_001032370.1	GAAAATGGAGGAACCTGTCTGCAAGATGTAATGGGAGGATTCATGTGCGA GTGTGCTGAAGGATGGATAGGTACAACATGTACGCAATCGACTGGTAAAG
Dlx	SPU_002815	TGACTGGGATGGGTGCGATGAGTCAGACGAGGCACTACGGCTATCACTTT ATGAACTCTTCCCTAGCCGGGCACCCTCCGATATACGACCATTCTGATCA
Dri	NM_214634.2	GACGTGAAACCAGCCACGCCCATATTGATCTGAACATGATGCGAGCCAAT GCCCACCTCAAGGAGATGATGGATAAGAACCGTCGATTTGTTTCTACCCG
E2a	SPU_016343	GCCAACAATGCAAGGGAGCGCATCCGCGTTCGGGACATCAACGAAGCATT CAAAGAACTGGGTAGGATGTGCCAGTTACACCTCAAGCAGGATAAGGCCC
E2F3	SPU_006753	CAGAATGGTTTGGACGACGACTTTGTTGCGCTCTCGCCACCGGCTGTGGA CGACTACCTCTTTGCCTTGAACGACAACGAAGGCATCTCTGACCTGTTCG
E2f4	SPU_028827	GCCACTGAGTAGACATGAAAAGAGCCTCGGTTTGCTCACTACGAAATTCG

		TTGGTCTACTTCAAGAAGCACCAGATGGAGTATTAGATCTTAAGCAAGCG
Ecr/fxr	NM_001129807.1	TGGAGATAAGGACCAGACAGTCGTGTCGATAGCGTATCAATCGTCAATGC AGATCAGCTGTGTTACACCATGCTGCGTGTTGACAACCTGTCACCGATCC
Egf2	NM_214531	AGCCCGAGACAGTGAGAACAGATGCTTAAGTGACACCAGCAACTGCGATG GCCATGGTATATGTCAACTGAGTACCTTTGGCCGGAATGAGCGGTACATC
Elk	NM_214627.2	ATCGTGAGAAATTCCAGTGGACGACGTCCTTACGCAGAGAAGGCGATGAG ATGTTCTCAAGACCACCCGTGACAGTGAATGTTCAGGAGGTCTACATCAC
Emx	SPU_002592	GGCACGGCGGCATCATCAAGCCTATGCCCACAGAGGCATCCTTCTTTAAC ACTCTCAGTGGAATGAAGGGACTCTACCACCCTGAGCCCATCTATCACAC
Endo16	NM_214519.1	TTCGCGGTTTTGGCCGTGGCGCGGTCAATGCCCACAGAGTTGCAATGTGA CATCGACGAGGCCCATACGAATGTTGGAACACCACGAGTAGTCGACGATG
Erf	XM_788779.1	GCGCTTCACCTACAAATTCAACTTCAGCAAGCTGATTCTAGTGAACTACC CGGGAACCGACCTCAAATATGTGCCTCCGTTCATCCCGCAATCTCGGGGT
Erg	NM_214668.1	TGTATAGGAAGCATTATCGGTAGTCTCGGACGACCAGACGCTATTTGAGA ACACATACCGTGAGGTGAATAAGTCCAACACTATCGTCTCGTCGCCACAC
Ese	SPU_024903	TCCGGGACCTTCTCAAGTCCGAATCCTACTGCCCGAAGTTTATCCGATGG GAGGATCGGGAAGCCGGAGTTTTCCGATTCGTCAACTCGGAAGCGGTCGC
Ets1/2	NM_214533.1	GCTCATCGCCACACGTACAGAAGCCGAGTAATTCCTGGTCGACGATTCGA GTTCGACCTCCAACAAGAGCCGCGCTGAGCATCAATCATGGCGTCTATGC
Ets4	NM_214534.1	TGTCCAGAAGTGGTTGTTATGCGTCGCCAATCGCTTTGAGCTAGGGGAAC TCGAGATGGGCCACTTTTACATCAATGGACCAACCCTTGCAACACTCCAG
Eve	NM_214651.1	CCGACGCCTACTCACTCCACCTCCGATCGCGAGCAGAGCTTCTACATGGT CTCGGTGCACACCCTTACGCAAGACCGTTTCCAAACCCTGCTGCCGCTCC
Fgf9/16/20	SPU_006242	GCGAGTAAGAGAGCCTCGCACGTCATCATCATCGGGTTCTTATGCGTAAC CCTGGCGGCAGGTTTATCAGACGATGGCGGGTTGCAAACGAGAAGAGAAC
Fic	SPU_000045	GTCACTGGCGCGACTACTCACGAGGGTTCGGCAATGTCAGCTCTGAGTTT TGGCTGGGCAATGAATTCCTTCACCACCTCACCATGCAGTATAACTACAT
Follistatin	XM_777947.2	CAAGGCCGTAGGCATCGCTCACGAAGGACGATGCGAAAATTTCACTTCAT GTTCTGCTTTGAGCTGTGTTCGCGGTAATCAATGTGTGATGGATCCCATC
