The Developmental Organization of Regulatory States in the Sea Urchin Larva

Thesis by

Jonathan Exiquio Valencia

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy in

Biology and Biological Engineering

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

Defended March 13, 2018

Jonathan Exiquio Valencia
ORCID: 0000-0002-3498-8980

ACKNOWLEDGEMENTS

In my time at Caltech, I received an incredible amount of support from a wonderful contingent of people to whom I would like to express my gratitude.

First and foremost, I would like to thank my parents, Dorothy and Exiquio Valencia, and my sister, Amber, for their unending support, encouragement, and understanding. Going home to visit always provided a haven of tranquility and relief from the grad school grind. I would also like to thank my grandma, Diet. Even though you think I'm attending MIT, you are still a source of inspiration.

I am honored to have been under the tutelage of two illustrious advisors. To my former and late advisor, Eric Davidson, who offered me a position in the lab shortly after showing him my in-situs. Thank you for accepting me as your student and believing in me. I will never forget the sound of your scooter cruising through the hallway and the fear, panic, yet awe it induced. You are missed. To my current advisor, Isabelle Peter, with whom I've shared this journey since the beginning of my time in the Davidson lab: we've been through quite a bit in the past few years, but I've learned quite a lot about science and life from you and I am grateful for your mentoring. Thank you for all the years of advice, guidance, support, encouragement, and the many helpful discussions.

I want to acknowledge the members of my thesis committee: Marianne Bronner, Lea Goentoro, and Paul Sternberg. I appreciate your time and energy. Thank you for your support and advice.

For the majority of my time at Caltech, I shared this academic journey with an incredible cohort of individuals whom I have the pleasure of calling my friends: Eric Erkenbrack, Mike Rome, Weston Nichols, Alex Webster, Justin Liu, Anna Abelin, and Avni Gandhi. You all mean the world to me. We all laughed, cried, and suffered together but yet had so much fun during our years as grad students. Only you truly know the experience and can appreciate the journey. Thank you for everything and the countless memories.

To my friends whom I met at Caltech: Tobias Heinen, Daniella Haming, Marcel Dammert, Carlos Gomez, Paul Minor, Brian Duistermars, Nate Pierce, Cory Tobin, James Lee, Alicia Rogers, Ruzbeh Mosadeghi, Rob Oania, and Kata Fejes-Toth. You made grad student life bearable and enjoyable. I always looked forward to meeting up for a coffee, a beer, a baseball game, or a workout. Thank you for the good times.

Special thanks to Finn Carlsvi, Kurt Reichermeier, Chris Somoza, and Chris Janover. You are incredible individuals who have taught me so much. As I've learned from you, it never hurts to add a little adventure in trying times. Thank you for sharing epic adventures and the feeling of success when summiting a mountain peak.

Many thanks to the people who I worked with closely over the years. To those in the Davidson lab: Sagar Damle and Stefan Materna – you both were instrumental in my laboratory-upbringing; Smadar Ben-Tabou De-Leon, Enhu Li, and Feng Gao – you were admirable post docs who created a wonderful and inviting lab atmosphere. To those in the Peter lab: Roberto Feuda – you inspired and encouraged me on the daily, thank you for being my friend, confidant, and fellow reggae enthusiast; Mike Collins – you were an integral part of my time in lab, thank you for a reliable dose of comic relief; Deanna Thomas, Erica Vielmas, and Ping Dong – you three made the last stretch of grad school absolutely delightful with the obligatory pho on Tuesdays and the weekly, random donut-runs. To the students who aided me in my research: Cathy Dong, Gracia Ng, and Eric Qiao – it was a pleasure working with you all. Together, you all have contributed to making my time in the lab special.

I also want to acknowledge Liz Ayala. You made the formalities and bureaucracy of grad school painless and effortless. Thank you for all your help.

I am indebted to the people I met outside of the Caltech circle, Rannie Amundsen and Zara Kunkler, who offered me escape and a breath of fresh air. Thank you for making my time in Pasadena special.

Last but not least, I would like to thank the countless number of people who have contributed to my personal and intellectual development both within and outside the academic setting over the years. An acknowledgments section is always an incomplete list as there are too numerous people to list but who all are no less significant and those who go unmentioned. You have my sincerest apologies as my memory failed to include you during this retrospection. This last section is to thank those unlisted and unmentioned.

My thesis would not have been possible if it were without the support of those mentioned above.

Thank you all.

ABSTRACT

Development is an inherently dynamic process where cell fate specification occurs continuously and in a progressive manner. Thus, a major focus in developmental biology is solving the gene regulatory networks (GRNs) that underlie specification of cell fates. GRNs specify new spatial domains of cells by controlling the expression of their changing regulatory states throughout development. Regulatory states are composed of combinations of expressed regulatory genes which encode transcription factors (TF) that form regulatory circuits which function to carry out the specific developmental tasks involved in cell fate specification.

To investigate the differences of GRNs operating in the embryo and their change over development, we sought to identify and characterize the regulatory states present in multiple developmental stages of sea urchin embryogenesis. We performed a genome-wide survey and embryo-wide annotation of regulatory gene expression by whole mount in situ hybridization at five consecutive developmental time-points in order to determine regulatory states and their developmental trajectory. We determined at least 74 distinct regulatory states expressed in discrete developmental domains which coincide with larval morphological structures and show that their progenitor domains foreshadow the ensuing larval morphology. Among these domains, we identified bilateral ciliary photoreceptors in the larva which express a distinct regulatory state that include factors known in ciliary photoreceptor specification. We show that this photoreceptor regulatory state does not express the genes of the retinal determination network that specify eyes in both flies and vertebrates. In addition, we show that though the sizes of regulatory states are comparable over developmental time, no two regulatory states are equal, even those expressed in a given domain at previous or subsequent developmental time-points. Lastly, we found that similarities among regulatory states reflect a common developmental function but not necessarily a common developmental history. The results suggest that the combinations of TFs defining regulatory states are both spatially and temporally dynamic in their progressive specification of cell fates during development and that regulatory state expression is tightly associated with the developing morphology of the larva.

PUBLISHED CONTENT AND CONTRIBUTIONS

Valencia, J. E. et al. (2018). "Bilateral ciliary photoreceptors in sea urchin larvae express a partially conserved regulatory state". [*submitted*]

J.E.V., R.F., and I.S.P conceived project; J.E.V. performed WMISH; J.E.V. and R.F. performed analysis; D.O.M. and R.D.B. contributed antibodies and performed immunostaining; S.Y. and R.D.B. performed and analyzed calcium reporter experiments; J.E.V., R.F. and I.S.P. contributed analysis of RDN circuit in the context of GRN evolution; R.F., R.D.B., and I.S.P. wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Number *Page*

CHAPTER 2

CHAPTER 3

LIST OF TABLES

Number *Page*

Chapter 1

INTRODUCTION

1.1 Specification

One of the most biologically fascinating aspects in animal development is the increasing organization of cell fates in time and space during embryogenesis. This unique and remarkable developmental feature is the result of developmental specification. Specification is a continuous, progressive, and irreversible process by which cells in diverse embryonic domains establish development states of specific gene expression in order to acquire their distinct identities[1,2]. During development, specification drives the continuous partitioning of the embryo into distinct, spatially organized domains of cells which express a specific set of genes that underlie their fate and function. As development progresses, cells within spatial domains undergo successive rounds of specification until they reach a differentiated state where they exhibit a distinctive cell type morphology and function. Thus, specification determines the formation of morphology and diversification of function in cells of embryonic spatial domains, providing an organized structure and complexity to the developing embryo.

The specification of cell fates in organized spatial domains throughout embryogenesis is a reproducible feature that is shared not only among individuals of a species but also across their generations[3]. This robust and conserved feature suggests that an intrinsic biological program exists to precisely regulate the developmental specification process. As we now know, this program is encoded within the DNA sequences of the genome via two types of components: regulatory genes which encode transcription factors that modulate gene expression by binding to their target DNA binding sites in a sequence specific manner and regulatory regions that contain binding sites within *cis*-regulatory modules which control the

expression of genes by integrating multiple regulatory inputs[1,4]. These sequences make it possible for the genome to direct the specification program in development by determining the specific set of genes expressed in cells of distinct spatial domains in the developing embryo. Though the instructions that direct the developmental program of specification are encoded in genomic DNA, how these instructions are executed is insufficiently understood and remains the most intriguing question in development. Fortunately, these instructions have been shown to be mediated by gene regulatory networks (GRN)[4].

1.2 Developmental Gene Regulatory Networks and their Functional Regulatory States

GRNs demonstrate how genomic programs operate in the specification of cell fates during development. They are a system composed of interconnected regulatory genes encoding DNA-binding transcription factors that modulate transcription of downstream genes and signaling molecules which mediate cell-cell interactions. These genes are linked together through their cis-regulatory modules (CRM) in particular topologies that determine the specific functions necessary for cells to acquire their distinct identities[1,2,5]. That is, these functions translate into discrete developmental tasks that solve the general problems common to cells within a spatial domain during specification. They include installing a new specification state by interpreting initial spatial inputs, locking down the new specification state, preventing the activation of alternative specification states, and activating differentiation genes for the terminal cell type specification. Thus, developmental GRNs control the progressive establishment of cell fates in each spatial domain of the developing

embryo through functions provided by interconnected regulatory genes and signaling molecules of its active state, i.e., its regulatory state.

The regulatory state is an important feature and an essential prerequisite of a GRN which provides the regulatory toolkit with which to perform developmentally specific functions of specification[1,6]. A regulatory state is defined by the total set of expressed transcription factors (TF) in a nucleus of a cell. For their functional operation, regulatory genes must be expressed at levels high enough to occupy their targeted DNA-binding sites on downstream genes[7]. However, the specific regulatory functions in the control of gene expression and genetic functions involved in the establishment of cell fates mainly depend on the combinations of TFs within a regulatory state[7]. At the cis-regulatory level, the transcriptional regulation of a downstream gene is molded by the discrete combination of other TFs within the regulatory state that work together in tandem on its CRM to execute either transcriptional activation or repression. At the circuit level, where the linkage architecture of TF genes within the sub-circuit is indicative of its discrete developmental job, different developmental domains express distinct regulatory states, hence TF function may perform differently in context-specific situations by operating with different TF genes in similar topological subcircuits[1,7]. Thus, the combinatorial nature of TF genes within regulatory states represents an underlying feature in their functional control of gene expression and in the common developmental tasks required for regulatory state specification.

1.3 Sea Urchin as a Model for Developmental Gene Regulatory Networks

Much of what we know concerning GRNs has been validated experimentally with the developing sea urchin embryo [1,5]. The sea urchin currently stands as a principle model organism for the studies of developmental GRNs. Its endomesoderm GRN model represents one of the best analyzed and nearly solved developmental GRNs. Numerous studies have contributed to the building and maintenance of the endomesoderm GRN model through the analysis of regulatory gene expression, the wiring of regulatory genes through the construction of circuit models, and the authentication of regulatory linkages within subcircuits [8–16]. Moreover, additional analyses have contributed to the building of GRN models in other territories of the sea urchin embryo further expanding our knowledge of developmental GRNs into an embryo-wide view [17–22].

Although the majority of analyses of developmental GRNs in sea urchin stems from the early developmental process, i.e., pre-gastrulation, little to nothing is known about the GRNs operating in the development of the larva. What is known is that the 72h echinopluteus larva bears more resemblance to bilaterian deuterostomes than its pentaradial adult form and yet it is also a completely functional organism. The early pluteus larva is characterized by two arms supported by an internal skeleton, a tripartite through-gut, and a brain, and uses the coordinated beating of cilia to freely swim, feed on phytoplankton, and thrive in the water column for two months before metamorphosis[23]. In addition, as we show here, it also possesses light-responsive photoreceptors[24]. Despite these interesting developmental features, most research on larval sea urchins focuses on ecological experimental studies

geared at understanding larval feeding, growth, and life span; however, few contributions on gene expression patterns and regulatory linkages do exist, albeit focusing on small subsets of genes in the context of few cell types and organ systems [25–29].

1.4 The Unknown Developmental Regulatory States of the Sea Urchin Larva

During sea urchin development, regulatory state expression changes in both time and space in order to define the larval body plan and progressively specify the constituent cell fates. At 24 hours post fertilization (hpf), the sea urchin embryo is a round, hollow ball of cells with several known domains of cells expressing unique regulatory states. By 72hpf, the sea urchin embryo has developed into an elaborate larval form with multiple morphological structures, indicating numerous changes in regulatory state expression. Understanding these changes in the domains of multiple cell fates of the developing sea urchin embryo requires the identification of the active regulatory states expressed by their operating GRNs.

Understanding the GRNs controlling the regulatory state expression in the specification of cell fates in the developing embryo and their change over time requires a system-level approach. In a time before system-level approaches, numerous single gene analyses using in situ hybridization have contributed much to both our understanding of pattern formation in various developmental processes and to the construction of GRNs in many model organisms. Nevertheless, studying network features such as regulatory states and their change in expression in development require the knowledge of all developmentally

expressed regulatory genes and their spatio-temporal expression patterns. Current studies using system-level approaches like single cell transcriptomics offer the advantage of knowing all genes expressed in cells or tissues thereby gaining incredible insight into the expressed composition of their developmental fates. However, the spatial information of sequenced cells is often lost. Studies surveying spatial gene expression are complementary to transcriptomics, which offer constructed gene expression maps of the embryo during development. While each of these studies have their merits in understanding development through gene expression, cell-fate specific gene expression is a consequence not just of a single regulatory gene or all expressed genes but of the combination of transcription factors, the regulatory state, expressed in a given cell fate. Thus, the identification and characterization of regulatory states within an embryo at multiple developmental stages in any model organism has so far not been determined.

1.5 Summary of Thesis

The goal of my thesis was to investigate how developmental gene regulatory networks change in time and space throughout the ever-changing landscape of the developing sea urchin embryo. Since the active output of any given GRN is its regulatory state, we aimed to identify and characterize the regulatory states expressed at multiple developmental times in order to determine the developmental trajectory of regulatory states. In order to perform this investigation at a system-level, we used a sequenced genome and developmental transcriptomes as resources to determine all developmentally expressed regulatory genes in

the relevant time window of sea urchin larval development. With the help of a few others, I performed a genome-wide survey of all developmentally expressed regulatory genes by whole mount in situ hybridization and annotated their expression patterns at five consecutive developmental time-points. As a result, we determined the location and composition of regulatory states expressed within the developing sea urchin embryo across five distinct developmental stages.

Importantly, we pioneered the regulatory state concept in a gene expression analysis, a novel feature that according to our knowledge has never been investigated until now. In our analysis, we introduced mapping out distinct regions where combination of regulatory genes are co-expressed, i.e., regulatory states, across all territories of the developing larval. In addition, we determined regulatory state expression across multiple developmentally distinct stages in sea urchin embryogenesis, creating a temporal profile to track the developmental progress of regulatory states in time and space. Hence, this work represents a first look at investigating changes of developmental GRNs across development by assembling their regulatory states and by tracking their developmental change. The implications of this work will be discussed in the following chapters.

An advantage in surveying gene expression is that it opens the door to the discovery of new spatial domains or novel cell types as we show in chapter two. There, we identified a bilateral cluster of cells expressing a distinct regulatory state that includes the retinal homeobox gene *Rx*, which is known to be expressed in ciliary photoreceptors. We show that this regulatory state is co-expressed with *opsin3.2,* a Go-opsin, and that this opsin is localized to the cilium

within these cells. We also show that these cells are neurons and are capable of responding to light. Lastly, we reveal that genes involved in the specification of eyes in flies and vertebrates, those of the retinal determination network *pax6, six1/2, eya,* and *dach*, are not expressed in the larva photoreceptors or in their progenitors at earlier developmental time-points.

The main body of my work is represented in chapter three. There, we investigated the regulatory states operated by the underlying GRNs responsible for producing the morphological complexity of the sea urchin larva. We found that the larva consists of at least 74 discrete spatial domains that express distinct regulatory states which associate with the morphological features of the larva. We show that the regulatory states established in these domains during development foreshadow the appearance of future morphological structures. We also found that a majority of developmentally expressed regulatory genes are expressed in specific patterns spanning multiple regulatory domains. Further, we show relatively little variation in both the number of regulatory genes expressed and the relative contribution of different DNA-binding domain families among the larval regulatory states of specific cell fates. Lastly, we reveal that the similarity of regulatory states as determined by hierarchical clustering reflects their association to specific morphological structures, suggesting that functionally and structurally related cell fates tend to express similar regulatory states. In summary, this chapter reveals that the spatial regulatory information driving the morphological complexity of the larva is provided by the unique combinations of transcription factors expressed in regulatory states of specific cell-fates.

1.6 References

1. Peter, I.S., and Davidson, E.H. (2009). Modularity and design principles in the sea urchin embryo gene regulatory network. FEBS Lett. *583*, 3948–3958.

2. de-Leon, S.B.-T., and Davidson, E.H. (2007). Gene Regulation: Gene Control Network in Development. Annu. Rev. Biophys. Biomol. Struct. *36*, 191–212.

3. Peter, I.S., and Davidson, E.H. (2015). Genomic control process: development and evolution.

4. Davidson, E.H. (2006). The regulatory genome: gene regulatory networks in development and evolution (Burlington, MA ; San Diego: Academic).

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

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

7. Peter, I.S. (2017). Regulatory states in the developmental control of gene expression. Brief. Funct. Genomics. Available at: https://academic.oup.com/bfg/article-lookup/doi/10.1093/bfgp/elx009 [Accessed August 24, 2017].

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

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

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

11. Oliveri, P., Carrick, D.M., and Davidson, E.H. (2002). A Regulatory Gene Network That Directs Micromere Specification in the Sea Urchin Embryo. Dev. Biol. *246*, 209–228.

12. Revilla-i-Domingo, R., Oliveri, P., and 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. *104*, 12383–12388.

13. Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., *et al.* (2002). A Provisional

Regulatory Gene Network for Specification of Endomesoderm in the Sea Urchin Embryo. Dev. Biol. *246*, 162–190.

14. Davidson, E.H. (2002). A Genomic Regulatory Network for Development. Science *295*, 1669–1678.

15. de-Leon, S.B.-T., and Davidson, E.H. (2010). Information processing at the foxa node of the sea urchin endomesoderm specification network. Proc. Natl. Acad. Sci. *107*, 10103–10108.

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

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

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

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

20. Li, E., Cui, M., Peter, I.S., and Davidson, E.H. (2014). Encoding regulatory state boundaries in the pregastrular oral ectoderm of the sea urchin embryo. Proc. Natl. Acad. Sci. *111*, E906–E913.

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

22. Su, Y.-H., Li, E., Geiss, G.K., Longabaugh, W.J.R., Krämer, A., and 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.

23. McClay, D.R. (2011). Evolutionary crossroads in developmental biology: sea urchins. Development *138*, 2639–2648.

24. Valencia, J.E., Feuda, R., Mellott, D.O., Yaguchi, S., Burke, R.D., and Peter, I.S. Photoreceptor specification in the absence of eye master regulators.

25. Andrikou, C., Iovene, E., Rizzo, F., Oliveri, P., and Arnone, M.I. (2013). Myogenesis in the sea urchin embryo: the molecular fingerprint of the myoblast precursors. Evodevo *4*, 33.

26. Annunziata, R., Perillo, M., Andrikou, C., Cole, A.G., Martinez, P., and Arnone, M.I. (2014). Pattern and process during sea urchin gut morphogenesis: The regulatory landscape: Pattern and Process During Sea Urchin Gut Morphogenesis. genesis *52*, 251–268.

27. Garner, S., Zysk, I., Byrne, G., Kramer, M., Moller, D., Taylor, V., and Burke, R.D. (2016). Neurogenesis in sea urchin embryos and the diversity of deuterostome neurogenic mechanisms. Development *143*, 286–297.

28. Perillo, M., Wang, Y.J., Leach, S.D., and Arnone, M.I. (2016). A pancreatic exocrine-like cell regulatory circuit operating in the upper stomach of the sea urchin Strongylocentrotus purpuratus larva. BMC Evol. Biol. *16*. Available at: http://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-016-0686-0 [Accessed August 24, 2017].

29. Sun, Z., and Ettensohn, C.A. (2014). Signal-dependent regulation of the sea urchin skeletogenic gene regulatory network. Gene Expr. Patterns *16*, 93–103.

C h a p t e r 2

BILATERAL CILIARY PHOTORECEPTORS IN SEA URCHIN LARVAE EXPRESS A PARTIALLY CONSERVED REGULATORY STATE

Jonathan Valencia[*,1], Roberto Feuda[*,1], Dan O. Mellott[2], Shunsuke Yaguchi[3], Robert D. Burke[2,*], Isabelle S. Peter[1,*]

[1] Division of Biology and Biological Engineering, MC156-29, California Institute of Technology, Pasadena, CA 91125
[2] Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada
[3] Shimoda Marine Research Center, University of Tsukuba, Shimoda, Shizuoka, Japan

* equal contribution

2.1 Abstract

The evolutionary history of animal eyes has received considerable attention, in part because comparative analyses of morphology, development, and molecular determinants offer seemingly contradictory evidence [1-4]. Molecular data from diverse clades across metazoan phylogeny will thus contribute to resolving these contradictions. Here we show that sea urchin larvae possess bilateral clusters of ciliary photoreceptors localized adjacent to the mouth and nervous system. Using immunostaining and optogenetics we determined that these photoreceptors are light-sensitive neurons expressing a G_O-opsin. Analysis of regulatory gene expression indicates that these photoreceptors express *rx*, *otx*, and *six3* encoding orthologs of transcription factors expressed in vertebrate eyes [5]. However, *pax6*, *six1/2*, *eya*, and *dac* (PSED), regulatory genes important for eye development in flies and vertebrates [6], are not expressed in sea urchin ciliary photoreceptors. Instead, these PSED genes are co-expressed in the hydropore canal of sea urchin larvae. These results together with molecular data from other deuterostomes suggest that the PSED circuit, although encoded in the genome of ancestral deuterostomes, is not required for the differentiation of ciliary photoreceptor cell types in basal deuterostomes but has been recruited to the upstream control of eye development during chordate evolution. Thus the PSED circuit represents an example of a homologous regulatory circuit that has been repeatedly co-opted into independently evolving gene regulatory networks.

2.3 Results

The transcription factor Rx plays an important role in the specification of ciliary photoreceptors in vertebrates and controls the expression of *pax6* and *opsin* genes during the development of eyes [7, 8]. Rx is also expressed in ciliary photoreceptors of the marine annelid *Platynereis dumerilii* [9], indicating that this transcription factor might play a conserved role in the formation of ciliary photoreceptors. We analyzed the expression of *rx* in 72h larvae of the purple sea urchin *Strongylocentrotus purpuratus* by whole mount in situ hybridization (WMISH), showing that this gene is expressed in bilateral clusters of 2-3 cells on the oral side of the neurogenic apical organ (Figure 2.1A). The particular location of these cells, and the expression of *rx*, suggested that these cells may correspond to photoreceptor cells. We decided to test this hypothesis based on four requirements for functional directional photoreceptor cells: i) expression of a photosensitive Opsin, ii) neuronal cell type identity, iii) response to light stimulation, and iv) presence of shading pigments.

To determine if the putative larval photoreceptor cells express Opsins, we analyzed expression of *opsin* genes based on available transcriptome data [10]. Out of eight *opsin* genes encoded in the genome, only *opsin2*, and *opsin3.2* showed expression at the larval stage (Figure S2.1A). However, when analyzed by WMISH, only *opsin 3.2* showed expression in the bilateral clusters similar to *rx*, which is consistent with earlier results (Figures 2.1B and S2.1B) [11]. To confirm the co-expression of *rx* and *opsin3.2* we performed double fluorescence WMISH, showing that these two genes are indeed expressed

in the same cells (Figure 2.1C). A phylogenetic analysis showed that Opsin3.2 belongs to the RGR/Go class of opsins that are also expressed in ciliary photoreceptors of scallops (Figure S2.2) [12-14]. According to this analysis, Opsin3.2 is an ortholog of Opsin1 of the annelid *Platynereis* that is sensitive to cyan light, wavelengths important for marine life [15].

The bilateral clusters of *opsin3.2* expressing cells are on the oral side of the apical neurogenic organ. To test whether these putative photoreceptor cells correspond to neurons, we generated rat polyclonal antibodies against sea urchin Opsin3.2. Expression of Opsin3.2 protein was analyzed by immunohistochemistry, showing expression in bilateral clusters of cells, similar to *opsin3.2* RNA (Figures 2.2A and B). To analyze whether Opsin expressing cells belong to neuronal cell types, we performed co-immunostaining using antibodies against Opsin 3.2 and the pan-neuronal protein synaptotagmin B in 72h larvae. The results demonstrate that Opsin3.2 is expressed in a subset of synaptotagmin B expressing cells (Figure 2.2C). Thus the putative photoreceptor cells indeed correspond to neuronal cell types. As expected, co-immunostaining of Opsin3.2 and Serotonin shows expression in separate cells, indicating that the photoreceptor neurons are not serotonergic (Figure S2.3A).