FoxA	NM_001079542.1	CACGCGCATACACATCAGTGGAGGCTGACACTATATACTTTAGCAGATAT ATACAAACATACACTCAGCTCACGATCGAGAGAGAGAGAGGAACACAACG
FoxB	NM_214632.1	ACAGTGATGCCAAGCCGCCATATTCCTACATCTCCCTAACGGCTATGGCT ATCCAGAGCTCGCAGGAGAAGATGCTACCCTTGAGTGATATCTATAAATT
FoxC	DQ286740.1	ATCAAGGAATGGCGATGCAACCTTATCATTCAGCGGCTCCTTCCATACCT GGCGTCATGCATTCGGTGTTTCCACCAGAACAGGGCTACTATCGGCACCC
FoxD	DQ286738	AAGAGGCAGCAACCAGACTTTTTCAGGGAAGCGGGGCATTTCATGACGGG TCCATATCCTATGGGGCGACCAGCATACGGTTATGCGATGGCAGGACTTC
FoxF	DQ286741	TATGATCGGTGGCCCGATGTCGCCACACCTTGCGAACAGCGCACCGAACG CCTGCTCGCCCAACGGTGGATTCATGACTAGCTGCTCGATGGCGGGTGGG
FoxG	DQ286739	GATGAGAATGATGGTGAGAAGGAGAAGGAGAAGGAAGGGGGAAATGATGA GGAGAAGAAAGATGACAAGGACGGTAACTCGACTACGAACAGTGTGCCAA
FoxI	SPU_023894	GAATGCAGCCGCCGCATCGACTTTAAATAGTGAGGTTCCTGCATCCCATC TCTCTCCCAACGATATAGCTCATAGCAACATGGCAGCTCTGAGGTACTTC
FoxJ1	DQ286742	GACAGCATGCCAAGCAGCATCCTCAAGGAGGACTTCAGTTGGAACTCCAT CTTCGAGTCGGAGATTGAGATCGAAGGCACGCGCATTAAGACGGAGGACA
FoxK	DQ286748	GTCAGAGGCTAAACTAATGGATCAAGCTTACAGAAGAAGACGCCAAAGGG GAGTGCCGTGCTTCAGGACACCCTTTAGCGGGGTCTCATCAAGCCGCTCC
FoxM	SPU_025590	AATGCTTTGCTCCTGAAGGCTCCTTTGACATGGCTGCGATGGCAAATCTT CAAATGCAAGCTCTGGTCGGCGATCCTAGCAGAATGCCCATGTTATCAGC
FoxN2/3	DQ286744	ATGATCCAATGGACCCGCACCGACACATCAATAGCAAACCTCCTTTCTCA TTTAGCTGCCTCATATTCATGTCGATAGAGGACTGCCCTCTTAAGCGTCT
FoxO	DQ286746	AGGCAAGTCATCAAGGAGAAGAGCATCCAGCATGGACACCACAAATTCAA AGTTTGAAAGGAAGCGGGGGCGCGTGAAGAAGAAGGTCCTGGAAGAGCGT
FoxP	DQ286749	CGCCTTGATGATTTGTGGGAGAGCATCTCCAACAAGTTCCTGCTGGTGAT TGTTGCATGGTACCAAACCAGACTCCATGAGTGAAACCCTGGCAAACAGT

FoxQ2	DQ286735	CATTTCTGGGCAATCCATCCCGCTAACCTTGAAGACTTCGCCCGTGGTGA TTACAGGAGACGGCAAGCTCGTCGCCGAGCACGATCAGTCAGTTATTCTC
FoxY	AF517552.1	GCCTACCTACCCAGCTTCACAAATCTCGCCTCATCGATTGCCTTACTCCT CATTTTCCATACCATTTGTAGTGTTCTGCACTTGGAGCCATGGAAGCGCA
Fxr	SPU_011348	TGGAGATACCAGGTCCTGAGGACCTCAAGAGAGTTACGCCATGGAAAGAG GGCCAGAACTCGGGGCAAGACTCTAGCGATGACAACGGGCCCAAATCGGT
GataC	NM_214539.1	TGGGCGCTCTGGCGTAGAGAACCCGAAACTATTTCGGACTAACTTATAGA ACTACTGCGAGCTGGACTCTGGGACAGATAAATGCTTTCATTGTGATTGA
GataE	NM_001005725.1	ACCTCTCATCAAAAACCCAAGGAGATTGCAGTCTGGCTCAAGACGAGAAG GAATTACATGTGCTAACTGTCACACTAGCACTACAACACTGTGGCGTAGG
Gcm	NM_214661.1	GTTCGGGAGAAATCGTGACTGTTCGGCCAGCCACTTCGGATCGTGCTCGG AAGAAGCAAGGCGATAAGAAATGCCCAAGGACCGGTTGCGATGGCAAACT
Gsc	NM_214498.1	GTTCAAGAACCGCCGAGCTAAGTGGAGGAAGCAGAAACGGGAGCAACAAG AGGCTGCCAAACGGGCTTCCGAAGCATATAAAACCGAGTACGGGTCCAAA
Hbn	SPU_023177	AGGCGCAGCAGCGGCTCACGACGGCGCCAAGCTCTCCCATTCCCTTCCAA TCTCCATCCCGAAACTCTACATGTCGCCTGCCGAGAAAGGAAGGAGAATG