Ciliary and rhabdomeric photoreceptor cells possess morphologically distinct cell surface structures that increase the photosensitive area, cilia and microvilli, respectively [4]. We used antibodies against α-tubulin to detect the presence of microtubules, the structural component of cilia, on the surface of the sea urchin photoreceptor cells. Our results show that Opsin3.2 and α-tubulin co-localize within the immotile cilia on the surface of

photoreceptor cells (Figure 2.2D). Taken together, the presence of Opsin presenting cilia and the expression of Rx and an RGR/Go Opsin indicates that the bilateral clusters of cells in sea urchin larvae represent ciliary photoreceptors. Interestingly, the tube feet of adult sea urchins have been reported to include rhabdomeric photoreceptors expressing an r-opsin [16]. Sea urchins therefore deploy alternative photoreceptor cell types during larval and adult stages.

To test if the ciliary photoreceptor cells are capable of detecting light, we used a fluorescent calcium sensor, GCamP6, to monitor the neuronal activity of photoreceptors in response to light [17]. In larvae injected with GCamP6 mRNA, photoreceptor cells showed oscillating fluorescence (Figures 2.2E and S2.3B). However, when these larvae were stimulated with 2s of white light, photoreceptor cell fluorescence sharply increased, followed by a refractory period in which exposure to light did not induce calcium release (Figures 2.2G and S2.3B; S2Movie1). As a control, we monitored fluorescent activity in cells that do not correspond to photoreceptors. Although oscillating fluorescence was detectable in these cells, exposure to light did not result in a refractory period, and these cells continued to emit fluorescence (Figure S2.3B). These results support the conclusion that the larval ciliary photoreceptors are light sensitive.

Photoreceptors are often associated with shading pigments enabling directional photoreception [4]. To test whether shading pigments are present near the larval photoreceptors, we performed immunostaining using SP1 antibodies detecting pigmented immunocytes. Indeed, pigment cells are found within 2-3 cell diameters of the

photoreceptors, embedded in the ectodermal epithelium (Figure S2.3C). Furthermore, pigment cells are interspersed within the entire aboral ectoderm, but absent from the oral ectoderm, potentially lowering the intensity of light perceived from the aboral side of the larva. Both mechanisms may support directional light perception.

An independent genome-wide survey of regulatory gene expression revealed additional nine regulatory genes that are expressed in a pattern similar to *rx* and *opsin3.2* in the 72h sea urchin larva: *awh*, *six3*, *foxg*, *hbn*, *otx*, *soxb2*, *tbx2/3*, *zic*, and *nkx2.1* (Figures 2.3A and S2.4). We confirmed the co-expression of *awh* and *six3* with *opsin3.2* by double fluorescent WMISH (Figure S2.5). Thus in addition to Rx, the sea urchin ciliary photoreceptors share the expression of Six3 and Otx with vertebrate ciliary photoreceptors. When analyzing gene expression during earlier development, we found that expression of *opsin3.2* is first activated at 60h (Figure S2.1A), indicating the earliest onset of photoreceptor differentiation. To determine the developmental origin of photoreceptor cells, we analyzed the expression of *rx* and other regulatory genes at earlier stages of development. The results show that *rx* is expressed broadly throughout the apical domain at 24h, as shown earlier [18], and expression becomes specific to photoreceptors by 60h (Figure S2.4). Similarly, *six3*, *nkx2.1*, *zic*, and *awh* show expression in photoreceptors by 60h with earlier expression in the apical domain. These results suggest that photoreceptors derive from precursors of the apical neurogenic domain, and become distinctly specified as photoreceptors by 60h when they activate expression of *opsin3.2*.

Given their important role in the specification of ciliary and rhabdomeric photoreceptors throughout bilateria, we analyzed the expression of *pax6*, *six1/2*, *dach*, and *eya*, in 72h sea urchin larvae by WMISH. Surprisingly, while all four regulatory genes are co-expressed in the hydropore canal, a mesodermal derivative for filtering and secretion of coelomic fluid, their expression was absent from larval photoreceptors (Figures 2.3B,C and S2.6;) [19, 20]. To analyze whether expression of these genes occurred during earlier stages of photoreceptor specification, we tested *pax6*, *six1/2*, and *eya* expression at 24h, 36h, 48h, and 60h (fig. S6). Our results indicate that the PSED module does not operate during development or differentiation of the sea urchin ciliary photoreceptors, even though a functional PSED module is encoded in the genome and expressed in the hydropore canal.

2.4 Figures

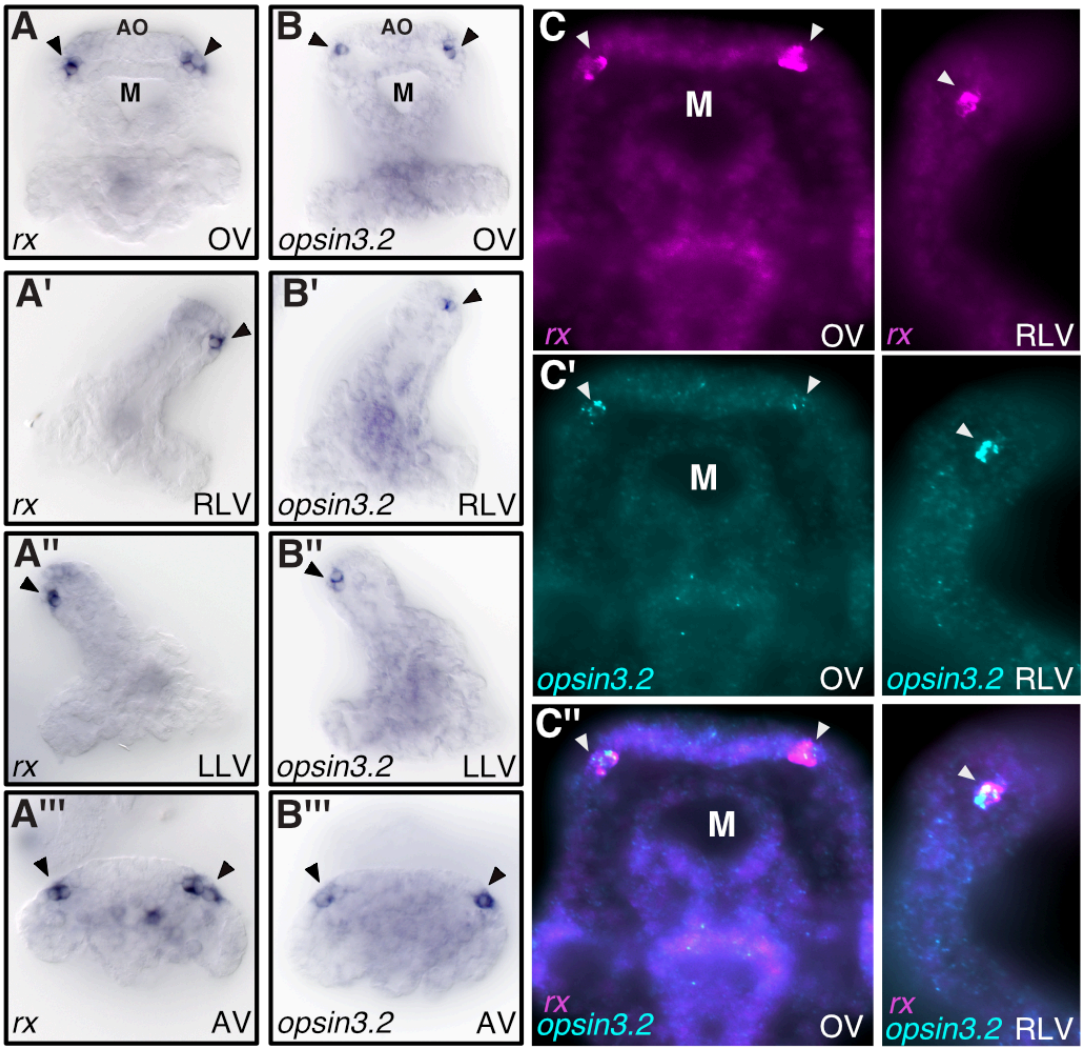


Figure 2.1: Spatial expression of rx and opsin 3.2 in putative PRCs of sea urchin larvae at 72h. (A-A’’’) WMISH showing rx expression in putative PRCs on oral side of apical organ. (B-B’’’) WMISH for opsin3.2 expression. (C-C’’) Double fluorescent WMISH for rx (magenta, C) and opsin3.2 (green, C’), overlay shown in C’’. Arrowhead indicates PRCs. M, mouth; AO, apical organ; OV, oral view; RLV, right lateral view; LLV, left lateral view; AV, apical view.

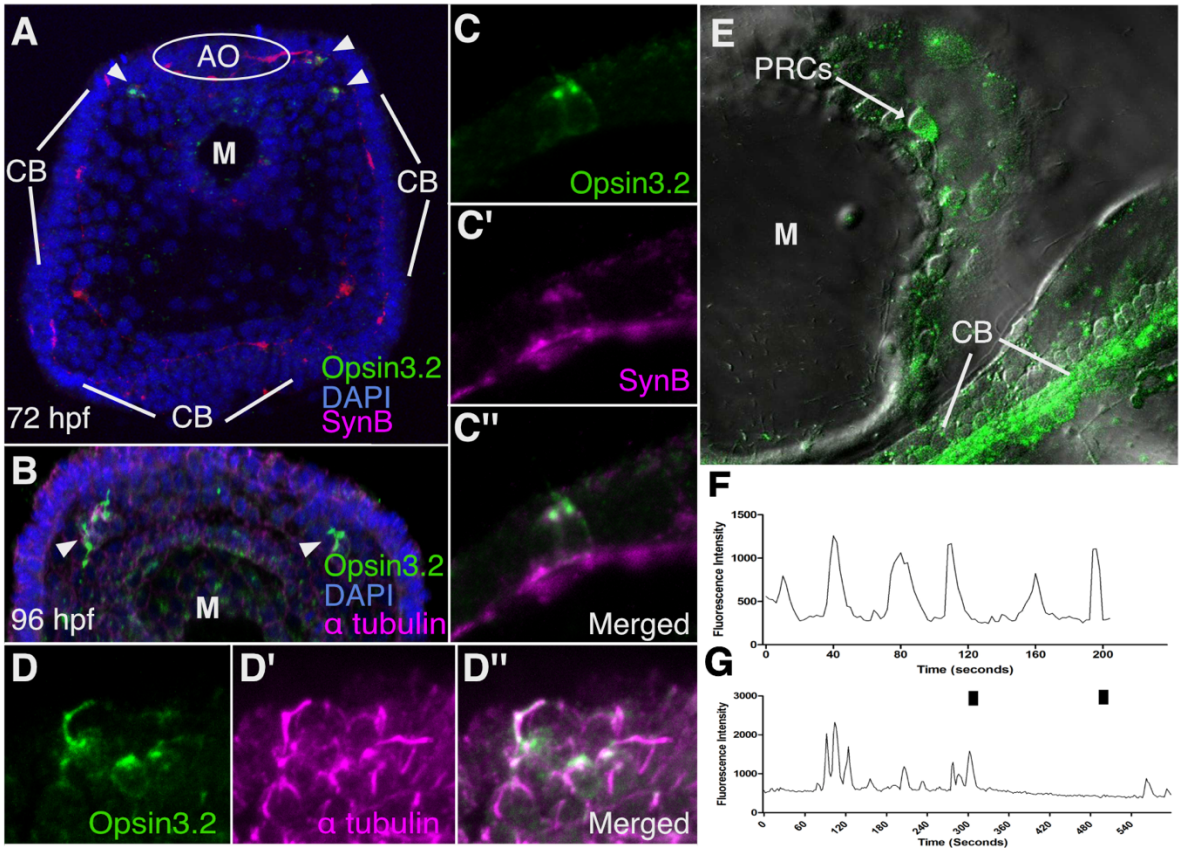


Figure 2.2: Immunostaining showing Opsin 3.2 expression in neuronal ciliary photoreceptors. Confocal laser scanning images of whole mount S. purpuratus larvae showing immunolocalization (A) of Opsin 3.2 and Synaptotagmin B at 72h, and (B) of Opsin3.2 and □ tubulin at 96 h. (C) Confocal images of co-immunostaining for Opsin3.2 (green, C) and Synaptotagmin B (magenta, C'), showing co-expression (C''). (D) Confocal images of co-immunostaining for Opsin3.2 with □-tubulin (D') showing co-localization in cell surface cilia of PRCs (D''). (E) 96h larva expressing GCamP6 fluorescent calcium reporter in which a PRC (PRC, arrow) is identified by the presence of a short, immotile cilium (projecting downward in image) and a fine axonal projection at the opposite pole of the cell. (F) Z-axis profile plots of fluorescence in individual PRC showing oscillation of cytoplasmic calcium in an individual PRC with 40 sec frequency. (G) Z-axis profile plot of fluorescence in a PRC showing initial oscillation, an immediate response to a 2 sec flash of white light (black bar), a 3 to 4 min refractory period in which there is no response to a second 2 sec flash of light (black bar), before resuming oscillatory behavior. M, mouth; AO, apical organ; CB, ciliated band.

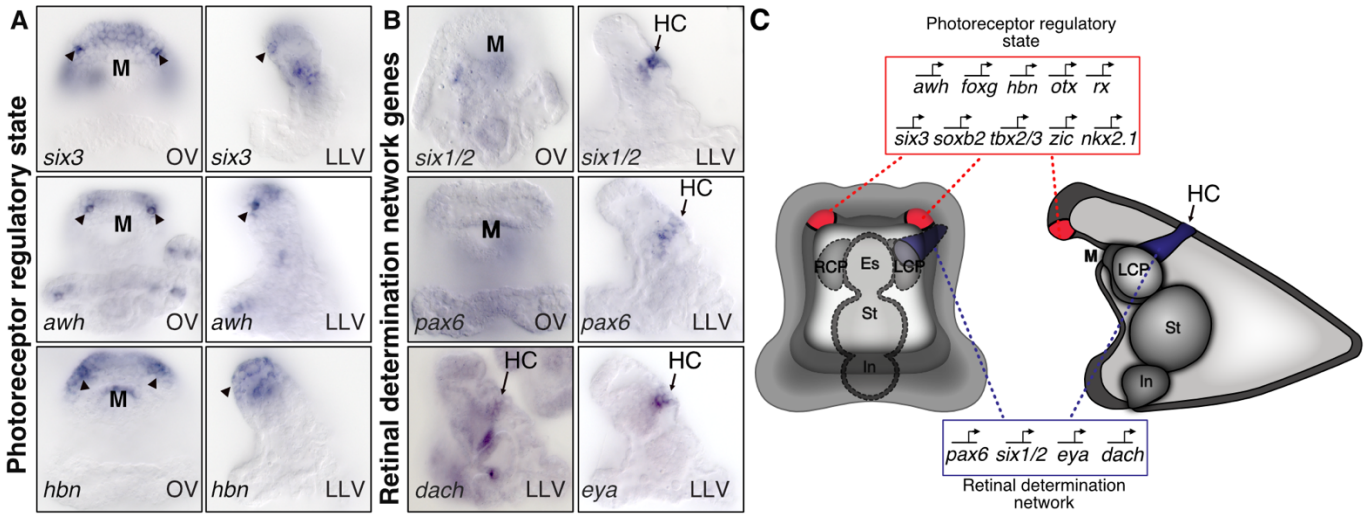


Figure 2.3: Expression of photoreceptor regulatory state and RDN genes in 72h larvae. (A) WMISH showing expression of six3, awh, and hbn in PRCs. Additional regulatory genes expressed in PRCs are shown in fig. S4. (B)WMISH for pax6, six1/2, dach, and eya showing expression in the hydropore canal (HC). (C) Summary diagram of sea urchin larva showing expression of regulatory genes in PRCs and expression of indicated RDN genes in the hydropore canal. OV, oral view; LLV, left lateral view.

2.5 Discussion

Taken together, our results show that sea urchin larvae possess bilateral clusters of light-sensitive photoreceptor cells that phenotypically belong to the class of ciliary photoreceptors. These larval photoreceptors are located in the vicinity to the neurogenic apical organ and gene expression data suggest that these cells are developmentally derived from cells of the apical plate ectoderm, reminiscent of the neural ectoderm origin of photoreceptors in Ciona and Amphioxus [21]. However, while the majority of ciliary photoreceptors characterized so far express c-opsins, sea urchin ciliary photoreceptors express an RGR/Go opsin, a class of opsins also expressed in the ciliary photoreceptors of scallops [1, 14]. This type of photoreceptor cell has been associated with relatively low resolution light detection.

The combination of transcription factors expressed in the sea urchin ciliary photoreceptors show a molecular signature that is similar to other Deuterostome ciliary photoreceptors. Most prominently, Rx has been shown to be expressed in ciliary photoreceptors of many species, including Amphioxus, Ciona, Platynereis, and in vertebrates [2, 8, 22-24]. In addition, photoreceptors express *otx*, which is an ortholog of *otx2* and *crx* that are both involved in photoreceptor specification and differentiation in vertebrates [25]. Otx is also involved in photoreceptor specification of amphioxus [2]. *Six3* is broadly expressed in anterior neural plate regions including photoreceptors of Amphioxus [2] and the mouse retina and its ortholog *six7* is required for the development of photoreceptors and expression of opsin in zebrafish [26, 27]. Furthermore, Tbx2b is required for the specification of UV-cone cells in

the zebrafish retina [28]. Thus a significant fraction of the regulatory state expressed in sea urchin larval photoreceptors is expressed also in ciliary photoreceptors in other deuterostomes.

Remarkably, a set of transcription factors involved in eye development in mice and flies are not expressed in the sea urchin larval photoreceptors. This includes the PSED factors Pax6, Six1/2, Eya, and Dach. Pax6 is one of the few transcription factors with the capacity to overwrite other developmental programs and induce ectopic eye phenotypes when expressed ectopically in flies and vertebrates [29, 30]. Pax6 is also expressed in photoreceptors of amphioxus and ciona [2, 31]. However, although the PSED factors are expressed in retinal progenitors and required for eye development in vertebrates, there is no evidence for a function of these factors in the differentiation of ciliary photoreceptor cell types in the vertebrate retina. Thus although during early eye development Pax6 is involved in the specification of retinal progenitors downstream of Rx, during later development Pax6 functions in the specification of horizontal and amacrine cells, and inhibits the differentiation of photoreceptor cells [32].

Besides their role during early eye development, various combinations of PSED factors control developmental processes in vertebrates including kidney development and specification of somitic muscle [33]. In amphioxus, PSED factors are co-expressed in several cell fates but not in photoreceptors [34], and similarly, they are expressed in sea urchin coelomic pouches [20] and, as we show here, the larval hydropore canal. PSED factors have been associated with basic cellular processes such as proliferation and apoptosis. In

Drosophila eyes for example, Six1/2 and Eya control expression of *stringer*, a gene involved in cell cycle control.

Thus in conclusion these data indicate that there is a common regulatory toolkit used for the differentiation of ciliary photoreceptors, including Rx, Otx and its ortholog Crx, Six3, and possibly Tbx2. The retinal determination network (RDN) however seems not be involved in the specification of ciliary photoreceptors in basal deuterostomes nor in the differentiation of these cells in vertebrates, although a functional PSED circuit is broadly shared among bilateria and must have been encoded in the genome of ancestral deuterostomes. Indeed, some components of the RDN are present and co-expressed in sponges [35, 36], suggesting that assembly of the RDN circuit precedes the appearance of photoreceptors. These results support the view that the PSED circuit has been co-opted into the gene regulatory network controlling specification of retinal progenitors during the evolution of complex vertebrate eyes. Similar co-options of entire circuits may have occurred also in the proximal/distal axis of insect and vertebrate legs, and in the co-option of the hox patterning system to the vertebrate limb [37-39]. But while the co-opted circuits are homologous and preserve some of their original structure and function within novel developmental contexts, the GRNs into which they were co-opted have evolved independently.

2.6 References

1. Arendt, D. (2003). Evolution of eyes and photoreceptor cell types. Int J Dev Biol *47*, 563-571.

2. Vopalensky, P., Pergner, J., Liegertova, M., Benito-Gutierrez, E., Arendt, D., and Kozmik, Z. (2012). Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye. Proc Natl Acad Sci U S A *109*, 15383-15388.

3. Vopalensky, P., and Kozmik, Z. (2009). Eye evolution: common use and independent recruitment of genetic components. Philos Trans R Soc Lond B Biol Sci *364*, 2819-2832.

4. Nilsson, D.E. (2009). The evolution of eyes and visually guided behaviour. Philos Trans R Soc Lond B Biol Sci *364*, 2833-2847.

5. Viets, K., Eldred, K., and Johnston, R.J., Jr. (2016). Mechanisms of Photoreceptor Patterning in Vertebrates and Invertebrates. Trends Genet *32*, 638-659.

6. Kumar, J.P. (2009). The molecular circuitry governing retinal determination. Biochim. Biophys. Acta *1789*, 306-314.

7. Bailey, T.J., El-Hodiri, H., Zhang, L., Shah, R., Mathers, P.H., and Jamrich, M. (2004). Regulation of vertebrate eye development by Rx genes. Int J Dev Biol *48*, 761-770.

8. Pan, Y., Martinez-De Luna, R.I., Lou, C.H., Nekkalapudi, S., Kelly, L.E., Sater, A.K., and El-Hodiri, H.M. (2010). Regulation of photoreceptor gene expression by the retinal homeobox (Rx) gene product. Dev Biol *339*, 494-506.

9. Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A.W., and Wittbrodt, J. (2004). Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science *306*, 869-871.

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

11. Valero-Gracia, A., Petrone, L., Oliveri, P., Nilsson, D.-E., and Arnone, M.I. (2016). Non-directional photoreceptors in the pluteus of Strongylocentrotus purpuratus. Frontiers in Ecology and Evolution *4*, 127.

12. Feuda, R., Hamilton, S.C., McInerney, J.O., and Pisani, D. (2012). Metazoan opsin evolution reveals a simple route to animal vision. Proc Natl Acad Sci U S A *109*, 18868-18872.

13. D'Aniello, S., Delroisse, J., Valero-Gracia, A., Lowe, E.K., Byrne, M., Cannon, J.T., Halanych, K.M., Elphick, M.R., Mallefet, J., Kaul-Strehlow, S., et al. (2015). Opsin evolution in the Ambulacraria. Mar Genomics *24 Pt 2*, 177-183.

14. Kojima, D., Terakita, A., Ishikawa, T., Tsukahara, Y., Maeda, A., and Shichida, Y. (1997). A novel Go-mediated phototransduction cascade in scallop visual cells. J Biol Chem *272*, 22979-22982.

15. Gühmann, M., Jia, H., Randel, N., Verasztó, C., Bezares-Calderón, L.A., Michiels, N.K., Yokoyama, S., and Jékely, G. (2015). Spectral tuning of phototaxis by a go-opsin in the rhabdomeric eyes of Platynereis. Current Biology *25*, 2265-2271.

16. Ullrich-Lüter, E.M., Dupont, S., Arboleda, E., Hausen, H., and Arnone, M.I. (2011). Unique system of photoreceptors in sea urchin tube feet. Proc Natl Acad Sci U S A *108*, 8367-8372.

17. Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295-300.

18. Burke, R.D., Angerer, L.M., Elphick, M.R., Humphrey, G.W., Yaguchi, S., Kiyama, T., Liang, S., Mu, X., Agca, C., Klein, W.H., et al. (2006). A genomic view of the sea urchin nervous system. Dev Biol *300*, 434-460.

19. Luo, Y.J., and Su, Y.H. (2012). Opposing nodal and BMP signals regulate left-right asymmetry in the sea urchin larva. PLoS Biol *10*, e1001402.

20. Martik, M.L., and McClay, D.R. (2015). Deployment of a retinal determination gene network drives directed cell migration in the sea urchin embryo. 1-19.

21. Oonuma, K., Tanaka, M., Nishitsuji, K., Kato, Y., Shimai, K., and Kusakabe, T.G. (2016). Revised lineage of larval photoreceptor cells in Ciona reveals archetypal collaboration between neural tube and neural crest in sensory organ formation. Dev Biol *420*, 178-185.

22. Mathers, P.H., Grinberg, A., Mahon, K.A., and Jamrich, M. (1997). The Rx homeobox gene is essential for vertebrate eye development. Nature *387*, 603-607.

23. Swaroop, A., Kim, D., and Forrest, D. (2010). Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. Nat. Rev. Neurosci. *11*, 563-576.

24. D'Aniello, S., D'Aniello, E., Locascio, A., Memoli, A., Corrado, M., Russo, M.T., Aniello, F., Fucci, L., Brown, E.R., and Branno, M. (2006). The ascidian homolog of the vertebrate homeobox gene Rx is essential for ocellus development and function. Differentiation *74*, 222-234.

25. Musser, J.M., and Arendt, D. (2017). Loss and gain of cone types in vertebrate ciliary photoreceptor evolution. Dev Biol *431*, 26-35.

26. Sotolongo-Lopez, M., Alvarez-Delfin, K., Saade, C.J., Vera, D.L., and Fadool, J.M. (2016). Genetic Dissection of Dual Roles for the Transcription Factor six7 in Photoreceptor Development and Patterning in Zebrafish. PLoS Genet *12*, e1005968.

27. Ogawa, Y., Shiraki, T., Kojima, D., and Fukada, Y. (2015). Homeobox transcription factor Six7 governs expression of green opsin genes in zebrafish. Proc Biol Sci *282*, 20150659.

28. Alvarez-Delfin, K., Morris, A.C., Snelson, C.D., Gamse, J.T., Gupta, T., Marlow, F.L., Mullins, M.C., Burgess, H.A., Granato, M., and Fadool, J.M. (2009). Tbx2b is required for ultraviolet photoreceptor cell specification during zebrafish retinal development. Proc Natl Acad Sci U S A *106*, 2023-2028.

29. Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science (New York, N.Y *267*, 1788-1792.

30. Chow, R.L., Altmann, C.R., Lang, R.A., and Hemmati-Brivanlou, A. (1999). Pax6 induces ectopic eyes in a vertebrate. Development (Cambridge, England) *126*, 4213-4222.

31. Irvine, S.Q., Fonseca, V.C., Zompa, M.A., and Antony, R. (2008). Cis-regulatory organization of the Pax6 gene in the ascidian Ciona intestinalis. Dev Biol *317*, 649-659.

32. Remez, L.A., Onishi, A., Menuchin-Lasowski, Y., Biran, A., Blackshaw, S., Wahlin, K.J., Zack, D.J., and Ashery-Padan, R. (2017). Pax6 is essential for the generation of late-born retinal neurons and for inhibition of photoreceptor-fate during late stages of retinogenesis. Dev Biol *432*, 140-150.

33. Heanue, T.A., Reshef, R., Davis, R.J., Mardon, G., Oliver, G., Tomarev, S., Lassar, A.B., and Tabin, C.J. (1999). Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation. Genes Dev *13*, 3231-3243.

34. Kozmik, Z., Holland, N.D., Kreslova, J., Oliveri, D., Schubert, M., Jonasova, K., Holland, L.Z., Pestarino, M., Benes, V., and Candiani, S. (2007). Pax-Six-Eya-Dach network during amphioxus development: conservation in vitro but context specificity in vivo. Dev Biol *306*, 143-159.

35. Fortunato, S.A., Leininger, S., and Adamska, M. (2014). Evolution of the Pax-Six-Eya-Dach network: the calcisponge case study. Evodevo *5*, 23.

36. Rivera, A., Winters, I., Rued, A., Ding, S., Posfai, D., Cieniewicz, B., Cameron, K., Gentile, L., and Hill, A. (2013). The evolution and function of the Pax/Six regulatory network in sponges. Evol Dev *15*, 186-196.

37. Peter, I.S., and Davidson, E.H. (2015). Genomic Control Process, Development and Evolution, (Academic Press/Elsevier).

38. Lemons, D., Fritzenwanker, J.H., Gerhart, J., Lowe, C.J., and McGinnis, W. (2010). Co-option of an anteroposterior head axis patterning system for proximodistal patterning of appendages in early bilaterian evolution. Developmental biology *344*, 358-362.

39. Shubin, N., Tabin, C., and Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. Nature *388*, 639-648.

40. Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064-1071.

41. Huang, S., Chen, Z., Yan, X., Yu, T., Huang, G., Yan, Q., Pontarotti, P.A., Zhao, H., Li, J., Yang, P., et al. (2014). Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. Nat Commun *5*, 5896.

42. Dehal, P., Satou, Y., Campbell, R.K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D.M., et al. (2002). The draft genome

of Ciona intestinalis: insights into chordate and vertebrate origins. Science (New York, N.Y *298*, 2157-2167.

43. Small, K.S., Brudno, M., Hill, M.M., and Sidow, A. (2007). A haplome alignment and reference sequence of the highly polymorphic Ciona savignyi genome. Genome Biol *8*, R41.

44. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

45. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res *33*, W116-120.

46. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30*, 3059-3066.

47. Le, S.Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. Mol Biol Evol *25*, 1307-1320.

48. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol *32*, 268-274.

49. Minh, B.Q., Nguyen, M.A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol *30*, 1188-1195.

50. Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst Biol *60*, 685-699.

51. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286-2288.

52. Ransick, A., Ernst, S., Britten, R.J., and Davidson, E.H. (1993). Whole mount in situ hybridization shows Endo 16 to be a marker for the vegetal plate territory in sea urchin embryos. Mechanisms of development *42*, 117-124.

53. Chen, J.H., Luo, Y.J., and Su, Y.H. (2011). The dynamic gene expression patterns of transcription factors constituting the sea urchin aboral ectoderm gene regulatory network. Dev Dyn *240*, 250-260.

54. Garner, S., Zysk, I., Byrne, G., Kramer, M., Moller, D., Taylor, V., and Burke, R.D. (2016). Neurogenesis in sea urchin embryos and the diversity of deuterostome neurogenic mechanisms. Development *143*, 286-297.

55. Vielkind, U., and Swierenga, S.H. (1989). A simple fixation procedure for immunofluorescent detection of different cytoskeletal components within the same cell. Histochemistry *91*, 81-88.

56. Nakajima, Y., Kaneko, H., Murray, G., and Burke, R.D. (2004). Divergent patterns of neural development in larval echinoids and asteroids. Evol Dev *6*, 95-104.

57. Gibson, A.W., and Burke, R.D. (1985). The origin of pigment cells in embryos of the sea urchin Strongylocentrotus purpuratus. Dev Biol *107*, 414-419.

Acknowledgements

This work was supported by National Institutes of Health Grant HD 037105 (to I.S.P.), and by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (2016-03737), awarded to R.D.B. The supplementary materials contain additional data.

2.7 Materials and Methods

Phylogenetic analysis. Opsin dataset was obtained by merging the sequences from [15] and [13]. Furthermore, additional opsin genes were obtained from the genomes of *Branchiostoma floridae* [40], *Branchiostome belechei* [41], *Ciona intestinalis* [42] and *Ciona savignyi* [43]. Specifically, the dataset of [12] composed by 449 sequences was used as seed and potential homologs were identified using BLASTP [44]. Each sequence with a e-value < 10^{-10} was retained a good opsin homolog. To identify opsin genes, sequences were further annotated using interproscan [45], and only sequences with retinal binding domains were considered as Opsins. The final dataset includes 232 Opsins and 10 melatonin genes that have been used to root the trees. Alignment was performed using MAFFT [46] and phylogenetic reconstruction was performed under Maximum likelihood framework and Bayesian framework under LG-G_4 [47]. The ML tree was reconstructed using iqtree [48] and nodal support was estimated using ultrafast bootstrap [49] (1000 replicates) and the SH-aLTR bootstrap [50]. Bayesian inference was performed using Phylobayes4.1 [51] with two independent runs. Convergence was evaluated using tracecomp and bpcomp packages in Phylobayes (see Phylobayes manual). Alignment and trees are available at https://github.com/RobertoFeu/Opsins_phylogeny_Valencia_et_al.

Gene amplification and probe synthesis. The primer sets used for gene amplification are listed in TableS1. Gene models generated from sea urchin transcriptome analysis were used as a reference for primer design [10] using T7 tailed primers or cloning. cDNA prepared from various developmental stages was used as a template for PCR. For cloning, PCR

products were purified and ligated into GEM-T EZ constructs. Cloned genes were PCR-amplified using the primer flanking the insert region, and PCR products were used to synthesize RNA probes for WMISH.

Whole-Mount in Situ Hybridization. The protocol for whole-mount *in situ* hybridization (WMISH) to detect spatial gene expression has been described previously [52]. Briefly, sea urchin embryos were fixed in 4% paraformaldehyde solution. The fixed embryos were incubated in hybridization buffer [50% (vol/vol) formamide, 5× SSC, 1× Denhardt's, 1 mg/mL yeast tRNA, 50 ng/mL heparin, and 0.1% tween-20] with a concentration from 1 to 2 ng/μL digoxygenin RNA probe(s) at 60 °C for 18 h. Two Post hybridization washes were performed with hybridization buffer without RNA probe, 2× SSCT (2× SSC, 0.1% tween-20), 0.2× SSCT, and 0.1× SSCT, each 20 min at 60 °C. Subsequently, 5 washes were performed with a buffer of 0.1% Tween 20, 10% MOPS (1M), 10% NaCl (5M) and 80% DEPC water. Antibody incubations were performed at room temperature with 1:2,000 diluted anti-DIG Fab (Roche). The embryos were extensively washed before staining reaction, including six times with MABT buffer (0.1 M maleic acid, 0.15 M NaCl, and 0.1% tween-20), twice with AP buffer [100 mM Tris·Cl (pH 9.5), 100 mM NaCl, 50 mM MgCl2, and 1 mM levamisole]. 5-Bromo-4-chloro-3-indolyl-phosphate (BCIP) and nitro blue tetrazolium were used for staining. Fluorescent in situ *in situ* hybridization protocol was performed as described in [53].

Antibody production. Antibody production was as previously described [54]. Antigens were made using a pET28b (+) plasmid (Novagen) for expression of 6XHis tagged proteins.

An Opsin3.2 construct was prepared with PCR (opsin-cyto:F = 5'-CAGTCATATGGCGTCGGTAAAATAAG-3', opsin-cyto:R = 5'-AGTCAAGCTTCTGTAGATTTTTAATG-3') encoding the carboxyl cytoplasmic domain (844-1494 of the coding sequence and 282-498 of the protein). High fidelity PCR was used with a cDNA template prepared from *S. purpuratus* embryos and the 650 bp product was cloned using the pGEM-T Easy system (Promega). Protein expression was induced in E. coli (BL21). Bacterial lysate was prepared and protein was solubilized in binding buffer (6 M guanidine HCl, 0.5 M NaCl, 100 mM Na2HPO4, 100 mM NaH2PO4, 10 mM imidazole, 10 mM Tris, 1 mM 2-mercaptoethanol, pH 8.0) prior to affinity purification by immobilized metal ion affinity chromatography (IMAC) using Chelex 100 Resin (Bio-Rad). Purified protein in PBS was mixed 1:1 with Freund's complete adjuvant for immunization or with Freund's incomplete adjuvant for booster injections. A rat was immunized by subcutaneous injection of 100 mg antigen in 250 µl of adjuvant, and booster injections were done 21 days and 42 days after the initial immunization. Terminal bleed via cardiac puncture was done after 52-56 days. Blood was incubated at 37°C for 45 min and then 4°C overnight. Samples were centrifuged at 1,000 XG, and serum collected. Antibody specificity was established by pre-absorbing the immune serum with an approximately equimolar preparation of the protein used to immunize the rat. Pre-absorption eliminated antibody binding to 72 and 96 h larvae.

Immunofluorescence. *S. purpuratus* embryos were collected at the desired time point and fixed for 5-10 min in 4% paraformaldehyde in PEM buffer [55]. Embryos were washed with phosphate buffered saline (PBS), blocked for 1 h in SuperBlock (Thermo), probed with primary antibody, and washed 3 times with PBS. Alexa Fluor fluorescent secondary

antibodies (Invitrogen) were used to visualize antibody labeling on a Zeiss 700 LSM (Carl Zeiss) confocal microscope. All preparations were done at 4o C. Imaging and analysis was conducted using ZEN (2009) or ImageJ (1.44) software. Adobe Photoshop (9.0.2) was used to prepare figures and adjust image contrast and brightness. Antibodies employed anti-SynB [56]; Sp1 [57]; a-tubulin (Santa Cruz Biotechnologies, sc-23948).

Fluorescent calcium sensor. The GCamP6 calcium sensor (Addgene Plasmid 40753) was subcloned to pCS2+ and plasmid was cut with Not1 and transcribed using the SP6 mMessage mMachine kit (Ambion). Eggs were prepared for microinjection as described previously (Krupke et al., 2014) and injection solutions contained water, 120 mM KCl, and 200 ng/µL RNA. Larvae expressing GCamP6 were pipetted onto NewSilane Adhesive Coated Slides (Newcomer Supply Ltd.) and trapped under a glass coverslip attached along two edges with double-sided adhesive tape (3M Inc). Paraffin oil was applied to the open edges of the coverslip to reduce evaporation and the room temperature was controlled to 16°C. Larvae were imaged with a Zeiss LSM700 with a confocal channel and a DIC channel, at 2 sec intervals for 20 to 30 min. Larvae were maintained in the dark and imaged in a darkened room. Some larvae were exposed to 2 sec bursts of white light (LED) during imaging. Movie stacks were prepared using FIJI (ImageJ 1.51k) and rendered to MP4 format in Adobe Photoshop CS6. Photoreceptor Cells were identified by location and the presence of a short non-motile cilium, and registered stacks (Stackreg) were outlined with an elliptical selection tool and Z-axis profiles plotted. Plot values were transferred to GraphPad Prism (ver. 5.04) to prepare figures.

2.8 Supplementary Information

2.8.1 Supplementary Figures and Tables

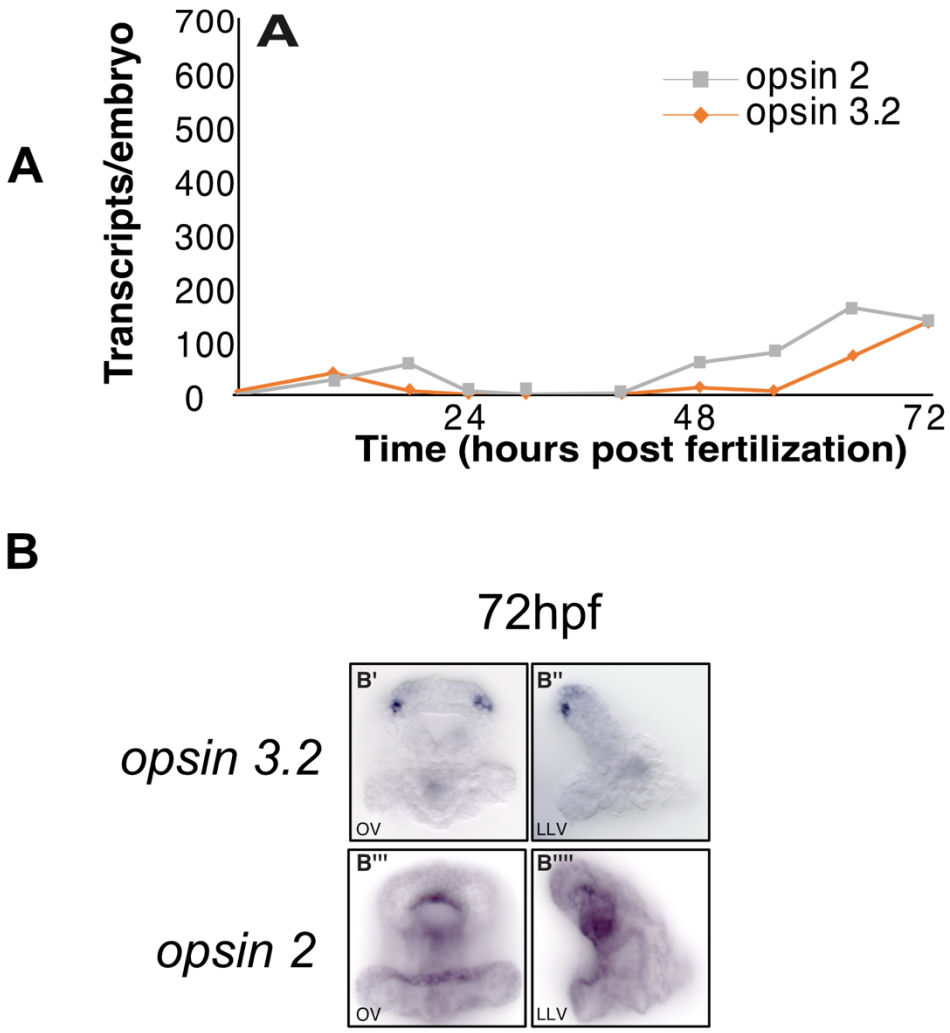


Figure S2.1: Expression of opsin genes in putative PRCs. (A) Developmental time course of opsin genes based on transcriptome data (11). (B) Spatial expression of opsin3.2 and opsin2 at 72h. OV, oral view; LLV left lateral view.

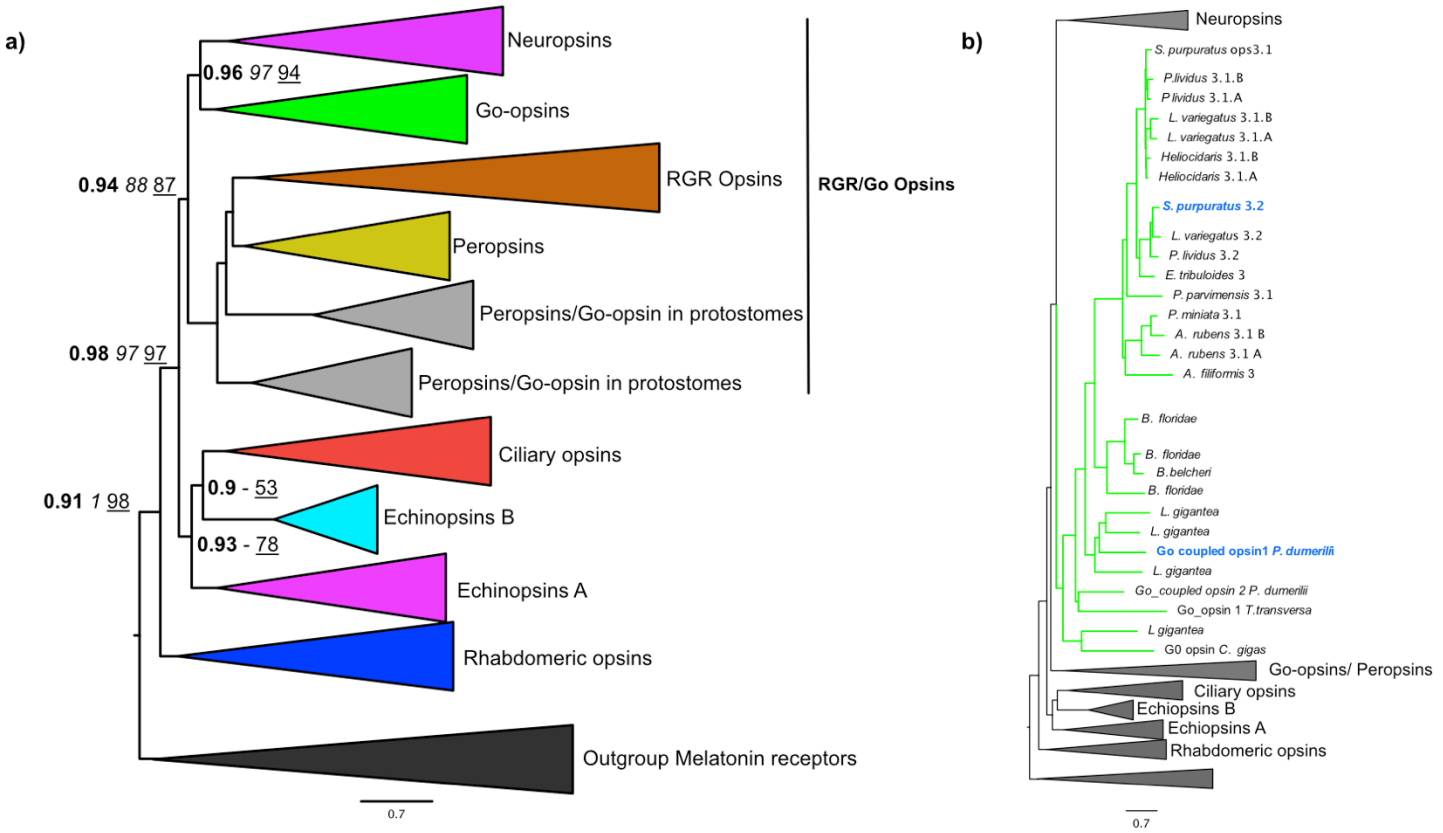


Figure S2. 2: Phylogenetic analysis of Opsins. (A) Phylogenetic tree of Opsins. Bold value indicates the Bayesian posterior probability, italic SH-aLTR bootstrap values, underlined values ultrafast bootstrap (1000 replicates). Major opsin clades are color coded. The three group-topology where R-Opsins are a sister group of C- and RGR/Go-Opsins is well supported (13, 44), however our results indicate a phylogenetic instability of Echinopsins A and B. **(B)** Phylogenetic tree as in **(A)** with focus on Go-opsins. Opsin 3.2 is a co-ortholog of Go-opsin1 of Platynereis dumerilii, which was recently shown to be sensitive to cyan blue light (15).

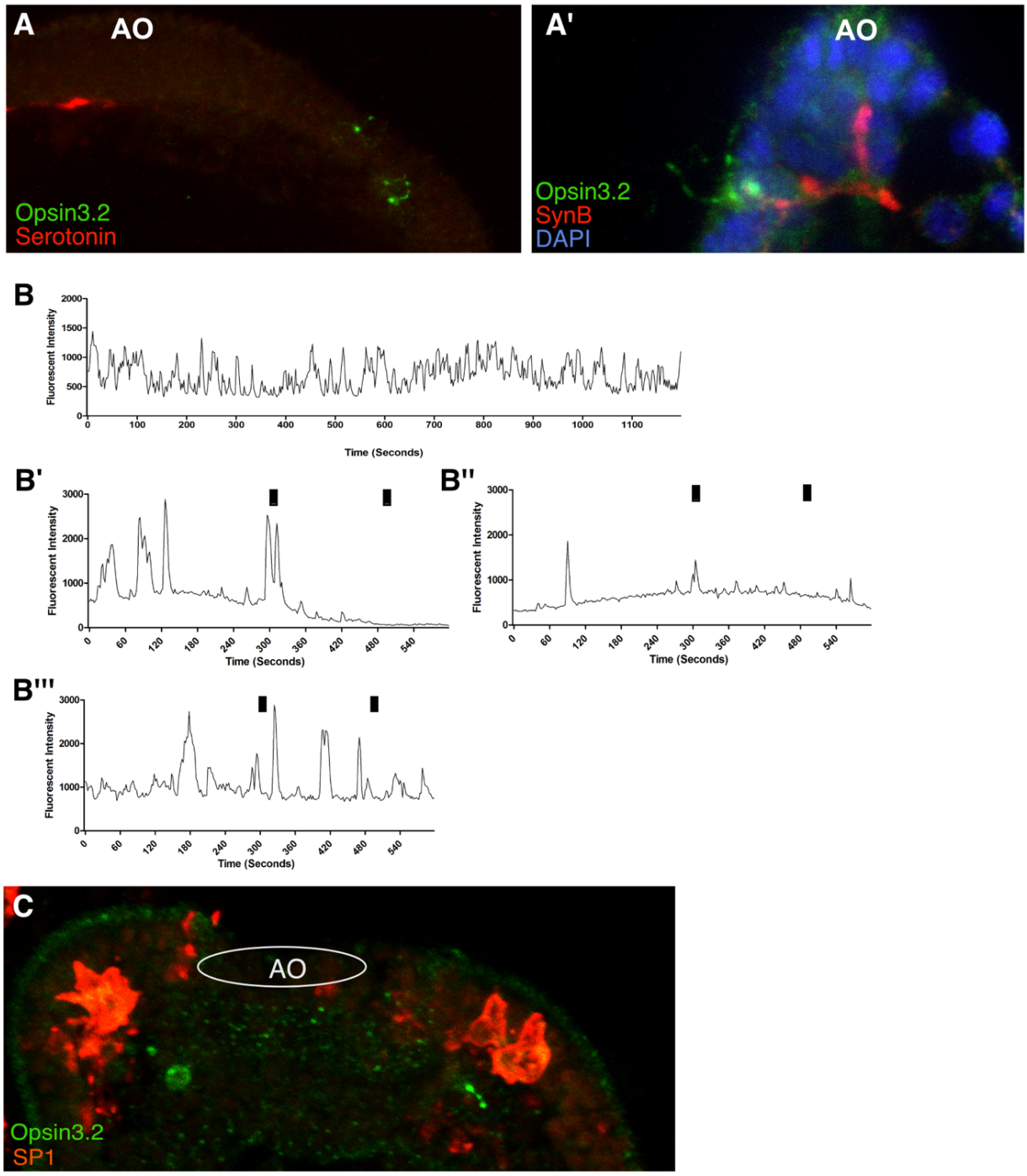


Figure S2.3: Immunostaining and calcium reporter analysis in PRCs. (A) Co-immunostaining of Opsin3.2 and Serotonin indicates that PRCs do not contain serotonin. The apical organ is known to include serotonergic neurons on the dorsal margin, and the photoreceptor cells project axons into the apical organ, but do not contain Serotonin. (A') Co-localization of Opsin3.2 and SynaptotagminB. Confocal projection from a left lateral perspective showing the projection of a SynaptotagminB containing axon into the basal neuropil of the Apical Organ (AO).

(B) Registered stacks enable z-axis profile plots of confocal images of individual PRC over 20-30 min intervals. This short profile plot shows a rhythmic oscillation of cytoplasmic calcium in an individual PRC with a 40 sec frequency. (B'B'') Z-axis profile plots of fluorescence in PRCs showing initial oscillatory behavior, an immediate response to a 2 sec flash of white light (black bar), a 3 to 4 min refractory period in which there is no response to a second 2 sec flash of light (black bar), before resuming oscillatory behavior. (B''') Z-axis profile plots of fluorescence in non-PRC cells showing oscillatory behavior of calcium release but no response to 2 sec flash of white light (black bar) and no subsequent refractory period. (C) Co-immunostaining of Opsin3.2 and pigment cell specific SP1 showing that PRCs are associated with shading pigments. Projection of confocal image stack of an apical view of a pluteus. In this orientation, cells expressing Opsin3.2 are on the ventral surface and clusters of pigment cells (Sp1) in the adjacent dorsal ectoderm. Although the pigment cells are not in direct contact with Opsin3.2 expressing cells, they are positioned so that they would shade the Opsin3.2 cells from light coming from the dorsal surface.