Hes	NM_001001768.1	GCTGTCAACAACACATCGCAGGTCACGCAACCGATACGCGTTCAGATCTC GCAGGCGACCTCAACGGCGTTAAGTTCATCAAACGGCATCGTGACATCAC
HesC	SPU_006814	GCAACATCGCCAGTCCAGCAGTTTGCTACAACACCAAACGGCATGGTTCT CCTCCTCCCAGCCCACACCGTCCAACACAACTCGGTCATCTCGGTTCGAG
Hex	SPU_027215	CACGTTCCCGGGGCATTCGTTCGGACCGTCGCCGTACGGACAGCCTCAGC ATCCGACCGGAGCTTACTACGATCCCGCCGCCTTACCTGGTGGGCTGGCT
Hh	NM_001012702.1	CAACTTATCAGGGTCGCTCAAACGGTGGTTATCGATGAGGCTCGCGGTAG GGAGGGGGAATGTAAATCCACGAGGTATCGGAAGTTGAGCAAGTGGATCA
Hlf	SPU_004414	GTGCACCTTTACCGCGGAGGAGCTCAAGCCGCAGCCTATGATCAAGAAGT CACGAAAGATCTACGTGCCCGATGAGCAGAAGGACGACAAGTACTGGGAG
Hmg1	SPU_027981	CAAGAAGGACCGAGACAGTTCAAAGCCTCGTGGACGTATGTCTGCATACG CTTACTTTGTACAGGATTCCAGAGCAGAGCATGGCAAGAACCACCCTAAC
Hmg2	SPU_005572	GTACAGGATCGGCGCAGTAAAGCTGAAGGGCAAGTGAATTTCACCGCTTT CTCTAAGGAATGTGCTGATAGATGGAAGCATATGGATGATGGAGACAAGC
Hmx	SPU_012490	CGATGTCGCCATCCAGTCCGTCGATGTCCAGCAATGGGTACAGTGAAGGG CCCATATACGGCTCCACTCGACACCCACTATCGGCGACCGTTCCGGGCTT
Hnf1	SPU_008196	GAGAAGTGGTAGATTCCACAGGTCTCAACCAATCACATCTCTCGCAGCAT TTGAATAAGGGCACGCCCATGAAGGGCATTAAACGAGCCGCTCTCTACAA
Hnf4	SPU_021192	TGCGTGGTGGACAAGGACAAGAGGAACCAATGCAGATACTGCAGGTTAAA GAAGTGCTTCAGGGCTGGCATGAAGAAAGAAGCTGTTCAGAATGAGCGGG
Hnf6	NM_214659.1	ACCTGATACCCATTCGTCTTCAGTAGCACCCGGCTTAAACCCGGGCGATT TCTTTAACGCCACGGATTTCGCGGTGAGTCGCCACTCCAGACCCGGTGGA
Hnrpr	XM_788184	GACGAAAGATGAGATCTTGGAAGAGTTCGCCAAAGTTGAAAAGGGTCTTC TGGATGTCATCATCTACAAGACAGAGGACAAGATGCGCAACCGTGGTTTT
Hox11/13b	SPU_002631	AACGCCGTCCATACTCCAAGTTACAGATCTACGAACTAGAGAAGGAGTTC ACAACTAACATGTACTTGACCCGAGATCGTCGCTCGAAGCTCTCACAGGC
Hox7	NM_214560.1	CGAGGATGGCCGCCACGTATAATTCAGCGTCGTGGGGCAGCACGGCGGCG GAGTTAGGGGACGGGAGCTACCGGGGTAGAGTGAGCGCACTGACTGCTGG
Id	SPU_015374	ACGCGCCGGAGTCTCGAACTTCACCATGTCCGATTGCTACGAGAAACTGA AGCAACTTGTGCCGACAATCCCGAAGAACAGGAAAGTCACACGAGTGGAG
IrxA	SPU_010351	AGGTCCTAAGGATTCTGAACACGGCGCACTTAGGGATGATTACCGAAGAA GAAGTGGAGAATTGTTAGCTGAACACCACCGACTGAGCGGATTGAATGGC
Jun	SPU_003102	AAAACCAGGAGCGCATCAAGGCTGAGAGGAAAAGGCTCAGGAATCGAATC GCCGCCAGCAAGTGCCGCAAGCGCAAGCTTGAGCGTATTGCCAGGTTAGA
Lefty	SPU_009911	CGGCGTCGCATTCAACATCGACCCGATCGAGAGAAGTCATCACGCCAGGC AGGTGGCCGACGAGATGGTCTTCGCAACCGACTTTTATCCCGATTCTCCG
Lhx2	SPU_021313	CACTACGAAACCCTCTTCGTTCCGTCCTCCGATCAGAGCACCTATCACCA TCTCCACCACGCCTCGCCTCCTCCTCCGCAACCAACCCTCCATAATCTCT
Lim1	NM_214645.1	GTCGCCGCGCGGAGAGACCTAGCGTCTCGTTCGCTCAACTTTTCGCCCGT AATATCTGTGCATTGAGCCACGGATACCAAGCAAAAGAGAGAGAGAAGAG

Lmx1	SPU_014157	ACCTCATGAAAGTAATGGATCACTGCTGGCACGAGCAATGTCTTCAATGC AGCGTCTGTAGAATAAGACTATCACACTCCTGCTTTGCCAGAGATCGCAA
Lox	NM_214650.2	ATGGAAGGGAACAATCCCGCGTACTATCACTACATCCCTAAGGGACAATT CCCGATACCTATCACCACGGGTGAAGTTAGCAGCTTTCCCTCTAGCTATA
Mad	SPU_006583	ATCTGAAGGTTTTACCTCGCACAAAGACCAGCTATGCAGAGAACAGCGCT TCCTCAGAAGGAGATTGGATTACTTACAGTCTACCTTGCACAGACAACGT
Max	SPU_022163	TTGGAGAGATCTTCGACAGGGGAAGTAGCCAAGACCAACGGCATCATTCT CAACACCAAATTATCTCACTACGACGCCGGTTCCGAATCCGACTCGAACT
Mef2	SPU_016168	TGAGTCCTACATCCTAACGCCACGCACTGAAGCCAAATACCAGAAGATCA ATGAGGAATTTGACAGAATGATGAATGGTGGAAGTGGGCAACCCATGAGC
Mitf/tfe3	SPU_008175	CAGGTCAAGACAGAACCGTTGGCTTACCCTGAAGAACATGTCCGAGCCCA TGCCAAAGAGCGACAGAAGAAGGATAACCACAATATGATTGAGCGAAGGA
Mlx	SPU_005787	GCTAGTTTTGCAGAGCTTTCGGCATGTGTATTCAGCTGGCTAGAGGAGTA CTGCAAGCCTCAGACACTCCGAGAACTGGTTGCTGAAGTCTTGAGGAAAG
Msx	NM_214613.1	CTCGCTCTCGAGCGCAAATTCCGCCAAAAGCAGTACCTCTCGATAGCCGA GCGTGCAGAGTTCTCCGCCTCGCTCAACCTCACCGAAACTCAGGTCAAGA
Myb	SPU_000861	ATGGCCCTAAGCGTTGGTCTCTCATCAGCAAGTTTCTGGTGGGGCGCACA GGCAAGCAGTGCAGGGAGAGATGGCATAATCATCTGAATCCTGATATAAA
Myc	NM_214579.2	CAGAGTTCCCTGCTACACATGGACATGACTATCTATAAGACTCTACGAGA GAGAACACTTTACCTCTTTGAGCGTATTGGATTTGGACCACTTACCCAGG
Nfe2	SPU_011174	AAAGAAGTCAATGAGATGCAGCAGAGGTATGCAGAACTCTGTGAAGAAGT CTTTGCCTCGGTTCAGGATGAGCATGGTAGCCCAGTTGATCCCAATGACT
Nfkb	NM_214654.1	CCTTGACCCTGTCATATCAGTGCCAGTCTTTGATAGCAAGGCACCCAACG CCACGACCCTCAAGATCTGTCGCATGGATAAGAGTGCAGGATGCTGTACA
Ngn	XM_780649.2	CTGACGAAGATTGAAACCCTTCGCTTCGCCCACAACTACATCTGGGCCCT GTCTCAGATGCTCAACATGGTGGACAGCAGCGAAAATGGTTGCCCAGGAA
Nk1	NM_001009577.1	CCAGGATGAACGAATGTCGCCTGACTTCTCGATGGCTCATCACGGAGTCG ACACAGCATCGCCGTTCTTTTCTCATCGGTCTCCGGGTCGGTTGCTAGCC
Nk2.1	NM_214635.1	CTTCGGCCAACAACCCGTATGCCCACATGCACGTCCCTCAGCTTTCCCAC AACTATTGCAACGGATCGGTCGGGGAGCTTAGCCAGCACTACAGCGACCA
Nk2.2	SPU_000756	ATCTACGACCAGAGTGACAATCCCTATACCAGATGGCTTCAAACTCAAGG GGGCGATGCAGTGCATTACTCATCTTTGGCACAGAACACCGTATCTCACA
Nlk	SPU_010846.1	CAGGTCAAAGAAGCGCTACACAAATTCATCACCGACAGACATCAGGGTAA CACCGTACCTCTATGCATCAACCCGCAATCAGCCGCCTTCAAGAGTTTTG
Nodal	NM_001098449.1	TGGACGTTGATTTCGGTCGAATCGGCTGGGACGAATGGATCATCTACCCA AAGCAGTTCAACGCCTACCGTTGCGTCGGTACGTGTAAGGGACCTCTTGA
Not	NM_214562.1	CCGTCGTGACACGGACCGAACATCGCCGTCCTCGTCGAATGACTCTTCGT CCTCCACCGCCAACTTCTCGGCCGCGTTGCAGTTGGAGGTGAAGAAGGTC
Notch	SPU_014131	CTGCGAGTCTGTCGTTAACTGGTGCAGCCCTCAGAACAACCCTTGCTACA ATGGAGCTAACTGCGTTGCAATGGGTCACCTGTATGAGTGTCGCTGTACA
Otp	SPU_019290	CCACCGTACTCGATTCACACCGGCCCAACTGAACGAACTGGAGAGAAACT TTGCCAAGACACACTATCCTGATATCTTCATGCGGGAGGAGATTGCGATG