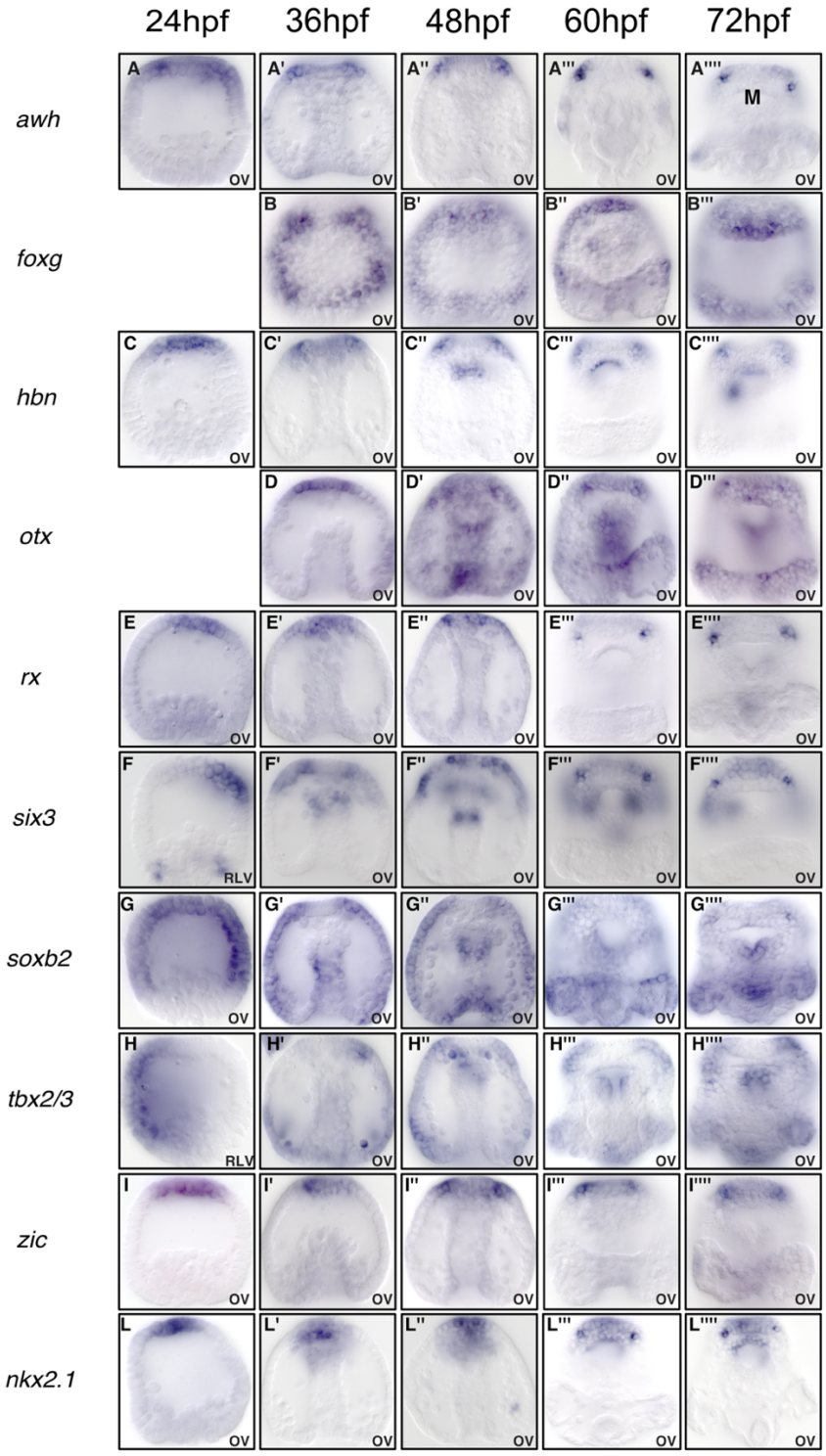


Figure S2.4: Developmental spatial expression of regulatory genes expressed in PRCs. Shown are images of embryos stained by WMISH detecting the expression of awh (A), foxg (B), hbn (C), otx (D), rx (E), six3 (F), soxb2 (G), tbx2/3 (H), zic (I), nkx2.1 (L) at 24-72h. OV, oral view.

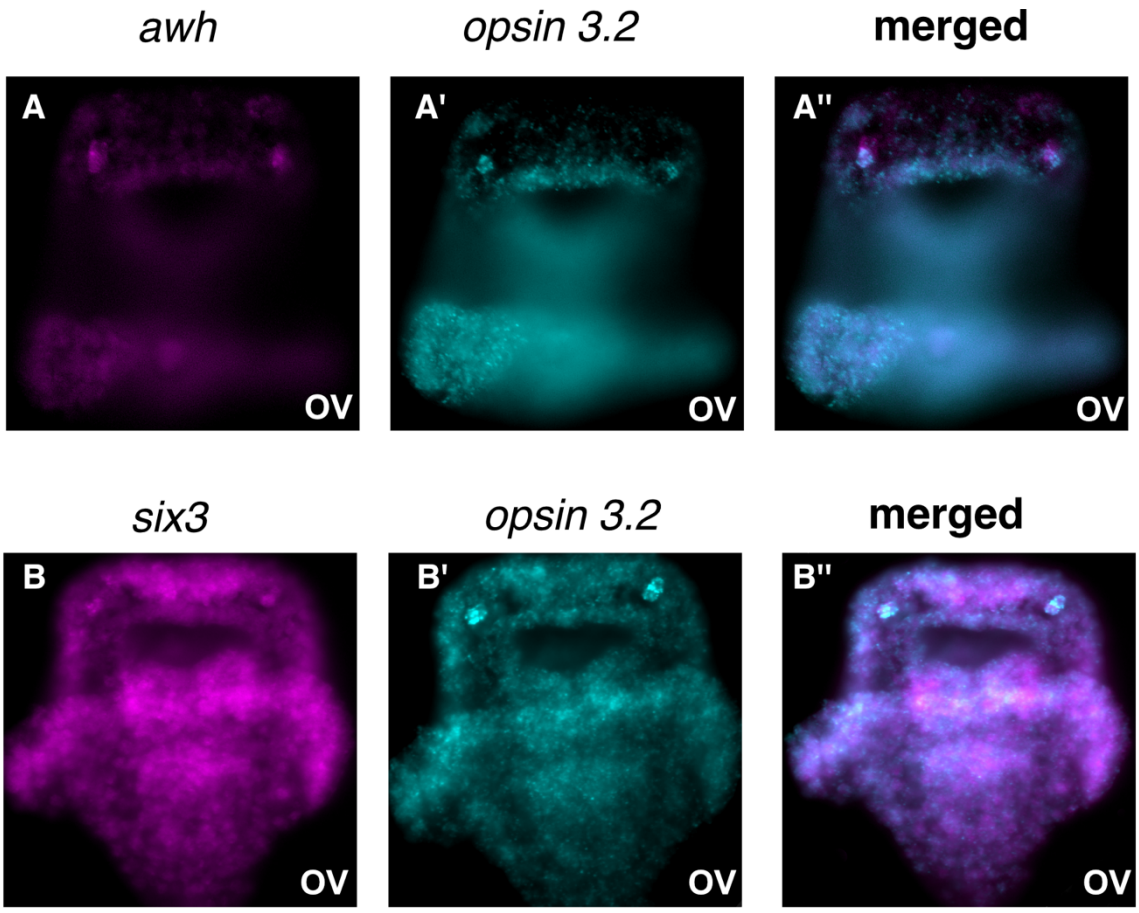


Figure S2.5: Expression of additional regulatory genes in PRCs. Double fluorescence WMISH of opsin3.2 and (A) awh or (B) six3 confirming expression in PRCs. OV, oral view.

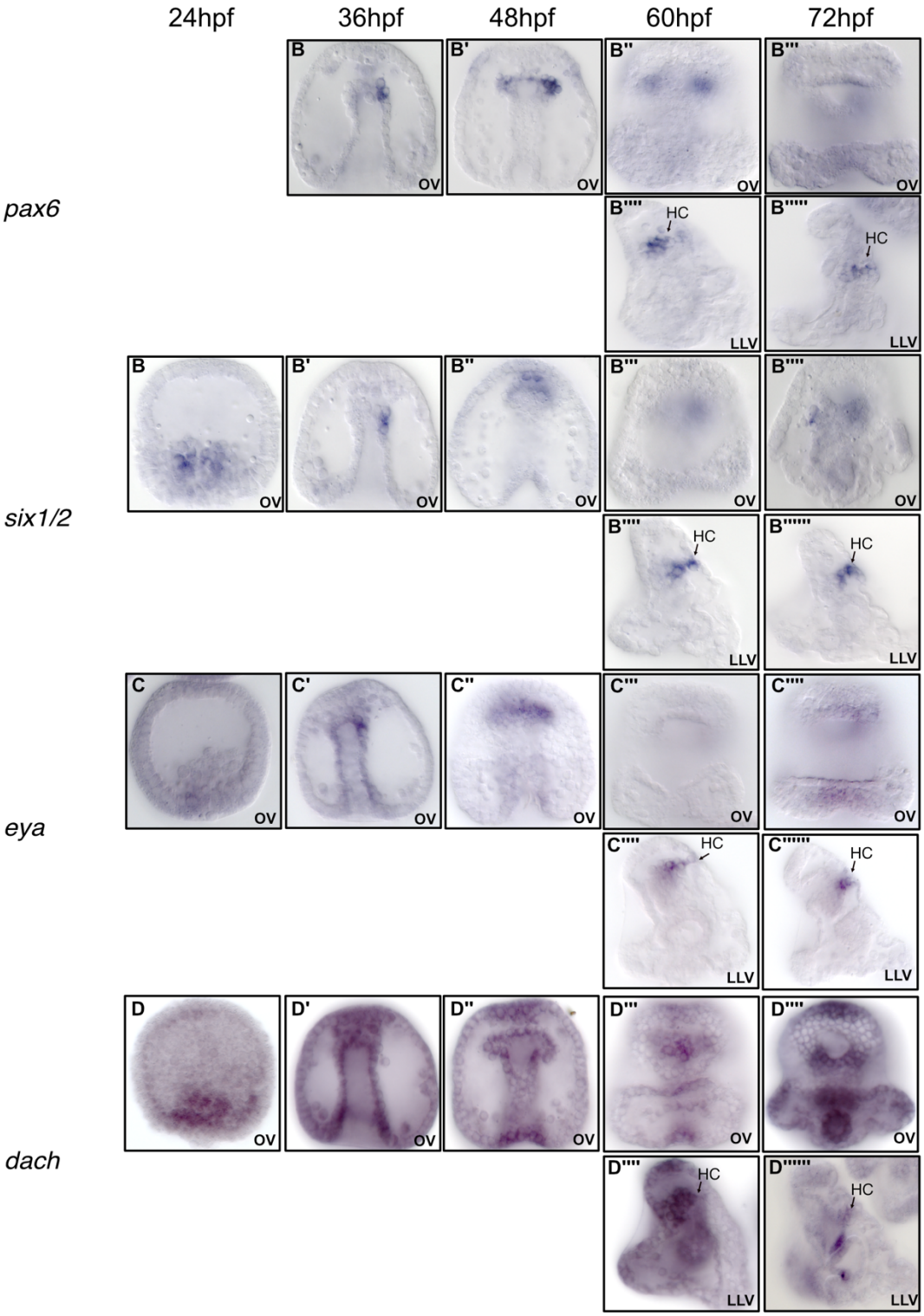


Figure S2.6: Developmental spatial expression of regulatory genes of the RDN. Images of embryos stained by WMISH for expression of pax6, six1/2, eya, and dach at 24-72h. OV, oral view; LLV left lateral view.

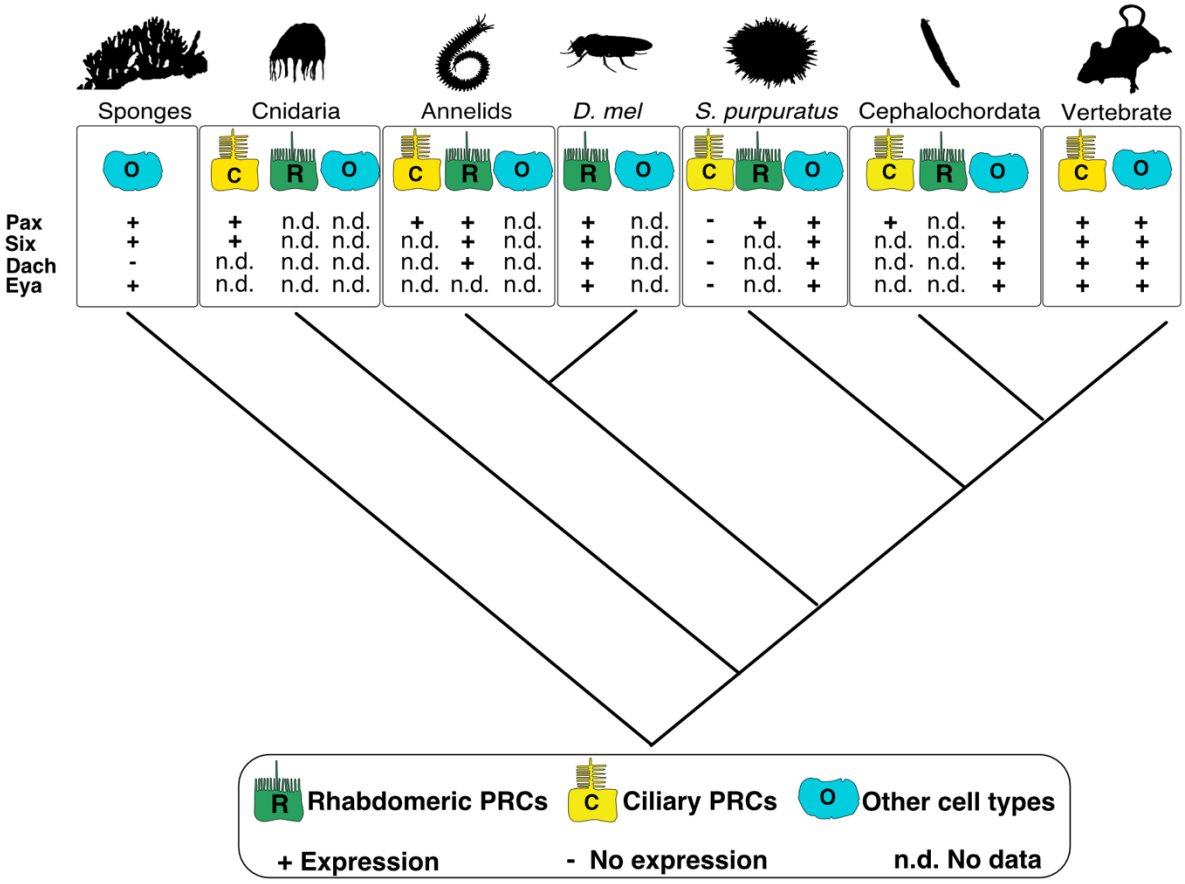


Figure S2.7: Phylogenetic distribution of PRCs and other cell types expressing RDN factors. Phylogenetic tree describing the expression of the retinal determination genes in ciliary PRCs, rhabdomeric PRCs, and other cell types. PRC distribution has been defined according to (1). Shown are summarized data from sponges (21, 22), cnidarians (45, 46), anellids (2, 45, 47-49), D. melanogaster (50, 51), S. purpuratus this work, B. floridae (52, 53), M. musculus and other vertebrates (50, 54-56).

Table S1: Primers used in this work.

Gene name	whl	Forward	Reverse
Rx	WHL22.523971.0	AAGAGCAACGGTGGAATAAAAAC	GCTGATTATACGTTCAGGCAAGA
Six3	WHL22.121654.0	CTCATAGACACACCCCAGCA	AGGATGGTGGGATCTTTCTTC
Awh	WHL22.9369.1	CATAACCATCCCATCAATAAATCC	TAATACGACTCACTATAGGGAGATGCACAGCACTCTATTTTCAATC
Ia1	WHL22.769122.0	ACCCTACAAGTGCAACTGAAACA	ATGGGCAAGTTGTGCAGTAATAA
SoxC	WHL22.622787.0	GTCACAAATCGAGAGGAGACG	TAATACGACTCACTATAGGGAGATCTGAGTCTATGAGTTCGCTTACC
FoxG	WHL22.389872.0	GCGCTTTACTCGTCTTATTCTACC	GTCCTTAGTTGAAATGGGAAACC
Hbn	WHL22.523959.1	TCATTACTTCGTTGGAGTTACCC	CATGAAAACGTCTGGATACTGG
Otx	WHL22.532435.1	AACAGCAGCAACAGCAACAG	AGAGCTGCGTTCAAGGTCAT
SoxB2	WHL22.104525.0	ATCAGAGACTTTCCCCATCATC	TAATACGACTCACTATAGGGAGAGTGTTGCACAGTCCTTGTTGAC
Tbx2/3	WHL22.457020.0	TCACAAAAGAGGAACAGAAATGG	TAATACGACTCACTATAGGGAGAGGGATGGGTGTCTAAATAACTCG
Zic	WHL22.331651.0	CAATCGCGTTTCAGTTGACTAC	ACGTACCATTCACTCAAGTTCGT
Nkx2.1	WHL22.739581.0	AAGCAGCAGAAGTACCTGTCG	TAATACGACTCACTATAGGGAGAATGACTATTGTGTGGTGCAAGC
Six1/2	WHL22.121485.0	GCAATAACTTCTCACCGCATAAC	TAATACGACTCACTATAGGGAGAGTTCATGTTTTTCTTTCCGACTG
Pax6	WHL22.585512.0	CGCAATCAGAGAAGACAGCA	TTAGCCAGCAAGAAGGGAAA
Eya	WHL22.168736	GTATTGGAAGAGGGCGTCAA	TAATACGACTCACTATAGGGAGAATGACTTGTTACCCGCCAG
Dac	WHL22.169355	GATGCGAACCTGTTCTACG	TAATACGACTCACTATAGGGAGACAATTCAAAAGCTTGTGGCA
opsin3.2	WHL22.338995	CGGTAACATCACCGTCCTTT	TAATACGACTCACTATAGGGAGACGGAATTTGGAGCTTGATGT
opsin2	WHL22.272775	CGTTAATGTCCCATGCTGTG	TAATACGACTCACTATAGGGAGACTTTGGGCAAGACAGCAGAT

Chapter 3

EXPRESSION OF DISCRETE COMBINATORIAL TRANS-REGULATORY STATES UNDERLIES DEVELOPMENTAL ORGANIZATION OF THE SEA URCHIN LARVA

Jonathan E. Valencia and Isabelle S. Peter

Division of Biology and Biological Engineering, MC156-29, California Institute of Technology, Pasadena, CA 91125

3.1 Abstract

Development is an inherently dynamic process during which embryonic space becomes defined continuously and in a progressive manner. Gene regulatory networks specify new spatial domains of cells by controlling the expression of regulatory genes throughout development. Here we assessed the global activity of gene regulatory networks in terms of expression of regulatory states, the combination of transcription factors expressed together in given spatial domains at given times. We performed a genome-wide survey and embryo-wide annotation of regulatory gene expression by whole mount in situ hybridization at five consecutive developmental time-points during sea urchin embryogenesis in order to determine the developmental trajectory of regulatory states. We report the identification of 74 discrete regulatory state domains which overlap with the morphological structures of the larva and show that their progenitor domains foreshadow the ensuing larval morphology. We show that these regulatory states are composed of distinct sets of regulatory genes wherein particular combinatorics define their spatial and temporal organization. We found that similarities among regulatory states reflect a common developmental function but not necessarily a common developmental history. These results suggest that larval morphology is tightly associated with regulatory state expression and that therefore the function of regulatory states is to determine the identity, fate, and function of cells in respect to the specific morphological structure it is expressed in.

3.2 Introduction

In animal development, the generation of increasingly complex organization of cell fates in space and time is driven by developmental gene regulatory networks (GRN) (Peter and Davidson, 2011a, 2015). Developmental GRNs driving this complexity represent a system composed of regulatory interactions between regulatory genes, encoding sequence-specific, DNA-binding transcription factor proteins. GRNs perform developmentally specific functions that control the progressive specification of cell fates. Several recent studies have experimentally characterized GRNs underlying early animal development and organogenesis. In the purple sea urchin *Strongylocentrotus purpuratus*, several distinct GRNs operate at the vegetal plate prior to the onset of gastrulation, distinguishing anterior endoderm from mesoderm, oral mesoderm from aboral mesoderm, and anterior endoderm from posterior endoderm (Davidson, 2002; Oliveri et al., 2008; Materna et al., 2013; Peter and Davidson, 2011b). In chordates, early embryonic GRNs have been analyzed for the specification of multiple developmental fates in *Ciona* (Imai et al., 2009; Satou and Imai, 2015), mesendoderm specification in *Xenopus* (Loose and Patient, 2004; Koide et al., 2005; Charney et al., 2017), and dorsal-ventral axis formation in zebrafish (Chan et al., 2009). Additionally, multiple GRNs have been elucidated in various vertebrate and mammalian organs: neural crest (Sauka-Spengler and Bronner-Fraser, 2008; Simoes-Costa and Bronner, 2015), neural tube (Nishi et al., 2009; Balaskas et al., 2012), lens (Cvekl and Zhang, 2017), retina (Wang et al., 2014), and heart (Cripps and Olson, 2002). While GRNs are typically composed of similar recurrent circuit structures, these circuits are operated by cell-fate specific sets of transcription factors (Davidson, 2010; Davidson and Levine, 2008; Peter and

Davidson, 2009, 2015). The combination of transcription factors that operate together in a GRN and that are co-expressed in a particular cell fate as a result of gene regulatory interactions define the cell fate-specific regulatory state.

The qualitative nature of regulatory states represents an underlying feature in their function as units in the control of gene expression (Peter, 2017). Cis-regulatory modules (CRMs) controlling developmentally regulated genes require the combinatorial function of multiple, distinct transcription factors (Xu et al., 2014; Yuh et al., 1998). Thus specific combinations of transcription factors are required to occupy a CRM to control the expression of downstream genes. An important consequence of this combinatorial gene regulation is that transcription factors can be pleiotropic and expressed in multiple developmental contexts of a developing embryo, where they regulate the expression of context-specific target genes. For example, expression of Tbx5 in the developing vertebrate heart, lung and forelimbs as well as Pax6 in the developing vertebrate eye, brain, neural tube, and pancreas, demonstrate that transcription factor function is also determined by other members of the context-specific regulatory state (Blake and Ziman, 2014; Papaioannou, 2014). Thus, to qualitatively characterize the combinatorial properties of regulatory states that define cell fates during development will prove useful in understanding how TFs perform in context-specific situations as well as shedding light on the underlying GRN.

Understanding the essential features of different cell fates and their change over time requires a system level approach. In order to characterize different cell types, many studies now employ single cell transcriptomics, wherein these data contain the entire breadth of expressed genes, not only the combination of expressed regulatory genes, but also including genes that are involved in cellular processes such as cell cycle regulation, cell metabolism, or structural cell organization. One caveat of single cell approaches, however, is that information on the spatial relationship among sequenced cells is lost during the procedure of isolating single cells, and the fate of a cell, its path along a differentiation trajectory and perhaps its origin, have to be reconstructed based on gene expression data. Several studies have solved this dilemma by combining massive single cell RNA-seq with reference gene expression patterns based on in situ hybridization data to computationally map sequenced cells back to their embryonic origin (Karaiskos et al., 2017; Satija et al., 2015). Due to the scale of acquired expression data, these studies are usually restricted to one or few developmental timepoints, limiting the knowledge of how the expression of regulatory genes changes over developmental time. An additional issue is that particularly regulatory genes that tend to be expressed at relatively low levels are not always reliably detectable at the single cell level. Other approaches involving the use of gene expression analysis by in-situ hybridization have thus been used to annotate regulatory gene expression patterns over several time-points in developmental time in order to create a comprehensive atlas of developmental gene and/or TF expression patterns (Bell et al., 2004; Visel, 2004; Pollet et al., 2005; Tomancak et al., 2007; Tassy et al., 2010; Diez-Roux et al., 2011; Spencer et al., 2011; Hammonds et al., 2013; Hu et al., 2017). While these data sets successfully detect spatial gene expression patterns even for low abundance regulatory genes at multiple

developmental time points, cell-fate specific gene expression is a consequence not just of individual transcription factors but of the combination of transcription factors, the regulatory state, expressed in a given cell fate. Thus, the global identification and characterization of regulatory states over developmental time in any model organism has so far not been determined.

Here we deployed a developmentally comprehensive approach to studying the combinatorics of regulatory gene expression during development of the sea urchin larva. We used whole mount in situ hybridization (WMISH) to detect the expression of 260 regulatory genes at five developmental time points and used these data to determine expressed regulatory states. We found that sea urchin larvae consist of at least 74 discrete domains expressing distinct regulatory states that are often associated with unique morphological structures. These regulatory domains are established during development such that the developmental expression of distinct regulatory states foreshadows the appearance of morphological structures. A vast majority of analyzed regulatory genes is expressed in specific patterns that include multiple regulatory domains. Furthermore, a comparison of cell fate specific regulatory states shows relatively little variation in the number of transcription factors expressed and in the relative contribution of different transcription factor families. Nevertheless, the similarity of regulatory states as determined by hierarchical clustering reflects the morphological structure such that functionally and structurally related cell fates tend to express similar regulatory states. Thus while the complexity of regulatory states is

comparable among cell fates throughout the early development of this organism, spatial regulatory information is provided by the cell fate specific combination of transcription factors.

3.3 Results

Spatial expression of regulatory genes defines regulatory state domains

The sea urchin genome contains approximately 400 regulatory genes encoding known transcription factors with DNA-binding domains (Sea Urchin Genome Sequencing Consortium et al., 2006). Notably, this set largely excludes zinc finger factors of unknown function, although some of these may execute transcriptional regulatory functions as well. A recent analysis of the sea urchin developmental transcriptome was used to generate 368 models of developmental regulatory genes (Tu et al., 2012, 2014). In order to determine the set of developmentally expressed regulatory genes, we selected genes expressed between 24 and 72 hours with at least 300 transcripts per embryo (Howard-Ashby et al., 2006) for at least 16 consecutive hours. We identified 267 regulatory genes that met these criteria. In addition, 23 regulatory genes expressed at less than 300 transcripts per embryo had previously reported developmental expression and were included in this analysis. In total, close to 360 RNA probes were designed for 290 regulatory genes (Table S3.1). The spatial expression of these genes was determined by WMISH at five developmental time points with 12 hour intervals, from just before onset of gastrulation (24h) to pluteus larva (72h; Figure 3.1). Of the 290 regulatory genes analyzed, we experienced technical difficulties in probe design or WMISH

for 30 genes, but spatial expression results were obtained for 260 genes. For each regulatory gene for which spatial expression was detected, microscopy images of stained embryos were obtained at each developmental timepoint from multiple views, angles, and focal depths to capture staining, accumulating a set of approximately 300,000 images (Figure 3.1).

Identification of larval domains expressing distinct combinatorial regulatory states

Although gene expression domains have been analyzed in great detail for the pregastrular (before 30h) sea urchin embryo, the spatial organization of cell fates after the onset of gastrulation has been characterized morphologically (Burke, 1978, 1980; Burke and Chia, 1980; Burke and Alvarez, 1988; Cameron and Davidson, 1991; Smith et al., 2008) and by lineage tracing (Cameron et al., 1987, 1991, 1993), but not molecularly. Thus before annotating the spatial expression of regulatory genes, it was necessary to generate a map of gene expression domains for developmental stages after 30h. We assumed that a maximum number of domains would be present at the last timepoint included in this study, and we thus first focused on 72h larvae. To simplify the analysis of spatial domains, we separately identified gene expression patterns in the 72h larva in each of the following morphologically distinct territories: apical plate ectoderm (APE), ciliated band ectoderm (CBE), oral ectoderm (OE), aboral ectoderm (ABO), skeletal mesoderm (SKM), non-skeletal mesoderm (MES), and endoderm (ENDO). Within each territory, we identified domains of cells that express a common regulatory state. Thus we searched for gene expression patterns that distinguish specific group of cells, either by expression or absence of expression of individual regulatory genes, or by overlapping expression of multiple regulatory genes. The principle

approach is shown in Figure 3.2 for regulatory state domains in the midgut. In this example, as few as eight expressed regulatory genes can be seen to divide the larval stomach into multiple regulatory state domains. Thus in a territory, each identified domain is delineated by a unique combination of multiple expressed transcription factors. We applied this approach to all territories in the larva in order to determine the expressed regulatory states. Importantly, throughout our analysis we abided by a conservative process in the identification of domains wherein we outlined domains that were easily discernible through manual comparison since our data consists entirely of single gene WMISH images. We only defined domains that were clearly distinct based on our data set of 260 regulatory genes, and cannot exclude the possibility that these domains consist of further sub-domains or single cells expressing unique regulatory states.

In total, we identified 74 unique regulatory domains from all seven territories in the 72h larva (Figure 3.3). Because of the high number of identified domains, we introduced a simple nomenclature to ID a domain for reference (Table S3.2). We named domains by an abbreviation of the associated territory plus a digit corresponding to the numerical count, e.g., APE1 domain corresponds to domain 1 of the apical plate ectoderm. Importantly, although perhaps not surprisingly, this analysis demonstrates that every part of the larval body plan showing morphologically distinct structure is also molecularly distinct. Thus the boundaries of morphological structures coincide with the molecular boundaries between different regulatory domains. However, most morphological structures are subdivided into several regulatory state domains, potentially indicating different cell fates (Figure 3.2, 3.3) (see below).

In analyzing the regulatory domains in each territory, we determined that the endoderm, which includes the larval tri-partite gut, the two muscular sphincters, and anus as containing the highest number of domains at 22 (Figures 3C and A.7). Interestingly, while most regulatory domains are organized along the anterior-posterior axis of the gut, we also identified several regulatory gene expression patterns that polarize the gut into an oral and an aboral portion, thus doubling the number of domains (Figures 3.2, 3.3C, S3.1 and A.7). Sea urchin mesoderm comprises two territories: 1) skeletogenic mesoderm composed of mesenchymal cells that give rise to the larval skeleton and 2) non-skeletogenic mesoderm, which has been shown by lineage labeling to contribute to pigment cells, blastocoelar cells, the coelomic pouches, and the circumesophageal muscles of the larva (Cameron and Davidson, 1991; Cameron et al., 1991). Within the skeletogenic mesoderm we identified 9 regulatory domains, each of which are composed of only a few to several cells, which can be distinguished based on their particular location and position along the various skeletal rods/spicules (Figures 3.3A and A.5). In the non-skeletogenic mesoderm, totaling 12 identified domains, we determined several domains within the coelomic pouches (CP), structures that arise from cells of the tip of the archenteron, where the left CP will give rise to the adult rudiment (Figures 3.3B and A.6). Additionally, we identified several blastocoelar domains which we distinguished by their unique locations within the larval blastocoel (Figures 3.3B and A.6). Within the ectoderm, ciliated band ectoderm, composed of a narrow strip of 3-5 columnar ciliated cells, separates the squamous epithelial cells of the oral and aboral ectoderm (Cameron et al., 1993). We identified 9 regulatory domains in the ciliated band, the most anterior of which overlap with the apical neurogenic domain (Figures 3.3D,

ciliated band ectoderm and A.2). The apical plate ectoderm, a neurogenic region housing the larval brain, i.e., the apical organ with clusters of neurons and axons (Burke et al., 2014), we identified 11 domains organized into concentric rings and grid-like regions along the oral-aboral and medial-distal axes (Figures 3.3D, apical plate ectoderm and A.1). The oral ectoderm comprises the mouth and its surrounding regions with 14 regulatory domains (Figures 3.3D, oral ectoderm and A.3). Surprisingly, we discovered a domain of 2-3 cells, also expressing *opsin3.2*, corresponding to light sensing photoreceptors at the junction between apical plate and oral ectoderm (Valencia et al., *unpublished*). All ectodermal territories except aboral ectoderm also include several single cell expression domains, possibly neuronal cell types (Figure 3.3D, all ectoderm). The aboral ectoderm, however, is rather a simple epithelium and includes only 5 distinct regulatory domains (Figures 3.3D, aboral ectoderm and A.4).

Developmental specification of regulatory state domains

As embryos develop, an increasing number of gene expression domains become specified. To identify the developmental origins of the regulatory state domains present in the larva, we therefore analyzed the expression domains present at earlier developmental stages. Rather than searching for domains de novo, we used regulatory gene expression patterns found in the larva to determine when and where molecular boundaries are formed in earlier developmental stages. Thus we used regulatory genes expressed in exclusive domains as markers to identify common progenitor domains at earlier stages. At 60h, we found the same number and location of larval domains present in 72h embryo, indicating that the larval domains are already established. At stages during gastrulation, we found that fewer domains

are defined in the 48h late gastrula ($n_{domains}$= 60) and the 36h mid gastrula ($n_{domains}$ = 40). In the pregastrular 24h embryo, we identified 26 regulatory state domains revealing that during the events of gastrulation, at least 52 new domains emerge and are specified prior to the formation of the post-gastrula pluteus larva. Using both the spatial location and the combination of expressed regulatory genes to identify developmentally related regulatory state domains, we determined the developmental ancestry according to which the progenitor domains at 24h give rise to all cell fate domains specified in the 72h lava (Figures 3.4 and A.1-A.7).

To determine if new regulatory domains formed synchronously across the embryo and continuously in time, we analyzed the timing when new domains formed in each part of the embryo. For each territory, we determined the sequence and developmental timing of domain formation from the early 24h embryo through gastrulation up to the 72h larva (Figure 3.4). We found that most domains within ectodermal territories are predominately established by 24h and undergo very little increase in spatial complexity during gastrulation. By 24h, apical ectoderm and ciliated band have each established 7 and 4 domains, respectively, but by 72h, each has specified a total of 11 and 5 domains, respectively, indicating that these ectodermal territories go through only few additional cell fate decisions during gastrulation (Figures 3.4 and 3.6). However, the situation is very different in endoderm and mesoderm. These territories significantly increase in spatial complexity during gastrulation. At 24h, endoderm and mesoderm each consist of two domains, but by 36h, each territory increases at least twofold in the total number of specified domains (n_{meso} = 4, n_{endo} = 10) and similarly again by 48h (n_{meso} = 11, n_{endo} = 16), suggesting that the

majority of specified domains within endoderm and mesoderm results from several cell fate decisions during gastrulation (Figures 3.4). Taken together, these data reveal that spatial complexity increases at particular times during development, for ectoderm, particularly between 12-24h (before the onset of gastrulation), suggesting that specification of ectoderm largely occurs before gastrulation, whereas for endoderm and mesoderm, a major increase is seen between 36-48h towards the end of gastrulation demonstrating that prior to gastrulation, the progenitors of mesoderm and endoderm which lie at the vegetal plate remain multi-potent.

Strikingly, these processes also revealed that cell fates are distinctly specified hours before morphological structures become apparent (Figure 3.3). For example, the archenteron of the mid gastrula (36hpf) shows no signs of constrictions along the anterior-posterior axis where the future sphincters develop nor are the future compartments of the gut recognizable at the morphological level (Figure 3.1). Yet, distinct regulatory state domains exist in the future locations of the foregut, cardiac sphincter, midgut, hindgut and anus, albeit, without any hint of structural change in archenteron morphology (Figures S3.1 and A.7), consistent with the idea that specification of domains precedes the establishment of morphological form.

Expression of unique regulatory states associated with morphological structure

To fully characterize the composition of identified regulatory state domains throughout time and space in the developing embryo, we manually annotated all developmental expression patterns of regulatory genes using a controlled vocabulary of 74

terms corresponding to identified domains (throughout developmental time). To annotate gene expression patterns, we developed a semi Boolean data matrix to assign an expression value for every regulatory gene in each domain and at every time-point examined. We annotated gene expression using values ranging from 0 to 3 to indicate following observed staining: 0, no expression; 1, strong expression; 2, weak expression or not consistently observed in all embryos; and 3, partial expression, meaning that expression was observed in a subset of cells within a domain. The distinction between "strong" and "weak" was made based on the following assumption. "Strong" expression was used to indicate expression that most likely resulted in functional levels of transcription factor expression, i.e., 10 copies of transcripts or more. On the other hand, "weak" expression was used to indicate cases where neither absence of expression nor strong expression was reproducibly observed, and where few transcripts are expressed per cell which may or may not produce functional levels of transcription factor. We used quantitative developmental transcriptome data from whole embryos together with approximate number of cells showing expression to estimate expression levels per cell (Bolouri and Davidson, 2003). Though regulatory genes annotated as weakly expressed might only produce very low levels of transcription factors, we cannot entirely exclude the possibility that these transcription factors might nevertheless contribute to transcriptional activity in these cells. The result of these annotations of regulatory gene expression in all regulatory domains in the sea urchin embryo at five developmental stages is shown in Figure 3.5 in the form of a Boolean expression matrix. To facilitate the comparison of regulatory gene expression across domains and territories, we clustered regulatory genes by similarity in expression patterns, while domains are shown as grouped by territory.

The gene expression matrix revealed a number of unexpected results. First, out of the 260 analyzed regulatory genes, only 2 show strong ubiquitous expression. In addition, about 15 regulatory genes showed no clear pattern, but are expressed at levels between 300-3000 transcripts/embryo. Since the sea urchin larva consists of approximately 1800 cells at 72h, this level of expression results in <5 transcripts/cell on average, and we interpreted this expression pattern as ubiquitous weak expression. However, all other regulatory genes are expressed in specific spatial patterns. Importantly, as revealed by the expression matrix, these regulatory genes contribute to the expression of spatially and temporally unique combinatorial regulatory states. Thus we found no two regulatory states to be equal, even those expressed in a given domain at two subsequent developmental time points. The number of regulatory genes expressed in each regulatory state is relatively small compared to the large number of genes not expressed. A comparison of regulatory states between domains of a given territory showed that there are territory-specific combinations of expressed regulatory genes. These combinations are consistently expressed throughout all regulatory states at all time-points of a territory and can be visualized in the matrix by a contiguous vertical block of expression throughout a territory, revealing that each territory identifies with a unique core group of regulatory genes that as a combination are expressed in all regulatory states throughout development.

Comparison of combinatorial regulatory states expressed in 72h larva

These results indicate that morphological structures and cell fates express unique regulatory states, and we next focused on identifying the similarities and differences between

spatial regulatory states expressed in the 72h larva. A relatively simple assumption would be that regulatory states are distinguished by transcription factors that are exclusively expressed in a given domain. However, only 15 regulatory genes are expressed exclusively in one regulatory domain at 72h (Figure 3.6A). The majority of regulatory genes are expressed in multiple domains. There are a total of 74 regulatory domains at 72h, and thus there are fewer exclusively expressed regulatory genes than there are domains, indicating that domain specification is rarely due to the novel expression of a single regulatory gene but mostly due to the expression of unique combinations of regulatory genes. In order to assess whether gene expression occurs in broad contiguous domains or in multiple independent areas of the embryo, we evaluated the number of territories each regulatory gene is expressed in. At 72h, about one third of transcription factors are expressed specifically in one of the seven territories, whereas the majority is expressed in two or more territories (Figure 3.6B), confirming that a majority of transcription factors contributes to multiple developmental processes.

We next asked whether the number of expressed regulatory genes differs between regulatory domains. On average, regulatory states express 27 regulatory genes, roughly one-tenth of all developmentally expressed regulatory genes in our dataset (Figure S3.2A). However, the complexity of regulatory states in terms of number of expressed transcription factors varies substantially between different domains at 72h (Figure 3.6C). For example, regulatory states of the apical ectoderm vary from 11 (APE 9, aboral distal region of the apical organ) to 54 (APE 4, central medial region of the apical plate) expressed regulatory genes. Additionally, regulatory states found within the mesodermal coelomic pouches are

considerably larger ($n_{average}$= 40) than those in migratory pigment (n =15) and multipolar blastocoelar cells ($n_{average}$ = 20). Before the onset of gastrulation, regulatory states are roughly similar in size, averaging 24 expressed regulatory genes (Figure S3.2 A,B, and C). However, after the onset of gastrulation, many regulatory states in APE, CBE, MES, and ENDO dramatically increase in size whereas most SKEL and ABO regulatory states undergo a constant reduction (Figures 3.6C and S3.2C). Taken together, these data suggest that regulatory state size varies considerably depending on cell fates and/or developmental process.

To gain additional insight into the differences and similarities between larval regulatory states in their respective territories, we compared the regulatory states expressed in each territory at 72h by hierarchical clustering. We found that functionally related regulatory states tend to be clustered together. For instance, in the endoderm, the eight regulatory states expressed in the midgut form a cluster that is separate from the clusters of four regulatory states of the foregut, and those of the hindgut (Figure S3.3). This comparison shows that the similarity of regulatory states in each cluster relies on a specific set of exclusively expressed regulatory genes, indicating that domains located in the same organs or morphological structures express similar regulatory states as well as a few exclusively expressed regulatory genes. Further, in territories with few domains like aboral ectoderm (n = 5) which lack clear morphological structures (and whose cells differentiate into squamous epithelium), domains form one cluster with highly similar regulatory states that are distinguished by as few as three expressed regulatory genes (compare ABO2 to ABO5, Figure S3.6), indicating that for a territory such as the squamous epithelium of the aboral

ectoderm, minor differences in regulatory states result in the specification of distinct regulatory state domains.

To investigate whether individual DNA binding domain (DBD) families are associated predominantly with particular developmental processes, we analyzed the fraction of transcription factor families composing each regulatory state at 72h. Overall, we found little variation in the relative contribution of different transcription factor families among regulatory states, although several states (e.g., APE8B and SKM7) exhibit uncommon ratios (Figure 3.6D). However, when we compared regulatory state domains within territories, we noticed that certain territories shared unique features among all or most of their expressed regulatory states. For example, all domains within aboral ectoderm lack expression of Basic Zipper, Ets, and Forkhead families, whereas all domains in oral ectoderm and ciliated band ectoderm are devoid of nuclear hormone receptor expression. In skeletal mesoderm, most domains lack expression of Sox genes, however, all domains do show greater contributions of BHLH members and relatively little contributions from homeodomain members when compared to regulatory state domains in other territories, particularly those in aboral ectoderm where the contribution of homeodomain families is expanded. Though minor differences in the relative contribution of transcription factor families do exist among larval regulatory states, to a large extent there are similar proportions of transcription factor families expressed in larval regulatory states, suggesting that a similar concoction of transcription factor families contribute to the specification of larval cell fates.

Temporal changes in regulatory state expression

Regulatory states change both in space and in time during development. To address the temporal change of regulatory states throughout developmental time, we assessed the total number of regulatory genes expressed in each territory between 24h to 72h. We found that the number of expressed regulatory genes continually increases in many territories over the course of development. Notably, during gastrulation, APE, CBE, OE, MES, and ENDO undergo an increase in expressed regulatory genes until 60h (Figure 3.6E). By 60h, when specification of larval domains is completed, the number of expressed regulatory genes plateau in these territories. Of these territories, mesoderm expresses the highest number of regulatory genes, with approximately 120 regulatory genes required for mesoderm formation. A similar extent of regulatory information is required for development of the gut, which involves approximately 100 regulatory genes. However, the specification of skeletogenic mesoderm or the aboral ectoderm involves a relatively low number of transcription factors. The number of regulatory genes expressed in each territory, and the rate of change, follows a trend that is also observed in the number and change of domains in these territories (Figure 3.6F). A Pearson correlation test confirmed that the number of expressed regulatory genes is positively correlated to the number of specified domains in all territories except in aboral ectoderm, which showed a strong negative correlation (Figure S3.4). Oddly, the number of expressed regulatory genes decreases consistently throughout development in aboral ectoderm but steadily gains new regulatory state domains in time (Figures 3.5A and 3.5B). These data suggest that the specification of new domains in mesodermal and endodermal territories occurs primarily during gastrulation, and that this increase in spatial requires an increase in expressed regulatory genes.

Similarity in regulatory states versus developmental history

Since regulatory states expressed within a morphological structure tend to express core sets of transcription factors, we wondered whether this similarity between functionally related regulatory states would reflect the morphological organization of the sea urchin larva. Furthermore, all 72h regulatory states derive from a much smaller number of progenitor domains at 24h, and we asked whether the combination of co-expressed transcription factors would represent not only spatial organization but also developmental history. To identify the similarity between larval regulatory states, we performed a clustering analysis on all 72h regulatory states (Figure 3.7). This analysis revealed that regulatory states expressed in a given territory mostly cluster together, exhibiting fewer differences among each other than to regulatory states expressed in other territories. Thus, regulatory states of the endoderm, ciliated band, skeletogenic mesoderm, aboral ectoderm, and non-ciliated band apical ectoderm all form territory specific clusters, consistent with the observation that territory-specific regulatory state domains express similar sets of regulatory genes. However, regulatory states of mesoderm and oral ectoderm are distributed into several non-adjoining organ-associated clusters i.e., oral ectoderm forms three separate clusters which include regulatory states expressed in the oral band, the mouth region, and the post oral ectoderm. Further, these clusters are widely distributed amongst all larval regulatory states exhibiting more similarities in regulatory gene expression to regulatory states from other territories (CBE, ABO, MES, APE, and SKM) than to each other. Like OE, mesoderm forms two distinct non-adjoining clusters, where regulatory states associated with the coelomic pouches including the inner circumesophageal muscle regulatory state together form a single cluster and a second is formed from regulatory states associated with migratory blastocoelar cells

and pigment cells, as well as the hydropore canal. These results revealed that although regulatory states expressed in a given morphological structure or organ tend to express a set similar regulatory genes and form a common cluster, this cluster does not necessarily include all regulatory states of the respective territory.

To test whether similarity between regulatory states reflects their developmental history, we assessed whether regulatory states associated with developmentally related domains tend to cluster together. We first focused on the similarity between 72h regulatory states descending from the same 24h progenitor state. For example, Figure 3.7 shows that all endodermal regulatory states cluster together and form a group of exclusively endodermal regulatory states. Within the endoderm, the regulatory states expressed in the midgut are different from regulatory states expressed in the fore- and hindgut, even though fore- and midgut domains are both derived from anterior endoderm and are developmentally more related to each other than to the hindgut. Similarly, the skeletogenic mesoderm regulatory states that are expressed in cells all deriving from a common skeletogenic precursor domain form a territory-specific cluster. On the other hand, non-skeletogenic mesoderm regulatory states that derive from a common endomesodermal progenitor domain are more similar to ectodermal than endodermal regulatory states. Within the mesoderm, coelomic pouch and circum-esophageal regulatory states all originate from the 24h oral mesoderm progenitor domain (Figure 3.4), and by 72h form an exclusive cluster that is separate from all other regulatory states of the larva (Figure 3.7). Interestingly, the coelomic pouch structures give rise to the rudiment forming the juvenile sea urchin, which is a structure developing independently from the rest of the larva, and the unique regulatory states expressed in this

structure reflect this difference in developmental function. Furthermore, we also found clusters of regulatory states formed from descendants of separate 24h progenitor domains. For example, 72h aboral ectoderm derives from two progenitor domains, the animal and vegetal ectoderm that are specified distinctly very early in development, yet regulatory states deriving from both (e.g., ABO2 from animal and ABO5 from vegetal, exhibit less differences in regulatory state expression than their sister states (Figure 3.7). Wildly, APE5B, a regulatory state specified late in apical plate development and expressed in serotonergic neurons (Figure 3.4), forms a separate cluster that is related to a larger inclusive cluster including regulatory states from mesoderm, skeletogenic mesoderm, oral and aboral ectoderm, and non-sister regulatory states from apical plate (Figures 3.4 and 3.7). Thus, taken together, clusters of regulatory states generally do not represent developmental origin, but in fact reflect the functional and morphological organization of the larva.

3.4 Figures

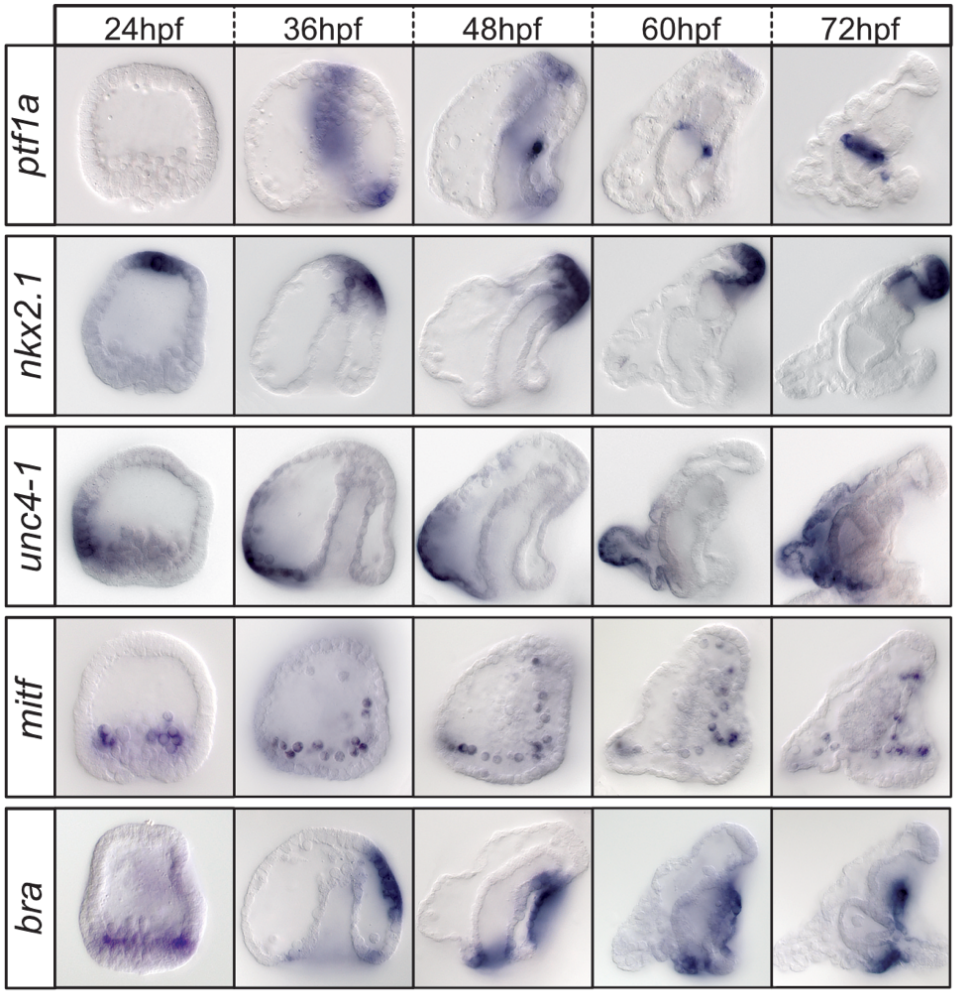


Figure 3.1: Spatial expression profiles for selected regulatory genes. Images of embryos stained by whole-mount in-situ hybridization (WMISH) at five developmental stages showing specific expression in the seven major embryonic territories. ptf1a is expressed in the ciliated band and apical plate at 36h and 48h and also in a ring of cells in the anterior midgut of the endoderm; nkx2.1 is expressed in the apical plate ectoderm throughout the time course and is also expressed in oral ectoderm starting at 36h; unc4-1 is expressed in the posterior aboral ectoderm; mitf is expressed in all skeletogenic cells throughout the time-course; bra is expressed in posterior endoderm and oral ectoderm throughout the time-course.