Otxa	NM_214588.3	ATTCTCCGCAGCGTCGATCCGACATTGAACAGCGTTATCAGCTGGACTAA GAACTGGTATCTCCTCTGCTGGCCATTGATTGCTTCACCATACACTCTAC
Otxb1/2	NM_001032368.2	GTCTATTGAACATCGCTTAATTATCGTCAGATTATCCTATCATACTTGCC GTTCATTGGCATTTTGTTTAGGAACGTCGCTGGAACATTTCTGCACTTCT
P3a2	NM_214593.1	AGTAACCACATTGCCAGAGGGCACTCAACTTGTTCTAGCAAGCGATGGAT CACTCCAAGCTATAAATGATGGCACAGCACAAGGTATTGTTATCCCGGCT
Pax2/5/8	CUFF_192615_1.1	TCTAGACCATAGAGTTGTATAGAGCGACCGCTAGACCCATTTTCTCTGTG AGATCGCCGGTGAATGTGGCTCCGGCTAATCGCTTCGCACCGTCAGTTCA
Pax4	SPU_017635	GGACTCTACCCGACAGAAGATAGTCGAGTTAGCACACAGCGGCGCCCGTC CGTGCGACATCTCCCGTATACTACAGGTTTCCAATGGCTGCGTCTCGAAG
Pax6	SPU_006786	GACTCGTGATTCCTTTATCCGAGTCGGCAGGGTTGCACACATTGACACCT GTGCAGAGTGATACGAACGTAAGGGGCGCTTATACGCGAGTGGTTGAAAG
PaxB	AF016886.1	AGGGACAAGTTCACTTCCCGTCGGTCCCTAGAGAGAACCAGTGTTATCCT ACTAATAGCATGATAGTGGGGAATCCAGATCTTGAGACTCAATGGGCGGG

Pea	SPU_014576	TGGAATGCGGCAGGGCGTTCTATGATGATGCAAGCGTACCGGAAAAGGCT CAAGAACAACACAATGAGCCGAGGCATGAGGTGGTGCGAGAAGGGCCTCT
Pitx2	SPU_004559	GGTGTGAAGGGTGGTCACGACGATGGCGACGATTCGGACAACGACCAGGA CAACGGCAAGAAGAAGCGTACTCGTCGACAACGCACACATTTCACCAGCC
Pks	SPU_002895	CCGTGGTAGGCATTGGAACTCGCCATGCTTGTGGCGCAAACACTACCGAT GACTTCTGGAAGGTTCTCAAGGAAGGTAAGGAGTGCATCCTGGACATTCC
Plod2	SPU_005537	TGTCCTTTTCCTGGCTGATGCTGATGAGATGCTGAGGAAGTTCAAAGCAT ACCAGATCAATCTACTCTTCTCGGCTGAGACGTACATCTGGCCTGAAAAA
Prox1	SPU_015984	TTTGTGCAGCGATACAGAATCACAAGATTCATTTGCGCAGGACAGAGTGG CGGAGGATCGCAGCAGTTTGCTTTATTCGGAAGAAGATATTGCGGAGAGC
Ptc	SPU_003313	AACTCAGGACATCAATGCTCTCTCATCTGCATCTCTGGAAGATCTCCTCC AAGACTTCTCACGGACAAGTGTTGTTCGGGTTGCCATGGGTTACGCTATC
Reverb	SPU_017492	TCAGATGAAGACGGAGCAGCTCATACTAAGCATCCATGAAGCACAGAAGA AGACCTTATGGGATTGGGGCATCCTGCAATGCAGGAGTTATAACCTTCTA
Runt1	NM_214614.2	CCGGTACGGAGGAACAACCTACATAACAGTTATAAGATGGCCGAAGGGGG CCAGAGGAATAAGGCCTCATCCGTCTTCAAAGGAGGAGAACGGTCTATTG
Rx	SPU_014289	GCACGCAGCAGCGGTCGCTGTCGCTGCAGCCGCCAGCCAAACAGCTATGT TTCCTTTGCTCAGCCCGACTCATCTCACATCGCCTCCTGGCCGCCCTGGG
Scl	SPU_028093	ACCGGTGCTGGAGCGGGTAGGCTCGGGTAAAGTCGTACGCCGAATCTTCA CCAATAGCCGAGAGCGATGGCGGCAGCAGAATGTTAACTCTGCATTTTCC
Shr2	SPU_008117	AGCACGCTAGCCAGCGTAGTCACATCCTTAGCCAATATGAACAAGAAGAC GGAAGAAGGGCCAAGTCATTCACAACAGATCTATTCTCCATCTCAGACCC
Sim1	SPU_013962	CAGTTACTTAAAGATGCGTGCCCTATTCCCTGAAGCATCGCCAGGTAATG GAGCGGTAAACCGGAGCACCGATCTTATTTGCAATTTTCGCTCCGATTGC
Sip/z81	SPU_022242	CAGCGACACTGCAATCATCTATCCTGAACCTGTAGAAGACATGGACGGAG TCAATGGCGATACAGATGACCCTGATACTCCTGAAGGAAATGACAACGAA
Six1/2	SPU_017379	GGTTGCCTGCGTCTGCGAGGTTCTCCAACAATCCGGCAACATCGAGAGAC TTGGCCGTTTTCTCTGGTCGCTACCTGCCTGTGAGCATCTCCACAAGAAT