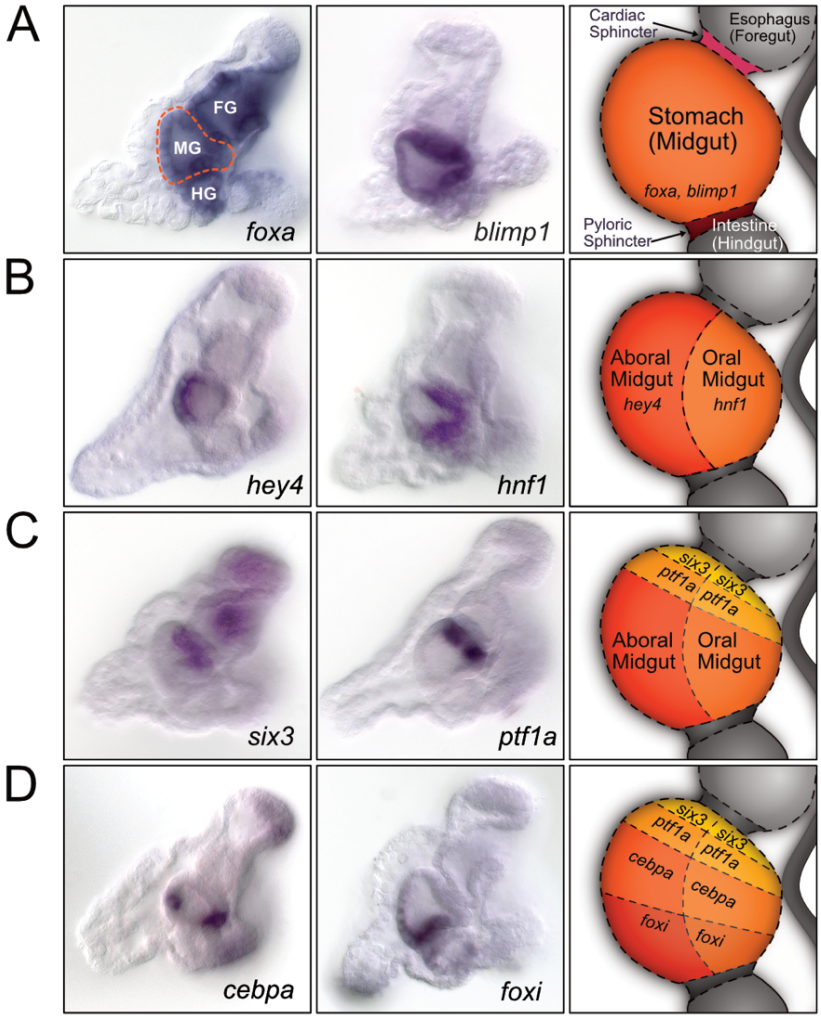


Figure 3.2: Identifying Regulatory State Domains by Differentially Expressed Regulatory Genes. Regulatory state domains were identified molecularly by manual comparisons of the spatial expression of regulatory genes at 72h. As an example, the larval stomach consists of eight regulatory domains that are defined in the following way: (A) foxa is expressed throughout the larval endoderm whereas blimp expression is restricted to the stomach, forming an initial molecular domain associated the larval stomach. (B) Expression of hey4 in the cells of the aboral midgut and hnf1 in the cells of the oral midgut polarize the larval stomach into two halves. (C) Expression of six3 and ptf1a are seen in two separate though adjacent concentric rings, where six3 expressing cells abut the posterior border of the cardiac sphincter and border ptf1a expressing cells from the anterior. (D) cebpa shows expression in center most regions of the midgut whereas foxi exhibits expression in the posterior regions of the midgut. Together, these regulatory genes partition the larval stomach into four divisions along the anterior-posterior axis and one division on the oral-aboral axis dividing the stomach into a total of eight discrete molecular domains exhibiting the differential gene expression of distinct combinations of regulatory genes (MG1: six3, hnf1, blimp1; MG2: six3, hey4, blimp1; MG3: ptf1a, hnf1, blimp1; MG4: ptf1a, hey4, blimp1; MG5: cebpa, hnf1, blimp1; MG6: cebpa, hey4, blimp1; MG7: foxi, hnf1, blimp1; MG8: foxi, hey4, blimp1).

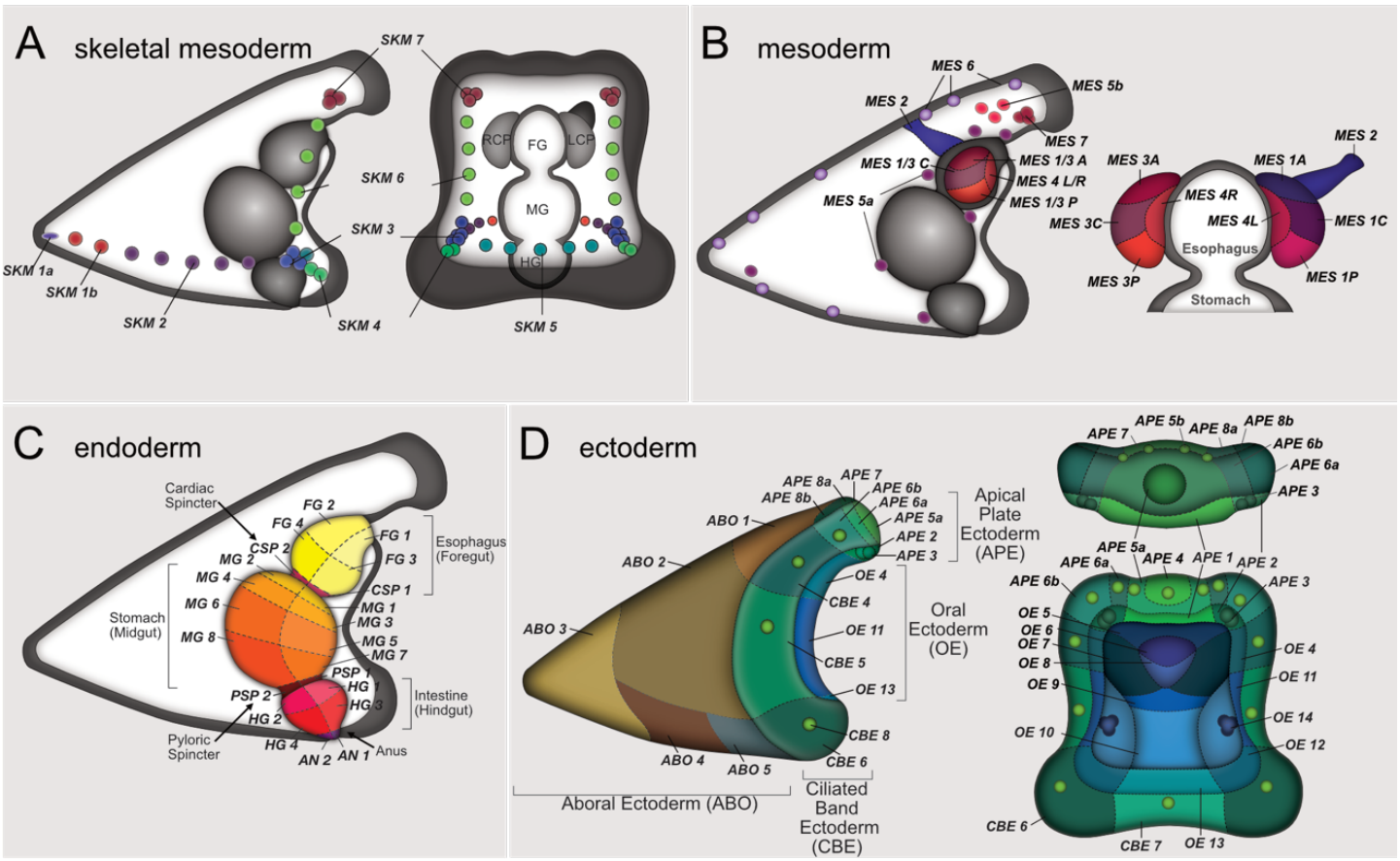


Figure 3.3: Regulatory State Domains in the 72h Larva. Identified regulatory state domains in the pluteus larva associate with its morphological structures. (A-D) Schematic representations of the larva are organized by regulatory state domains determined from combinatorial analysis of regulatory gene expression patterns. (A) Lateral and oral schematic views of skeletogenic mesoderm. Seven regulatory state domains within the SKM are organized in single cells that contribute to the various rods of the larval skeleton. (B) Lateral and zoomed oral schematic view of mesoderm. Several regulatory state domains within the mesoderm represent migratory pigment cells (MES6/13), blastocoelar cells (MES5A/7, MES5B/8, MES7/12), regions of the coelomic pouches (MES1/3A, MES1/3C, MES1/3P, and MES4) and the hydropore canal (MES2). (C) Lateral schematic view of larval endoderm. A total of 22 regulatory states domains are organized in several geometric configurations within the larval gut. The organs of the tri-partite gut are organized by regulatory states domains as quadrants defining four regions (i.e., ant-oral, ant-aboral, post-oral, post-aboral) each in the esophagus (FG1-4) and intestine (HG1-4), as four concentric regions bisected coronally to form eight regions in the stomach (MG1-8), or as two regions, either oral or aboral in the sphincters (CSP1-2, PSP1-2) and anus (AN1-2). (D) Lateral, apical, and oral schematic views of aboral, ciliated band, oral, and apical plate ectoderm. The aboral ectoderm is arranged by five regulatory state domains in an anterior to posterior fashion (ABO1-5). Nine regulatory states domains organize the apical plate as single cells (APE3, APE5B), concentric circles (APE4-5A) and in a grid (APE1-2, APE6A-B, APE7-8B). The ciliated band which separates oral from aboral ectoderm is organized by several regulatory state domains in the apical portions of the band (CBE1-3) which overlaps with apical plate domains

(APE4-6B), vertical regions (CBE4-5), posterior regions (CBE6-7), and single cells (CBE8). Like the ciliated band, regulatory state domains organize the perimeter of oral ectoderm into a band of several regions (OE1-4, OE11-13), in which the apical regions overlap with the apical plate (APE1-3). Within the band, many regulatory state domains are associated in and around the mouth (OE5-9), a single posterior to the mouth (OE10) and in single cells post orally (OE14). Regulatory state domains and their larval morphological associations are listed in TableS2. APE, apical plate ectoderm; CBE, ciliated band ectoderm; OE, oral ectoderm; ABO, aboral ectoderm; SKM, skeletal mesoderm; MESO, mesoderm; RCP, right coelomic pouch; LCP, left coelomic pouch; ENDO, endoderm; FG, foregut; CSP, cardiac sphincter; MG, midgut; PSP, pyloric sphincter; HG, hindgut; AN, anus.

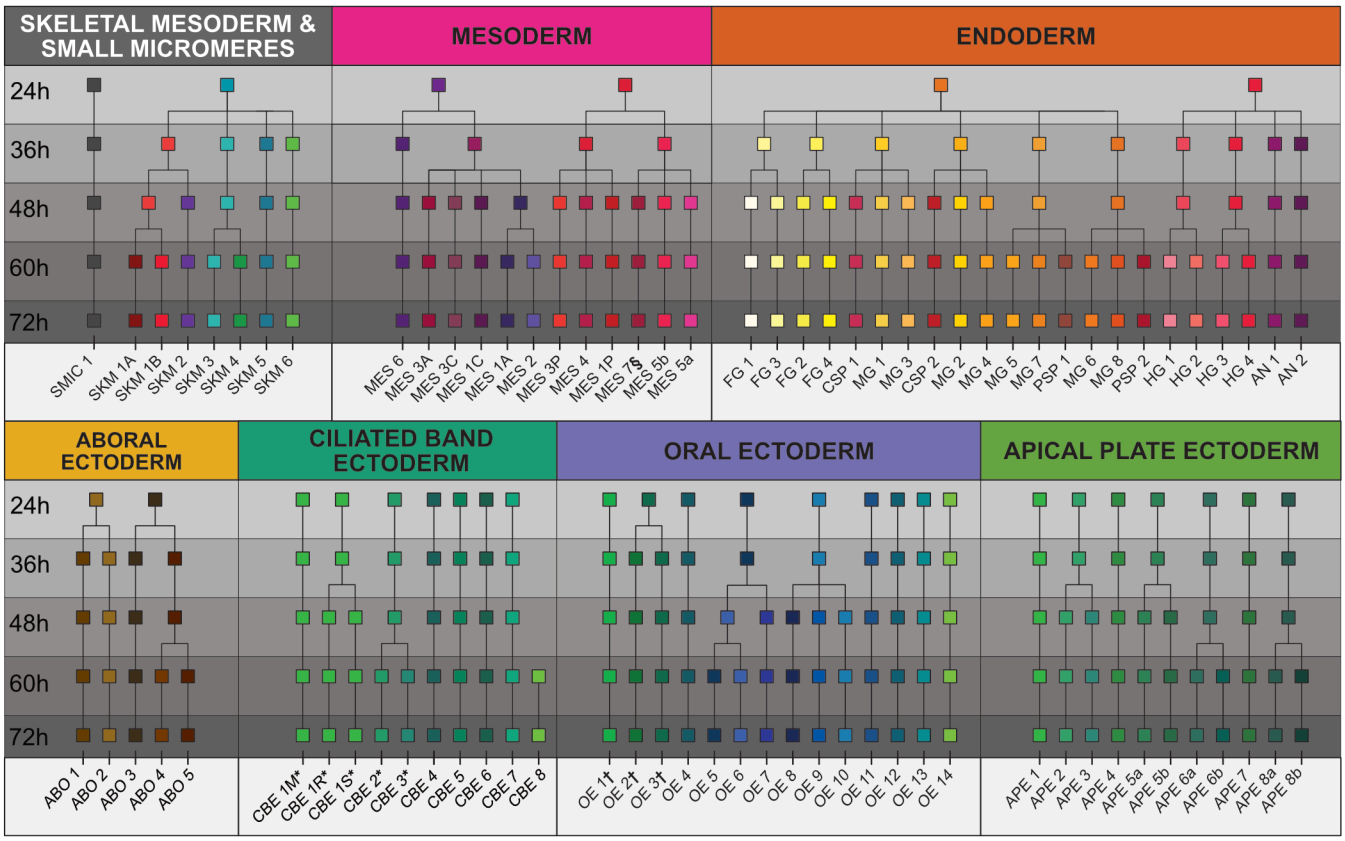


Figure 3.4: Developmental Ancestry of Regulatory State Domains. Diagram showing the developmental ancestry of regulatory state domains and the timing of cell fate decisions determining spatial organization. Developmental progenitor domains of the larval domains were identified based on spatial expression of regulatory genes defining the 72h regulatory states (see text). Shaded horizontal bars stacked along vertical axis represent developmental time-points. Regulatory state domains are represented as colored boxes within their embryonic territory (color code as in Fig. 3). Vertical lines indicate temporal linkages between regulatory state domains, while horizontal lines mark bifurcation events of progenitor domain, whereby a parent domain undergoes spatial subdivision into two or more sister subdomains at the subsequent time-point. Asterisk marks the ciliated band domains (CBE1M, CBE1R, CBE1S, CBE2, CBE3) that are synonymous with the apical plate domains APE4, APE5A, APE5B, APE6A, and APE6B respectively. Cross denotes oral ectoderm domains (OE1, OE2, and OE3) that are synonymous with the apical plate domains APE1, APE2, and APE3 respectively. Section symbol indicates the mesodermally-derived domain, MES7/13, is synonymous to SKM7. APE, apical plate ectoderm; CBE, ciliated band ectoderm; OE, oral ectoderm; ABO, aboral ectoderm; SKM, skeletal mesoderm; MES, mesoderm; FG, foregut; CSP, cardiac sphincter; MG, midgut; PSP, pyloric sphincter; HG, hindgut; AN, anus.

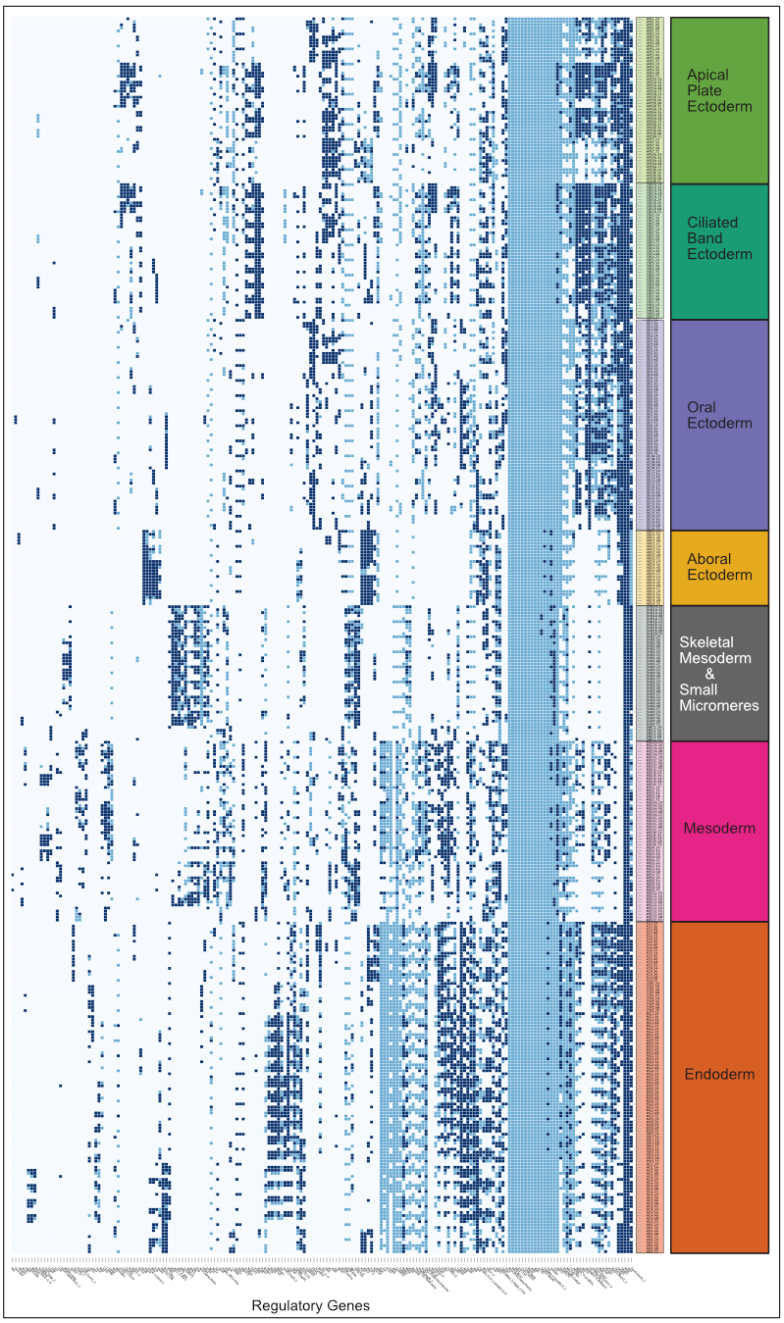


Figure 3.5: Clustered Boolean Matrix showing Combinatorial Regulatory States. Spatial expression of regulatory genes was annotated using semi-Boolean terms in all regulatory domains and at all developmental time points. Regulatory states are shown in rows for each domain at 24, 36, 48, 60, and 72h, and domains are represented within their territory. The expression of regulatory genes is shown in columns, and regulatory genes are clustered by similarity of expression pattern. Colored boxes represent strong expression (dark blue), weak expression (light blue), and no expression (white) of regulatory genes. Regulatory genes exhibiting ubiquitous expression are seen as vertical bars of light blue (weak expression) or dark blue (strong expression).

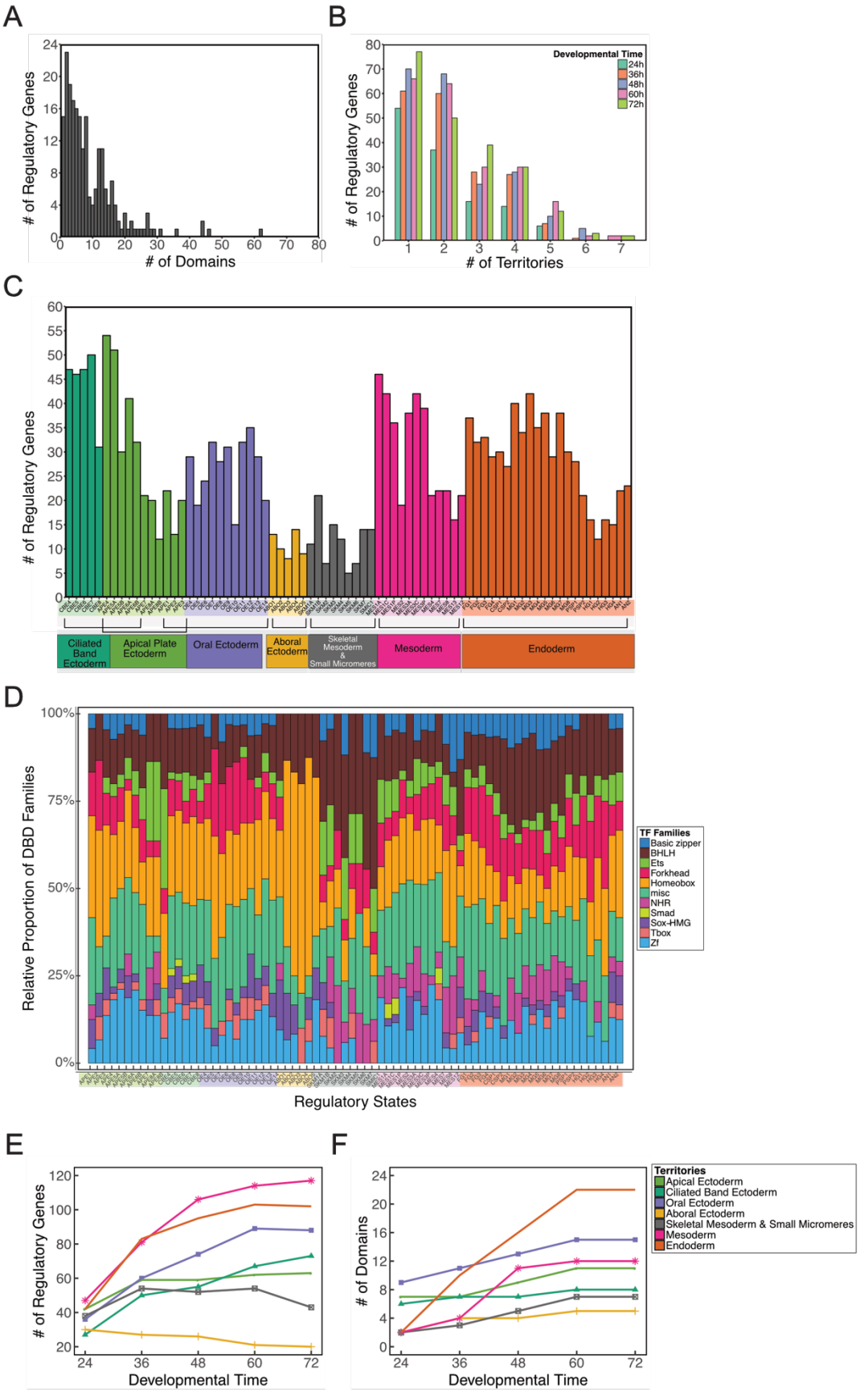

A
# of Regulatory Genes
# of Domains
B
# of Regulatory Genes
# of Territories
Developmental Time
24h
36h
48h
60h
72h
C
# of Regulatory Genes
Ciliated Band Ectoderm
Apical Plate Ectoderm
Oral Ectoderm
Aboral Ectoderm
Skeletal Mesoderm & Small Micromeres
Mesoderm
Endoderm
D
Relative Proportion of DBD Families
Regulatory States
TF Families
Basic zipper
BHLH
Ets
Forkhead
Homeobox
misc
NHR
Smad
Sox-HMG
Tbox
Zf
E
# of Regulatory Genes
Developmental Time
F
# of Domains
Developmental Time
Territories
Apical Ectoderm
Ciliated Band Ectoderm
Oral Ectoderm
Aboral Ectoderm
Skeletal Mesoderm & Small Micromeres
Mesoderm
Endoderm

Figure 3.6: Comparison of combinatorial regulatory states expressed in larva. (A, B) Spatial distribution of expressed regulatory genes in larval domains and territories. (C) Size of larval 72h regulatory states grouped by territory. (D) Relative proportion of specifically expressed DNA-binding domain family members in larval regulatory states. BHLH, basic helix loop helix; NHR, nuclear hormone receptor; Zf, zinc finger; misc, miscellaneous includes single to several members of various TF families which include: AP2, bright, COE, DM, E2F, GCM, Grouch/TLE, HMG-box, ipt, IRF, LAG1, Lim, LZTF1, Mads-box, Myb, NFI, Nrf1, pcg, polycomb, PWWP, RHD, RRM, Runx, SAM, ski/sno, SRF, STAT, Tea, TrxG, TSC-22, tulp, and yeats. (E) Size of sea urchin territories measured by the number of expressed regulatory genes as a function of time. (F) Size of sea urchin territories measured by the number of regulatory states gained over developmental time.

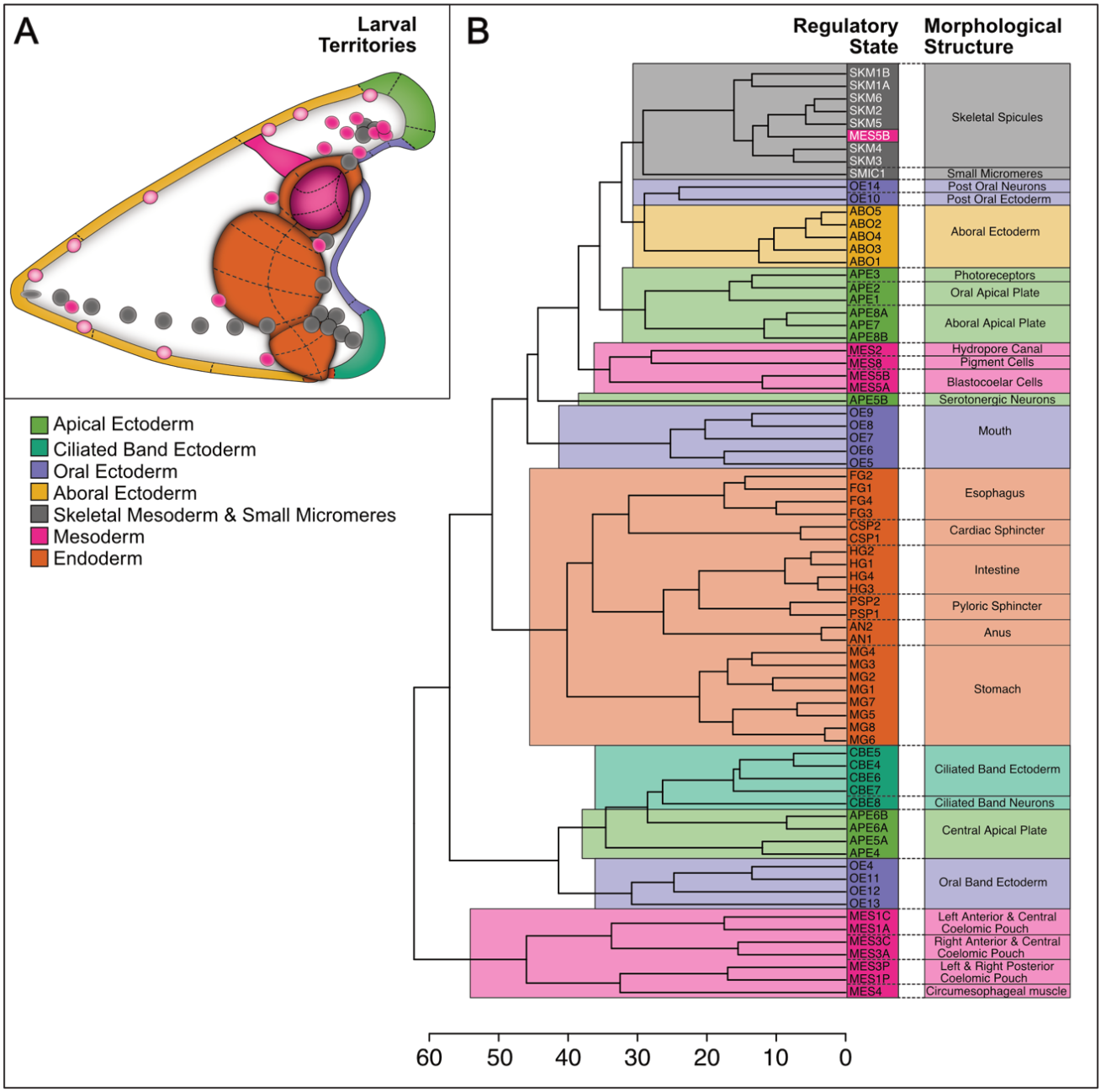


Figure 3.7: Hierarchical Clustering of Regulatory States Reveals functional similarity. (A) Larval schematic illustrating the location of larval territories. Color scheme denotes territory and is the same as in B. (B) Dendrogram shows hierarchical clustering of regulatory states expressed at 72h. Clustering reveals high similarities of regulatory state expression among domains within same territory (color code as in Fig. 3). For example, all regulatory states expressed in the gut form a cluster that is further subdivided into three subgroups containing foregut (FG), midgut (MG), and hindgut (HG) regulatory states. Regulatory states expressed in the coelomic pouches, where the rudiment of the juvenile sea urchin develops, are distinct from all regulatory states expressed in the larval structures of the sea urchin.

3.5 Discussion

This work represents a system-wide analysis of regulatory state expression during sea urchin embryogenesis. Although maps of regulatory gene expression have been generated in other developmental systems (Tomancak et al., 2007; Tassy et al., 2010; Diez-Roux et al., 2011), our focus on combinatorial regulatory states in embryonic space over time to describe the global process of developmental specification represents a novel undertaking. At each developmental stage, unique regulatory states determine all developmental functions and are considered the active states of the GRN (Peter and Davidson, 2015; Peter, 2017).

In our identification of combinatorial regulatory states and successive annotations of regulatory gene expression, expression of regulatory genes in discrete domains of cells was annotated using Boolean values. Earlier studies using Boolean logic to capture regulatory interactions have successfully recapitulated aspects of the developmental process thereby denoting its value in investigating and solving GRNs (Peter et al., 2012). By employing semi-Boolean logic, we treated manual annotations with expression values of either strong, weak, or non-expressed. Though these annotations are mostly based on non-quantitative WMISH expression data, they have been corroborated with quantitative whole embryo transcriptome data. However, one question that remains is whether annotated genes with weak values actually have any functional value in a regulatory state, a prospect that is debatable yet testable. Further, by using comparative single chromogenic in-situs, our

analysis is limited in its depth of resolving finer domains, that is, those that may exist as single cell domains in a larger field or those that lie at a particular border.

The identification and molecular characterization of 74 regulatory state domains shows that sea urchin larval morphology and function is associated with expression of distinct regulatory states. Thus functionally and morphologically distinct structures are associated with distinct molecular signatures. We found that regulatory states associated with the larval endoderm, mesoderm, or ectoderm, either overlie with the structures and organs of these larval territories or assemble in multiple geometric configurations within them creating finer suborgan/substructure molecular domains. Strikingly, each of these domains expresses a changing combination of regulatory genes over the course of development, such that no two regulatory states expressed in different domains or at different times are equivalent. Furthermore, we found that distinct regulatory state domains during early developmental stages foreshadow the ensuing larval morphology several hours before the formation of distinct anatomical structures. This systems-level view of pattern formation processes reveals the formation of discrete domains during development of the sea urchin larva. With the exception of a few isolated gene expression patterns, relatively little was known so far about the cell fates established during gastrulation and formation of the larva. Thus, in any particular area, space is defined not by one but by many regulatory genes which are associated in determining morphology and specification of cell fates. Importantly, these associations validate the approach of using regulatory states to define embryonic space. This

study identified many novel spatial domains in the embryo. Because we define domains exhibiting differential gene expression, such patterns may not have been identified previously through the annotations of single regulatory gene expression patterns.

Our results suggest that regulatory states predominantly consist of combinations of transcription factors that are each expressed in a spatially restricted manner. Our analysis of regulatory states revealed very few genes with ubiquitous expression among developmentally expressed regulatory genes whereas most exhibit specific spatial information. Though the size of regulatory states is different in all domains, they average to 27 regulatory genes in larval states. In principle, the differences between regulatory states expressed in different spatial domains of the sea urchin larva could be reflected by differences in the composition of transcription factor families represented in each regulatory state. However, we found that to a large extent, TF families are equally represented in larval regulatory states, with few exceptions. Results of TF family comparisons across regulatory states revealed similar ratios of family members in each state. Where differences were observed, they usually did not just affect individual regulatory states, but usually presented features shared among all domains of a territory. For example, all five regulatory states expressed in the aboral ectoderm (ABO) lack expression of Basic Zipper, Ets and Forkhead family factors, while most regulatory states in the oral ectoderm and ciliary band ectoderm lack expression of nuclear hormone receptors. Thus to a large extent, regulatory states

operate using a similar recipe in the fraction of regulatory genes of specific families for determining developmentally-specific and cell type-specific functions.

The identification of distinct sets of regulatory genes within regulatory states highlights the importance of combinatorial expression of multiple regulatory genes. The observation that regulatory states expressed in a given morphological structure such as for example the stomach tend to share common sets of transcription factors implies an element of modularity in regulatory states. The current view of modules in GRNs is that individual subcircuits execute the various biological functions necessary for correct developmental operation. We offer an additional view of modularity in regulatory states, where modules are defined by the distinct sets of TF within regulatory states that identify its region at multiple levels in embryonic space. For example, in any one regulatory state there exist a module to define its place in a territory, another in a morphological structure, and one in its specific regulatory state domain. Whether or not these two views of modularity are congruent remains to be tested.

Conclusion

In summary, our work reveals that the sea urchin larva is composed of 74 discrete, morphology-associated regulatory state domains that are the result of cell fate specification processes throughout sea urchin embryogenesis. Based on the analysis of all

developmentally expressed regulatory genes encoding known transcription factors, the change of regulatory states during developmental specification has been revealed. This work presents a substantial advancement from classic pattern formation studies and provides a molecular definition of embryonic and larval space at the level of combinatorial expression of regulatory genes that determine morphology and specification of cell fates. The results enable a view on the overall operation of developmental GRNs over developmental time laid out in embryonic space, providing the basal framework for relating the regulatory genome to developmental function.

3.6 References

Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K.M., Briscoe, J., and Ribes, V. (2012). Gene Regulatory Logic for Reading the Sonic Hedgehog Signaling Gradient in the Vertebrate Neural Tube. Cell *148*, 273–284.

Bell, G.W., Yatskievych, T.A., and Antin, P.B. (2004). GEISHA, a whole-mount in situ hybridization gene expression screen in chicken embryos. Dev. Dyn. *229*, 677–687.

Blake, J.A., and Ziman, M.R. (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. Development *141*, 737–751.

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

Burke, R. (1978). The structure of the nervous system of the pluteus larva of Strongylocentrotus purpuratus. Cell Tissue Res. *191*.

Burke, R.D. (1980). Morphogenesis of the digestive tract of the pluteus larva of *Strongylocentrotus purpuratus:* shaping and bending. Int. J. Invertebr. Reprod. *2*, 13–21.

Burke, R., and Alvarez, C. (1988). Development of the esophageal muscles in embryos of the sea urchin Strongylocentrotus purpuratus. Cell Tissue Res. *252*.

Burke, R.D., and Chia, F.S. (1980). Morphogenesis of the digestive tract of the pluteus larva of *Strongylocentrotus purpuratus* : sphincter formation. Int. J. Invertebr. Reprod. *2*, 1–12.

Cameron, R.A., and Davidson, E.H. (1991). Cell type specification during sea urchin development. Trends Genet. *7*, 212–218.

Cameron, R.A., Hough-Evans, B.R., Britten, R.J., and Davidson, E.H. (1987). Lineage and fate of each blastomere of the eight-cell sea urchin embryo. Genes Dev. *1*, 75–85.

Cameron, R.A., Fraser, S.E., Britten, R.J., and Davidson, E.H. (1991). Macromere cell fates during sea urchin development. Development *113*, 1085–1091.

Cameron, R.A., Britten, R.J., and Davidson, E.H. (1993). The Embryonic Ciliated Band of the Sea Urchin, Strongylocentrotus purpuratus Derives from Both Oral and Aboral Ectoderm. Dev. Biol. *160*, 369–376.

Chan, T.-M., Longabaugh, W., Bolouri, H., Chen, H.-L., Tseng, W.-F., Chao, C.-H., Jang, T.-H., Lin, Y.-I., Hung, S.-C., Wang, H.-D., et al. (2009). Developmental gene regulatory networks in the zebrafish embryo. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1789*, 279–298.

Charney, R.M., Paraiso, K.D., Blitz, I.L., and Cho, K.W.Y. (2017). A gene regulatory program controlling early Xenopus mesendoderm formation: Network conservation and motifs. Semin. Cell Dev. Biol. *66*, 12–24.

Cripps, R.M., and Olson, E.N. (2002). Control of Cardiac Development by an Evolutionarily Conserved Transcriptional Network. Dev. Biol. *246*, 14–28.

Cvekl, A., and Zhang, X. (2017). Signaling and Gene Regulatory Networks in Mammalian Lens Development. Trends Genet. *33*, 677–702.

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

Davidson, E.H., and Levine, M.S. (2008). Properties of developmental gene regulatory networks. Proc. Natl. Acad. Sci. *105*, 20063–20066.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A High-Resolution Anatomical Atlas of the Transcriptome in the Mouse Embryo. PLoS Biol. *9*, e1000582.

Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., and Celniker, S.E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. Genome Biol. *14*, R140.

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

Hu, Y., Comjean, A., Perrimon, N., and Mohr, S.E. (2017). The Drosophila Gene Expression Tool (DGET) for expression analyses. BMC Bioinformatics *18*.

Imai, K.S., Stolfi, A., Levine, M., and Satou, Y. (2009). Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. Development *136*, 285–293.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science *358*, 194–199.

Koide, T., Hayata, T., and Cho, K.W. (2005). Xenopus as a model system to study transcriptional regulatory networks. Proc. Natl. Acad. Sci. U. S. A. *102*, 4943–4948.

Loose, M., and Patient, R. (2004). A genetic regulatory network for Xenopus mesendoderm formation. Dev. Biol. *271*, 467–478.

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

Nishi, Y., Ji, H., Wong, W.H., McMahon, A.P., and Vokes, S.A. (2009). Modeling the spatio-temporal network that drives patterning in the vertebrate central nervous system. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1789*, 299–305.

Papaioannou, V.E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. Development *141*, 3819–3833.

Peter, I.S. (2017). Regulatory states in the developmental control of gene expression. Brief. Funct. Genomics.

Peter, I.S., and Davidson, E.H. (2009). Modularity and design principles in the sea urchin embryo gene regulatory network. FEBS Lett. *583*, 3948–3958.

Peter, I.S., and Davidson, E.H. (2011a). Evolution of Gene Regulatory Networks Controlling Body Plan Development. Cell *144*, 970–985.

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

Peter, I.S., and Davidson, E.H. (2015). Genomic control process: development and evolution.

Peter, I.S., Faure, E., and Davidson, E.H. (2012). Predictive computation of genomic logic processing functions in embryonic development. Proc. Natl. Acad. Sci. *109*, 16434–16442.

Pollet, N., Muncke, N., Verbeek, B., Li, Y., Fenger, U., Delius, H., and Niehrs, C. (2005). An atlas of differential gene expression during early Xenopus embryogenesis. Mech. Dev. *122*, 365–439.

Ransick, A. (2004). Detection of mRNA by in situ hybridization and RT-PCR. Methods Cell Biol. *74*, 601–620.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Satou, Y., and Imai, K.S. (2015). Gene regulatory systems that control gene expression in the *Ciona* embryo. Proc. Jpn. Acad. Ser. B *91*, 33–51.

Sauka-Spengler, T., and Bronner-Fraser, M. (2008). A gene regulatory network orchestrates neural crest formation. Nat. Rev. Mol. Cell Biol. *9*, 557–568.

Sea Urchin Genome Sequencing Consortium, Sodergren, E., Weinstock, G.M., Davidson, E.H., Cameron, R.A., Gibbs, R.A., Angerer, R.C., Angerer, L.M., Arnone, M.I., Burgess,

D.R., et al. (2006). The Genome of the Sea Urchin Strongylocentrotus purpuratus. Science *314*, 941–952.

Simoes-Costa, M., and Bronner, M.E. (2015). Establishing neural crest identity: a gene regulatory recipe. Development *142*, 242–257.

Smith, M.M., Cruz Smith, L., Cameron, R.A., and Urry, L.A. (2008). The larval stages of the sea urchin,Strongylocentrotus purpuratus. J. Morphol. *269*, 713–733.

Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J., et al. (2011). A spatial and temporal map of C. elegans gene expression. Genome Res. *21*, 325–341.

Tassy, O., Dauga, D., Daian, F., Sobral, D., Robin, F., Khoueiry, P., Salgado, D., Fox, V., Caillol, D., Schiappa, R., et al. (2010). The ANISEED database: Digital representation, formalization, and elucidation of a chordate developmental program. Genome Res. *20*, 1459–1468.

Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *8*, R145.

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

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

Visel, A. (2004). GenePaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res. *32*, 552D–556.

Wang, S., Sengel, C., Emerson, M.M., and Cepko, C.L. (2014). A Gene Regulatory Network Controls the Binary Fate Decision of Rod and Bipolar Cells in the Vertebrate Retina. Dev. Cell *30*, 513–527.

Xu, Z., Chen, H., Ling, J., Yu, D., Struffi, P., and Small, S. (2014). Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. Genes Dev. *28*, 608–621.

Yuh, C.-H., Bolouri, H., and Davidson, E.H. (1998). Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science *279*, 1896–1902.

3.7 Material and Methods

Gene amplification and probe synthesis.

The primer sets used for gene amplification are listed in TableS1. Gene models generated from sea urchin transcriptome analysis (Tu et al., 2014) were used as a reference for primer design using T7 tailed primers or cloning. cDNA prepared from various developmental stages was used as template for PCR. For cloning, PCR products were purified and ligated into pGEM-TEZ or pCRII plasmid vectors. Cloned genes were PCR-amplified using the primer flanking the insert region, and PCR products were used to synthesize RNA probes for WMISH.

Whole-Mount *In-situ* Hybridization

The protocols for animal culture, collection, fixation, and whole-mount *in-situ* hybridization (WMISH) to detect spatial gene expression has been described previously (Ransick, 2004). Slight modifications to this protocol were adapted to allow for a high through-put production of spatial expression by *in-situ* hybridization for the entire developmental range of sea urchin embryos. Briefly, sea urchin embryos from a large culture were fixed for 48-72 hours in 4% paraformaldehyde solution at the time-points of 24, 36, 48, 60 and 72hpf. For storage, embryos were washed in 1M mops solution and put through an ethanol gradient and stored at -20C in 70% ethanol. Fixed and rehydrated embryos were incubated in hybridization buffer [50% (vol/vol) formamide, 5× SSC, 1× Denhardt's, 1 mg/mL yeast tRNA, 50 ng/mL heparin, and 0.1% tween-20] with a concentration from 1 to 2 ng/μL digoxygenin RNA probe(s) at 60 °C for 18 h. Two Post hybridization washes were performed with hybridization buffer without RNA probe, 2× SSCT (2× SSC, 0.1% tween-20), 0.2× SSCT, and 0.1× SSCT, each 20 min at 60 °C. Subsequently, 5 washes were performed with a buffer of 0.1% Tween 20, 10% MOPS (1M), 10% NaCl (5M) and 80% DEPC water. Antibody incubations were performed at room temperature with 1:2,000 diluted anti-DIG Fab (Roche). The embryos were extensively washed before staining reaction, including six times with 9 MABT buffer (0.1 M maleic acid, 0.15 M NaCl, and 0.1% tween-20), twice with AP buffer [100 mM Tris·Cl (pH 9.5), 100 mM NaCl, 50 mM MgCl2, and 1 mM levamisole]. 5-Bromo- 4-chloro-3-indolyl-phosphate (BCIP) and nitro blue tetrazolium were used for staining.

Image Capture

For each regulatory gene in the dataset, we imaged at least 3 embryos per developmental time-point (24h, 36h, 48h, 60h, 72h) and captured a standard set of images from a variety of focal depths taken at a lateral view, with additional images taken from oral, aboral, apical and anal views depending on the complexity of expression patterns. Expression patterns were validated by comparison to RNA-seq and any available primary literature. *In-situ* hybridization experiments were repeated in cases of confounding results or conflicting data. Images were captured using a Zeiss AxioSkop microscope equipped with a Zeiss AxioCam camera. Images shown in Figure 1 were downloaded from our database with minor image processing adjustments to correct for variations in color balance and orientation.

Databases

Sea urchin expression database (http://mandolin.caltech.edu/ExpressionData/index.php) and raw image archive (http://mandolin.caltech.edu/JonathanImages/index.php) were created for use as an image repository and as an analytical tool to aid in identifying regulatory state domains and annotating regulatory gene expression patterns.

Pairwise Comparisons

To calculate pairwise comparisons between two variables, we used Pearson's correlation method using the function cor() in the "stats" R package to generate a matrix of correlation coefficients (r). Visual correlation matrices were created using "corrplot" package in R.

Hierarchical Clustering

For all distance comparisons between regulatory states, we performed hierarchical clustering using the base functions of dist() and hclust() in the standard R library ("stats" package). A distance matrix was computed from expression values of Regulatory genes in regulatory states using the absolute distance (method = "manhattan") between any two regulatory states. We used the simple agglomerative method (method = "average") of unweighted pair group method with arithmetic mean (UPGMA) with each distance matrix to generate a hierarchical

cluster. To visualize clustering, dendrograms and heatmaps were created using the "dendextend" and the "d3heatmap" packages in R.