Six3	SPU_018908	AAAGAGAGAACGCGGAGTTTGCTCCGGGAGTGGTACCTCCAAGACCCTTA TCCTAACCCTACCAAAAAACGTGAACTGGCTCAAGCCACAGGACTTACAC
Sm50	NM_214610.1	TTCCCTACAGGATGGCCTCCGAATTCTGTGAAATGGTTACACCTTGTGGA AATGGACCAGCAAAAATGGGTGCTCTGGCTTCAGTTTCGTCGCCTCAGGA
Smad4	SPU_004287	TGCATATCACTATGAGCGAATTGTGTCACCTGGTATTGATCTGACTGGGC TCACATTACAACACACAGCTGGCCCTCCACGTGTGGTTAAAGACGAGTTT
Smo	SPU_006789	GAACTGGAAAGAGGATGGCCACATTTCTTGCAGTGCAACGAAGAGCACAT GCCTAGAGATTGCAGTGAACCAAACACCTACAAGGCAGTGACATTTAACT
Snail	AY372519.1	CCCGTCAACATTCCGCATCCAGTTATCCACAAACCCGAGCCTCTCCAGGC CATTCCCAACCCATCGGCCTACTGGCGACACCATCCGAATGTGATCTACA
SoxB1	NM_214474.1	ATGATGCACCACGCACCGTCTGGCACCTCAGCCGCCAAGTCCGAACAAGC GTCGATGCCAACAGAGCGAACGCAGTCAGCAGCCACCGCCGTCTCCGCTG
SoxB2	NM_214473	TGCCTTTATGGTTTGGTCAAGAGGACAGAGGAGGAAACTTGCTCAGGAAA ACCCTAAAATGCACAACTCCGAAATCAGCAAGAGGTTAGGAGCCGAATGG
SoxC	SPU_002603	GCTGCTAAGACCACCAGCAGCAAACCAAAAGCCAACAAGCCCAAGTCATC ATCGAAATTGACGAAGATGAATGGCATCGTGATCGACCAGATGCACCCGC
SoxD1	NM_214472.1	ACTATGAGGAACAAGCCCGACTTGACAAAGCTCATCTGGAGAAATATCCG GACTACAAGTACAAGCCGCGCCCTAAGCGAACGTGCATCATCGACGGCAA
SoxE	SPU_016881	GCTGGTGTCGATGTCCGTGACTTTGGTGGCGACATCATGGGCATGGAGGA GTTTTCATCGGAGGAGCTGGACCAGTACATCGTGCAGACCATAGCTAGTG
Spec1	NM_214603.1	TCAGTCCTCAAGAGCTGCGTGAAGCGTTGTCAGCAAGCAAACCACCGATG AAGAGGAAGAAAATCAAAGCAATCATCCAGAAAGCCGACGCCAATAAGGA
Spz12	NM_214616.1	GAGTGTCTTCATGGAGTGAGGCCCTACGAATGCAAAACTGCGGGAAACGG TTCGCCCAGAAGAGCTGTCTTACGCGTCACACCAAGATCCACACCGGAGA
Srf	SPU_027774	GAAGCTGAAAAATGGTATCAGGGACCATGGATCATGTTTACCGGACAGTG ACAGTAGCTCTGTTGGTACGGGCTCCGAGCTAGCGGACCGGGACCATGAC
Su(H)	XM_001200121.1	GAGATCATCAACGATGGAGCTTCTTGGACTATTATCAGTACAGACCGGGC AGAGTACAGCTTCTATGAGGGCATGGACCCTGTTAAGATGCTTGTTACAC

Tatabp	SPU_012621	AAGAGATTTGCAGCAGTGATCATGAGGATACGTGAGCCAAGAACCACAGC TCTAATCTTCAGCTCAGGGAAGATGGTCTGTACTGGTGCAAAGAGTGAAG
Tbr	SPU_025584	TCGAATTCACGTCCTAGAGTTGAGCGAGAGCCGCTCTATCCAAACCCATA GCTTTCCCGAGACGCAATTCTTCGGCGTGACGGCCTACCAAAATACTGAT
Tbx2/3	SPU_023386	TCCATTCGCTAAAGGTTTCAGAGATACAGGGGCAGGGAAGAGAGAGAAAA GGAAATACATTGGCGCGACGGGCACATACGAAATCGACCACGGTGATGAC
Tbx6	SPU_020346	ACGGTTGCGATGCGAGGTTGTCTGATTATAACACGGGAAATGGAATGGTC GATACGTTATCGCCCTGTTACACGCAGGACGGTGTCATGGATGACAACGG
Tead4	SPU_021210	TGAGATCTCGCGGTCGCCCATGTGTGAATACATGATCAACTTCATCCACA AACTCAAGCATCTCCCAGAGAAGTACATGATGAACAGTGTTCTAGAGAAC
Tel	SPU_028479	CAGCGCATGTTCGAGAAACAACCCTTGCGACTTCCCGGCGGGAAAAACCT CCACGACAACTTCAGCGAGTCGGTGATGGATTCCTCTCAGGTTCCGCCGT
Tgif	SPU_018126	AGCTCTACCTATCTCGCTTGGCTAACCTCACTCTTCTCCAGGTGTGCAAT TGGTTCATCAACGCTCGAAGACGTATCCTGCCAGAGATGATTCGTCGCGA
Thrb	SPU_018861	ATGGATCTGAGTAATGGCCCTGAACCGTGTGTAGTATGCGGAGATGCTGC TACTGGCTATCACTACAGGTGTATGACATGCGAAGGATGCAAGGGTTTCT
Ubq	SPU_021496	TTGTCAAGACCCTCACCGGCAAGACCATCACACTCGAGGTCGAGCCAAGT GACTCCATCGAGAACGTAAAGGCCAAGATCCAGGACAAAGAAGGCATCCC
Unc4.1	SPU_001739	AGCGAGATCGATTCAGAATGCATGGACAGTATGAGTATTGATACAACGGA GGAGCAGCGAAAACTCCCGGATTCGCCGCGCTTCAAGTTAGCGGCACGAA
Univin	NM_214628.1	GGAGTTGATGTCAACTTCAGGTAACAACCGACGTGGCAGTCAAGTCATCA AAGAGCTTGGAGCGATAAGCAAAAAATGTACAGCGAATCTGATTGTGACG
Usf	NM_214653.1	GCTTCACAGAGAACCATCGCTCCGAGGACACACCAGTTCAATACGAAAAT TGACAATTCAAGAACGGTTCGTGATGAGAGACGGAGGGCGACTCATAACG
Vegf3	SPU_030148.1	CTCTCTGGAACAAATGTACTAAGTGGCACGAGCACCGGCGTGAGAGACTC CTGCAAATGTTCCCATTACGACGCCGATGGAAGGCGCGTCTTCGAGCGCG
Vegfr	XM_779931.2	TTCAATGTTACCTGCAATGTAGAATCACCCCTTGGAATCCAAATCAGGTG GAAATGGGACTATCCAGGTCTTCATAACCCCAACGCTAATCCCAATAATT
Vitellogenin	SPU_016052	AAACGTTGTTAGTCCTCGCTTGCGTAGCGGCTGGATCTCTAGCTGGACCA GTGAGGGAAACGCAATGTGCAGAAGAATGCAATAGTCGAGATGACGAGGG
Wnt1	SPU_011756	GGTCTCACAGAATCGAGGGACCTTAGCCGCTATTAACAGAGCAGTCAAGA TGGCTGTATCAGAGTGTCGATACCAGTTTAAAGAGAGGAGGTGGAACTGT
Wnt16	SPU_011130	CTCTCTGTCAAAGATTCCCGGGACTTACTGTTGAACAGCGGAGAGTATGT TCGAGTACACCCGAGATCATCAACATGATAAGTGAAGGTGCTAAGGTCGG
Wnt4	SPU_023065	TGCTGGAAATCGATGCCGACCTTCGGCGACATTGGACAGGTCCTCAAGGA GAAGTTCGACGGTGCTACTGAAGTCCAGTCCCTGAAGATCGGATCCAGAC
Wnt5	SPU_026277	TCCTTCACGAGAGCACAAGATACAACATGGATAAATCTAGGACTGGACAC GAGAGTCCAGCAATTCGATGCCTTCCGGAACCCAGAGCTGTTCATCCTGG
Wnt8	NM_214667.1	CAGTGACAATGTACGATTCGGTGAACGAATGGCGAGCGATATCATGGACG ACGCAGAGAGTTCACAGGGCGCCATCTCCGTTATGACTCTACATAACAAC
Wnta	SPU_024946	GCTAAATTCTGCAAGTCCCACACGTGGTTCACCAGGCAGCAGCTAAAGCT ATGTCAGCTCCATCCCGACATCATTCCAAGTGTGACCCAAGGCGCCATGC
Z108	SPU_003378	TGGTGGTGCATACGGGTAGAAAAGACTTCTTGTGTCAGACGTGTGGGCAA AGGTTTGGGCGCAAGGACCATTTAGTGCGGCACACCCGCAAGAGTCACAA
Z115	SPU_000440	ATAATTCTTCTCCAGCCAAAGATAGTAAAGCTGAAAGCCTAGTCCCGTCA AAGAGTCAATCAAAAAGGCTTACCAAGCAAAGGCAACCTGTACAGAAAGC
Z121	SPU_024877	ATAGCACCCTGCGGCGCCTAGTACTCAGCGGCTCCTCGTCAGTGCCATCA TCGCGCCAGCACACATCATTGGGGTCTTCGTCGGGGGCACCGCCCAAGAA
Z13/Krl	NM_214497.1	AGCTCACCGAGACGCTCCGACGTTCTCCGACGGGTTCCGCTTCATCAACG GACATATCCGCGATGCCTACGAAGCGTCAACGATTAGATGAGACTTCATC
Z133/fez	SPU_019089	TCACCCAAACTCGCATTTTCAATTGATAGCATCTTAGGAATCAAGAAGAT GAAGAAATCTAAACTTGAGCCATCGGTGAACGAGGATGCATTCCATAAAG
z141	SPU_011189	CAATGCCGTGGATAAACTGACCCGTCGACCTCCCAAGAACGTGCTCAGGA TCAAGATGGGCTGCACTGACAAGCGTAAGGTACGCAGCGGCAAGCTCAAG
z157/ovo	SPU_012448	ATACTGCTACAAGCAGCGCCGTTCCAAGATATTTGTCTGTGAGGAATGCG GCATCACGACCGAAACCGCCGATTTTCACTACGACCACATCAAGGAGATA

Z166	SPU_012465	GAGGTTCACGTGCGACGTTCCCACAGCGGTAAGCGTCCTTTCGCCTGTGA CATTTGCCAAAAGACGTTCGGCCACGCGGTCAGTCTCAGTCAGCACCGCG
Z188/klf13	SPU_023727	TCACATCTTAAAGCTCATCTCAGAACACACACAGGAGAGCGACCATTCCC GTGTAACTGGCAGGACTGTAACAAGCGTTTTGCGCGGTCTGACGAGCTTG
Z199/sp5	SPU_024189	AGTCTATGGCAAGACCTCGCATCTCAAAGCTCATCTGCGTTGGCACACAG GCGAGCGACCCTTCGTTTGCAACTGGCTTTTCTGCGGGAAGTCGTTCACG
Z204	SPU_006484	TCAATGTACATCACTCAGGGACCAACTCAGCAAGGAACGGGAGGAGAAAG GTGTGTTGAAGGAGGAGGTGGATCGTCTCCGGGAGGCCCTTGTCAACTTT
Z214	SPU_026418	AAACGAGCCAATCAGTGAAAGACATCCCTACTGTCACCGCTACTGCTGGA TGCTTCACCATGAGTCCTCGTCCCATCGAGAAATCTGACTTCGACGCCGT
Z22/gli1	SPU_017627	GCTTACTCTCGACTTGAGAACCTCAAGACTCATCTTCGATCCCATACCGG TGAGAGACCGTATGTCTGTGAGTTCCAGGGATGTACCAAAGCCTTCTCTA
Z244/zic	SPU_028583	CTCGGCAGCCGAGACCACCGCGCTCAGCTCGGGCTTATCTGATAGCACGG GGATGACGCTGAACATGAATTCCTACGGCATGGACACCTCTCATATGACG
Z30	SPU_017348	CGGCGATCAGGGAGGACAAGGTTTACAGACTACCAGATTAAAGTTCTACA GGAGTTTTTCGAAAGCAATGCCTACCCTAAAGATGATGACCTCGACCACC
Z400/klf3/8	SPU_017470	TCTGACGAACTCACCAGACACTATCGCAAGCATACGGGTGACAAGCCATT CAAATGTACTCACTGTGACAGGGCGTTCTCAAGATCGGACCATCTCTCGC
Z48	SPU_009642	CACCATGTGTGGTAAGAAGCTGAGTTCTACCAGTTCCTTGGACCGCCACA TGTTGATCCATTCTGGAGAGCGTCCCTTTACATGTCCTCTGTGCGCCATG
Z487	SPU_002140	ATCTGATGGCAGCAGAGACTGTTCAGTCATTCCCAGAAGTCAAGCAAGAA ATGGAGAGAGGGACACTTGATCATCACCAAACTACATATAGAAGTGTAGC
Z54/spalt	SPU_014793	ACCAAGGAGCGTCCATTCCGCTGTGATTGCTGTAGCAAGGGATTCTCCAC CAAGGGTAACCTCAAGCAGCACATGCTGACTCACAAAATCAGAGATATGC
Z55	SPU_014197	ACATGGACGAGTGTACCAGTGTAAGCTATGTACACACATTTCTCCTACGA AGACACACTTAAACGAGCACATGAATGTTCATTCAGGGAAGAAGCCCTAC
Z57	SPU_015767	TGTGCAGCGTTTGCCTGAAGTGGTTTAGCCACGAGTCCAAGTTCTTGCTT CACATCCGAAGACACAGCAGCATTAAGCCTTTCAAGTGTAACATGTGCGG
Z60	SPU_015358	AGTTTGATCAACGAAGCGGTATCGGCTCAGCCTATCGTATCCTGCGAGAC GGCTTGTTTCACACCTAGCTGCACCTCAAGCTTTCAGATCGCCAGCACTC
Z62	SPU_012914	CCTTCAACTTGCGTATACACTACAGGCGTATCCATGCTTCAGAAGATGAG AAGAAACACCATTGCATGTCCTGTGACTACAAGTGTGCTGACAAAGGCAT
Z85/klf2/4	SPU_020311	GGGAAACCACTGCGACTTTTCATACGAGATTTGGCCTGATATGGAGAGCT TTCTTCAGGAGATCTGTCCGCCAAACAGCTTTGAGGAGCCATCATTTGCC
Z86	SPU_012772	AACAGAAAATTACAAGGCGGCGAGAATGATATCTGAAGCGACCCTCACGC CACCATCATCACCTGAACTTGGTCGTAGCCGCTTAGTGAGTGTGCCTGTT
Z92	SPU_012913	CCCGAATGTGACTTTAGAGGTGCAACGAACAAGATCATCTCGGAGCATGT GATGTGCAAGCATGCCCGCGTCCGCCCGTACACCTGTAAAATATGTGGTT