3.8 Supplementary Information

3.8.1. Supplementary Figures

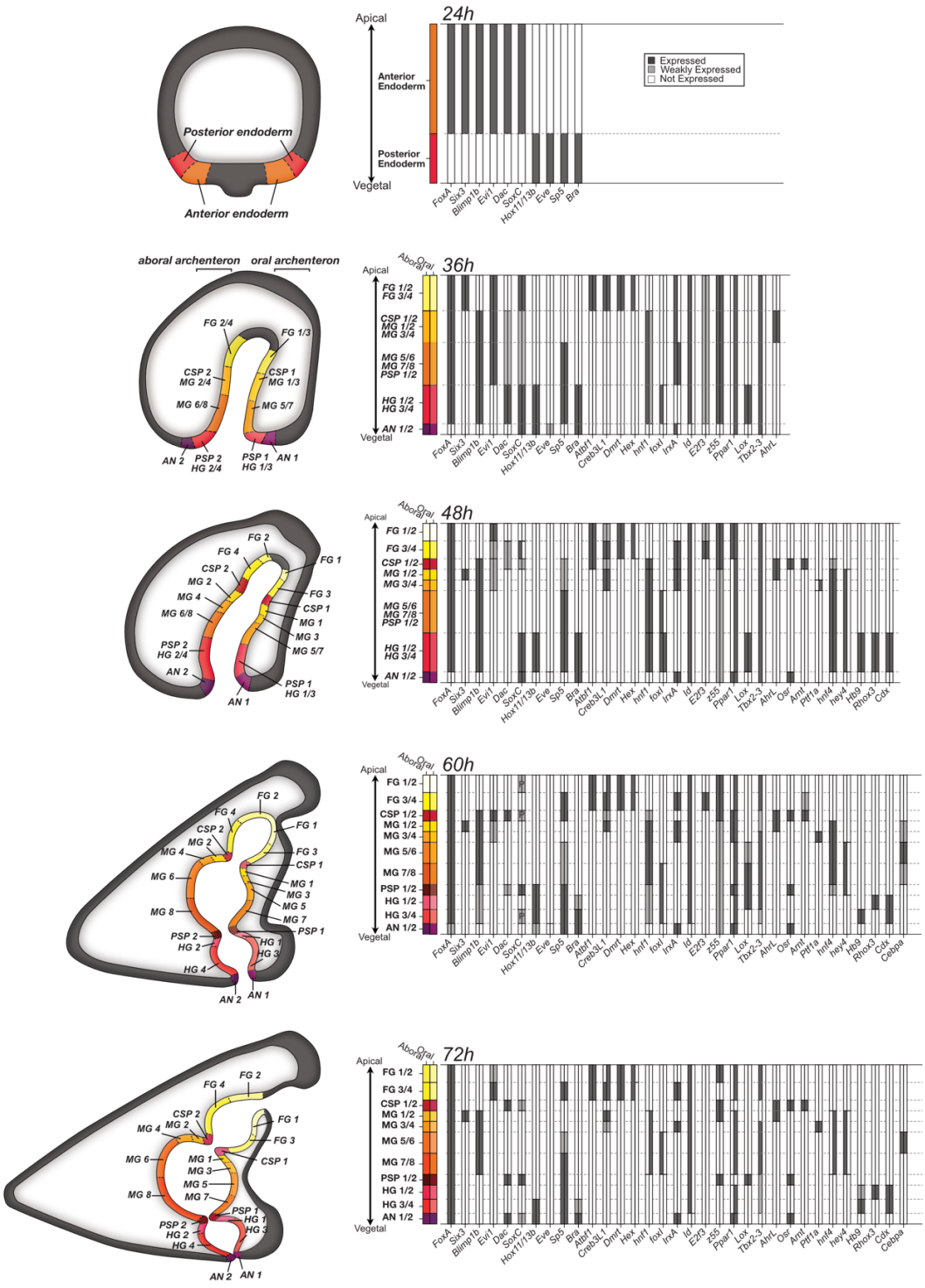


Figure S3.1: Endoderm Specification. Embryonic staging schematics on the left represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Schematics are drawn in a lateral view and are oriented with the oral ectoderm facing right and the vegetal pole lying down. Identified regulatory state domains within the endoderm are highlighted in color. Regulatory state domains were identified and mapped using WMISH (Figure 2). For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. On the right of each schematic is a graphic representation of an expression table of the regulatory genes expressed within the endodermal regulatory states of the associated developmental time-point. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the endoderm and represent the composition of regulatory genes responsible for their specification. For the developmental stages ranging between 36-72hpf, the endoderm is divided into oral and aboral sub-territories along the archenteron/gut. This subdivision is represented in the table as two adjacent columns for every gene, oral on the right and aboral on the left. Those regulatory states can be read in rows accordingly. As a whole, these schematics demonstrate the specification of domains within the endoderm along developmental time. The emergence of a novel domains (left schematic) can be attributed to the establishment of a new regulatory state (right expression table), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

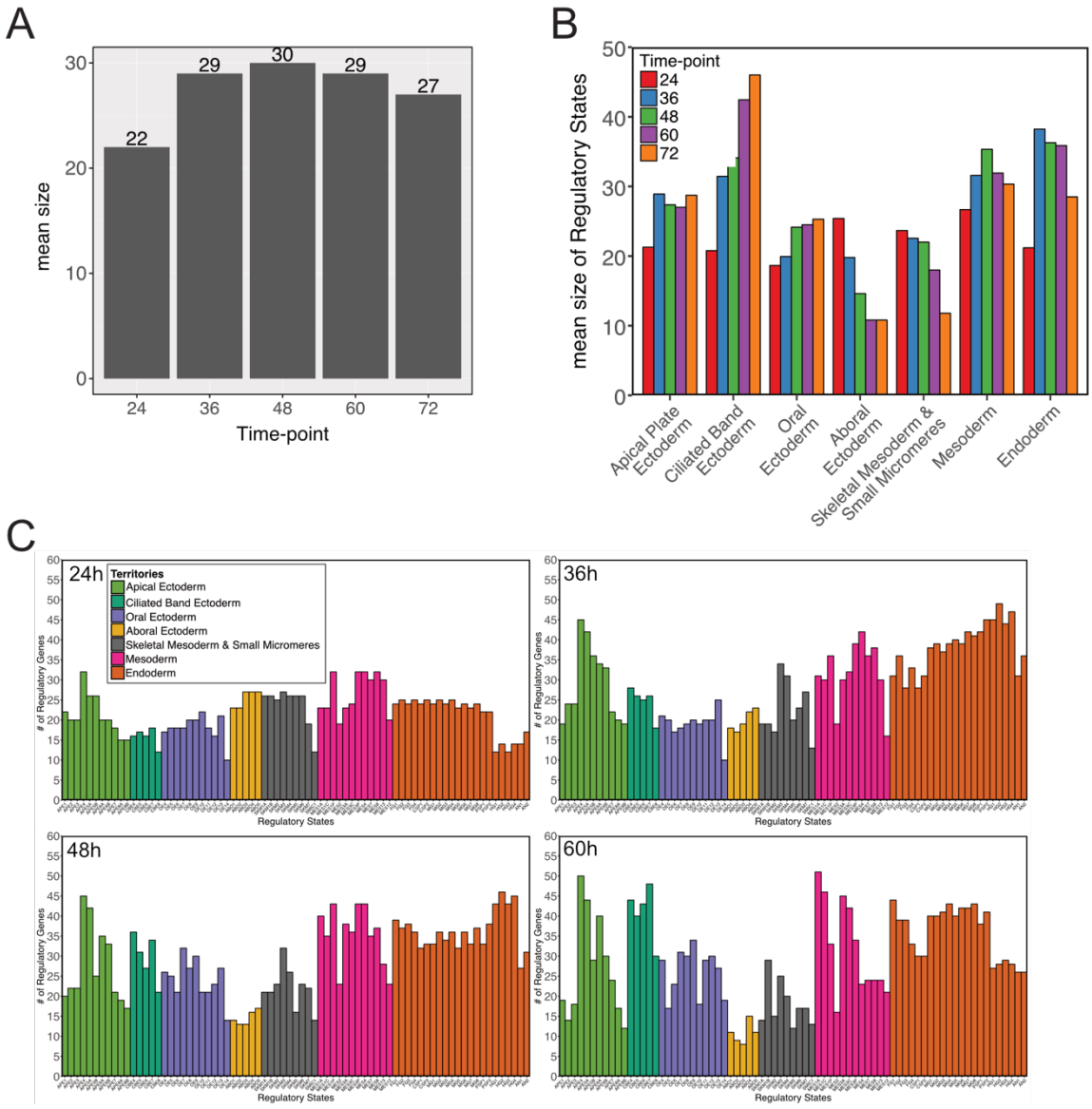


Figure S3.2: Size of regulatory states. (A) Mean size of regulatory states across developmental time. (B) Mean size of territorial regulatory states across development time. (C) Sizes of regulatory states at 24h, 36h, 48h, and 60h grouped by territory.

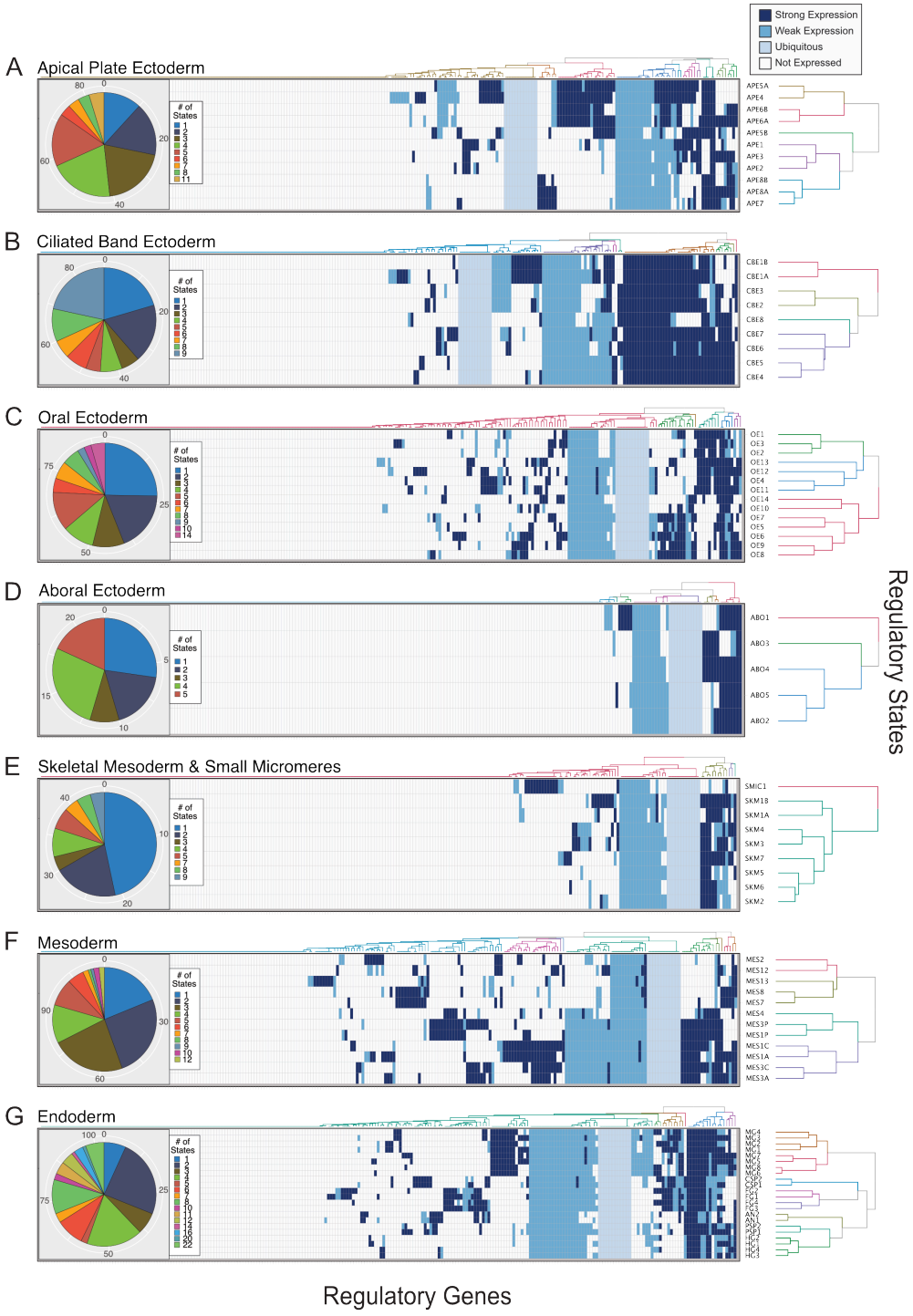


Figure S3.3: Clustered Boolean matrix of regulatory states grouped by larval territories. (A-G) Territory-specific regulatory states clustered by regulatory gene expression and grouped according to their larval territory. Pie charts as insets of clustered matrices, detail the distribution of specifically expressed regulatory genes in the number of regulatory state domains within a territory.

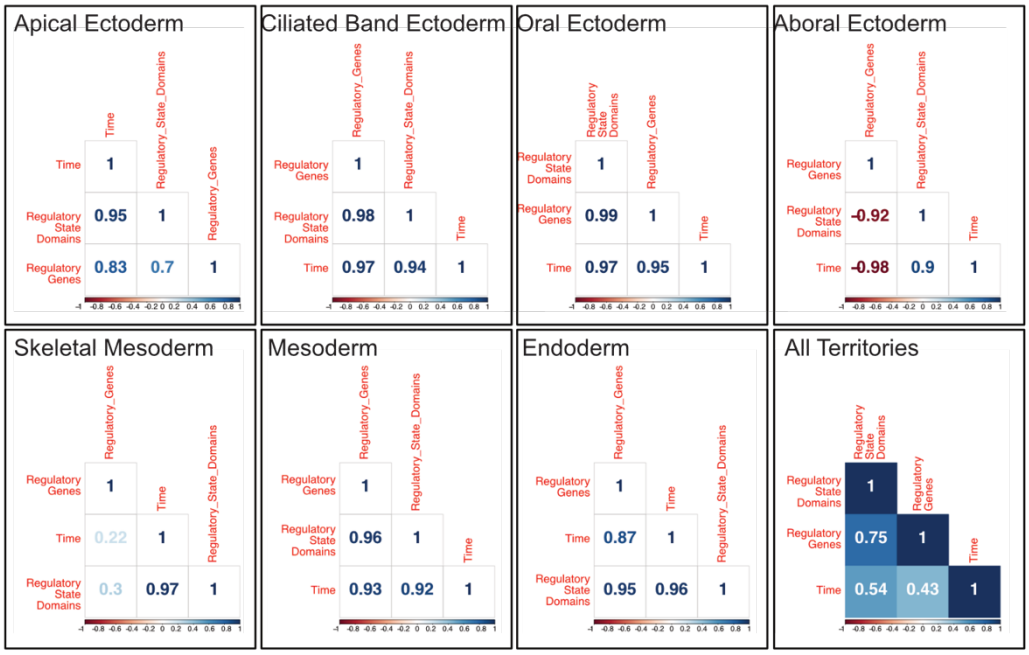


Figure S3.4: Correlation Plots. Correlation plots of embryonic territories with time, number of regulatory genes, and number of regulatory state domains as variables.

3.8.2 Supplementary Tables

Table S3.1: List of regulatory genes and their primer sequences used to construct RNA probe for WMISH.

Gene name	WHL transcript model	Forward Primer	Reverse Primer
Ac/Sc	WHL22.128311.0	AGTGACGTAGAGCCAGGTTAGG	TCATAAGCAACAAGTGTGACTGC
Af9	WHL22.754828.0	CAATGTTTGTCAGAGGACCAGAG	TAATACGACTCACTATAGGGAGACTTCTCTTTCTCGCTTGATGGAT
Ahr	WHL22.676940.0	CTGGATACAAAGGGCTCTGAAGT	CTGCTGTCTCTGACCACTCATCT
AhrL	WHL22.256860.0	TGCCATCTCTTTCTTCAGGATTA	CCATTGACTATATACGACGCACA
Alx1	WHL22.731056.1	GATGCCAAGAGGAAGAAGAGG	TAATACGACTCACTATAGGGAGAGGTGGTATATTTGATTTGGATTGG
Alx4	WHL22.731149.0	CCACAAGGACCATACAATTAACG	TAATACGACTCACTATAGGGAGAATCTGTCTCGATGGGTTCCAC
Ap2	WHL22.151633.6	AAAAGCTGCAAACGTCACCT	ACTTGTCCGCGATTGTCTTT
Ap4, Tcfap4L	WHL22.28828.0	ATCAATGCTGGCTTCCAGTC	CTACCGGGCTTCACTCTCAG
Apa, Hypp_2213	WHL22.771305.0	CTACGCAGGGAAACCTCAAG	TGACATGGCTGTCTGAAACC
Arnt	WHL22.478770.0	CTAACTTCCATGACCTCTCCTGA	TGATCTGTCAAGGTATCTGTTGG
ArxL	WHL22.134213.0	GCTATCTTCATTCATTCGTCGTC	GGTCGAGACACAAGTTTTCGTAG
Ash1	WHL22.755846.0	GTTAGGGAAGGGAAAGAAAGTGA	CTGTATTTGTGCTGGTACATCCA
Ash2	WHL22.96450.4	CTGAACCTGCACAATGACCTTAC	AGAGCGAAAACAAACACAGAGTG
Ash3L	WHL22.523564.0	Low EXP	Low EXP
Atbf1	WHL22.369567.7	AGGCAAAGGAATAAAGACTCACC	TGGATTGCTGTCTTCTTGGTATT
Atf2	WHL22.59755.0	ATGAGTGACGATAAACCATTTGC	TAATACGACTCACTATAGGGAGAATCTCTTTTCCAGATCAACTACCC
Atf6b	WHL22.176487.0	TAATTCTGGACTGGATGAAGAGG	TAATACGACTCACTATAGGGAGATTGATTGACTGTAGAGCTTTGAGG
Ato	WHL22.459901.0	GAGATGGTGAAAATGGAAGTGG	TAATACGACTCACTATAGGGAGAAGCTGTAGATGATGGTGAAGAGC
Awh	WHL22.9369.0	CATAACCATCCCATCAATAAATCC	TAATACGACTCACTATAGGGAGATGCACAGCACTCTATTTTCAATC
Bbx	WHL22.356734.3	TATAGATGGAGTGGTCCTCATCG	TAATACGACTCACTATAGGGAGAAGCTTTCTTGCAGTTGAATGG
Birc6/Hif1a	WHL22.609698.0	ATCCCTTTACTAGCGAGACTTGG	TAATACGACTCACTATAGGGAGACCTAAGAGCTGGGTGATACTGC
Blimp1	WHL22.5073.1	AGAGAAACCAGTTGTCGCGT	ATCTTGCTTCATGCTGGCTT

Bmal	WHL22.117405.0	TCCTGTTCTATGTAATCTTCGTCGT	GTAACGGAACATGGTTCTTTCCT
Bra, Bra_1	WHL22.600041.0	AAAACTCTCCACCCTTTCACC	CATCAACAGCCATTCAGTTACC
Brn1-2-4	WHL22.40221.0	CTTGATTCTCAGTTCAGCTACCG	TCTTCCATATTTTCAGCAACAGG
BsxL	WHL22.497248.1	TCGCTGTTCGTTGTCATATCTTT	TAATACGACTCACTATAGGGAGATTCTGTTTCATCCTCCTGTTCTG
Cdx	WHL22.642075.0	CATTACCAGGGCGTCACATCTAT	TCATCGTTGTTATGATCGGGAGT
Cebpa	WHL22.255599.0	CACAAAACTGCATATTGTCCAGTAG	CTAAGCCCTCGACACGTTTCTT
Cebpg	WHL22.744360.0	CCTGTTCCTGACTCATGCTAATG	GCTCAAGGATGTTCACAAGTCAC
Cic	WHL22.183787.0	ACTGGTCCACCAGAGACACC	AGGGGTGCTAGAGGCTTGAT
Cp2	WHL22.553348.0	GGTGCAGACAGGAAACACAA	ATTGTGGCAACACTTGGTGA
Creb	WHL22.278056.0	GTCCATTCCCATCGCTTCTA	GCCAGTGATTTTCCTTTCCA
Creb3l1	WHL22.19719.0	AAGCTCTTCTCATCCTCCTCATC	AGTGGTCCCAAGTACACAAGAAG
Creb3l3	WHL22.743120.0	GGAAGCATAGCAAGCAGACC	AGTCTCCGTTGAATGGGTTG
Crebzf	WHL22.126978.1	ATGGTCGAGAGTGCTTTATCTCC	TAATACGACTCACTATAGGGAGACCTGTCGCTTTGATATTTTTCAG
Ctcf	WHL22.295495.0	GCAGAGCTACAACCCGTCTC	TCGTGGCACTCTTCACATTC
CutL	WHL22.223201.0	AGGAGTTTGCTGAAGTTAAAAACC	TAATACGACTCACTATAGGGAGATCCCTCTGACTGGATACAATAGG
Dac	WHL22.169355.0	GATGCGAACCTGTTCTACG	TAATACGACTCACTATAGGGAGACAATTCAAAAGCTTGTGGCA
Dlx	WHL22.107309.0	GTATGAGGAACATTTACTGCTTGG	TAATACGACTCACTATAGGGAGAGTGATGCTGTTGAATGAGATGG
Dmrt	WHL22.521135.0	CGTCGAGAAACACGAACTACC	TAATACGACTCACTATAGGGAGAGGGCATTCACAAACATTTACG
DmrtA2	WHL22.114846.0	AAGGCCACAAGAGATACTGTCG	TAATACGACTCACTATAGGGAGAGATCATTGCAGAGGTGTAGGG
Dmtf	WHL22.529855.1	CAGAACGAGGTCATAAATGGAAG	TAATACGACTCACTATAGGGAGACTTGTGTTTCCGTCTGTTCCTAC
Dp1	WHL22.476402.0	TTAGAAGATACAGGGGCAAGTCC	TAATACGACTCACTATAGGGAGAAAATCTCCACTAATCTCGGTTGG
Dr	WHL22.544154.0	CACACAATCCTTCTCTGTTAGTCC	TAATACGACTCACTATAGGGAGATCTCATCTTTCAACCCAATATCC
Dri	WHL22.544150.0	CCCTTCTAAACGCTCTCTACTGG	TAATACGACTCACTATAGGGAGAATGAATACTGCATGGTGAAAAGG
E12	WHL22.548599.4	GACCTCAATTCTGGACAACCATA	CAGATAAAGCTGGATGCAGAAAT
E2F3	WHL22.307059.0	AAGGAAATCCATTCCATCTTGTC	TGTCATTCAAGGCAAAGAGGTAG
E2F4	WHL22.195483.0	CGTACAATACAGCCACGATCTTC	TGCCTGTGATTACATAACCCAAA

E78 bridge	WHL22.454455.1	ACGAGAAAAATCATAGGGAGGTG	TAATACGACTCACTATAGGGAGATTTCAGTAAGAGCAGTTGGTCATC
Ebf3	WHL22.113329.0	AACTCTCCAAACGCTCACCTC	GTTGGTCACTTTCGTGTTCTAGC
Egr	WHL22.280477.0	TAAGAAGTCCAAGGAATCAATCG	TAATACGACTCACTATAGGGAGAGATGTCTGCTGGTGATATGTGG
ElfA	WHL22.629740.0	TCCTTACCTTTTGTTTCTTGGTAGG	AAATCTTCCAGAACTTGACCCATA
ElfB/A	WHL22.629454.0	GTTGTGAACATGGGCTTTGTAAT	ACCTCTCATTGTTTCCTCATCAA
Elk	WHL22.626998.1	TCGAAGAACAGAACTTGATAGCTG	TAATACGACTCACTATAGGGAGATCAGAACCCTTGATAGTTGATGG
Emx	WHL22.113468.0	TGACCTGACTTGCTCTAACACAA	CCATAAACAGCAGGAAACTCAAG
Enz1/2	WHL22.223228.0	AACAAGAGCAGACGTTCATTGAG	TGATCTCCGTTTACCATCATCAC
Erf	WHL22.429793.2	GCGCTTCACCTACAAATTCAAC	GGTCGATTAAACCCTATGACGAG
Erg	WHL22.552472.5	TCTTTCGTTACTTTTTCCTCACTTC	TAATACGACTCACTATAGGGAGAATCACGTTCATGTGTCTTCTGAC
Err	WHL22.91797.0	ACACCAAACTATGCCCAGCTATC	GGGTTTCAACGATAGTCCATTTC
Ese	WHL22.110532.1	ATGTACCATGACCACGTCTCC	TAATACGACTCACTATAGGGAGAGTGCGTATCCAGCATCTTGAG
Ets1-2	WHL22.293603.0	AAACAAACCCAAGATGAACTACG	TAATACGACTCACTATAGGGAGATCGCATGACAAAAATAAGTTGC
Ets1/2	WHL22.238821.1	AAAACAGAACCAAAGGCACAAG	TAATACGACTCACTATAGGGAGATCGTAAATGTGAGGGTTGTTGTAG
Ets4	WHL22.613078.0	CGAAGTTTGCATTGTTTCACC	TAATACGACTCACTATAGGGAGATCTGTGTGCGTGTTCATATCG
Eve	WHL22.442145.0	ACGACTGAGAACCATCATCAAG	TAATACGACTCACTATAGGGAGAGAGTGGGTGGTAGTAGCTGAGTG
Evi1	WHL22.227650.0	ACCCTTAACCAAGAGTTCAAAGC	AAAGTAGGCCTCATCCTTCTCAC
Fog	WHL22.591129.0	CTGTTCAAGTGCCATCTGTAAGC	TAATACGACTCACTATAGGGAGATTTGTCTGCTAGTGGTTCAGAGC
Fos	WHL22.538480.0	LOW EXP	LOWEXP
FoxA	WHL22.439762.0	CATGGACTTGTTTCCCTTTTACC	TAATACGACTCACTATAGGGAGAGATTGATTTAACCGAGTGTTTGG
FoxABL	WHL22.615153.0	GACTCCGTCATCGTTATTTAAGC	TAATACGACTCACTATAGGGAGAGATGCAGAGCCCAATAGTTACC
FoxB	WHL22.743430.0	LOW EXP	LOW EXP
FoxC	WHL22.639627.0	GATGATAAGAAACCAGGCAAGG	TAATACGACTCACTATAGGGAGAGTAGCTGTAATGCGCTGATGG
FoxD	WHL22.41742.0	GACCTGATGGAAGGGATAAACAT	AGTTACACGTTGAGGTGGGTTG
FoxF	WHL22.639684.0	AAGGATGGAAAAACTCTGTTCGT	TAATACGACTCACTATAGGGAGACCGATGATGTGATGTAATTGGTAG
FoxG	WHL22.389872.0	GCGCTTTACTCGTCTTATTCTACC	TAATACGACTCACTATAGGGAGAGTCCTTAGTTGAAATGGGAAACC

FoxI	WHL22.535569.0	GTTTATCAGACAGGCAACATCCA	TCACAATGAGGCATCTTTACCTG
FoxJ1	WHL22.468365.0	CATCTTCATCAGAGTCCCATCC	TAATACGACTCACTATAGGGAGAGTTTCTTTTAGCGGGAGTTCG
FoxJ2	WHL22.714669.0	GTTGCAAGCAGCTACAATTTCTC	CCGTTAACACTCACACACACTCA
FoxK	WHL22.490768.0	AGTATTTGATCCGCAAGAACAGA	TAATACGACTCACTATAGGGAGACTGAGCACCTGTCGTAGTAGTGA
FoxL1	WHL22.639477.0	CAACCAAGACACATCTCACCTTC	GAGCAAATCGTTCATGCTAACTG
FoxM	WHL22.548104.0	ATCTGCTTCAAAGAAAGGCTCTG	GAATGTCCCACTCGGTTATTAGG
FoxN1/4	WHL22.604594.1	TGGCTCATTGCATGGATATTACT	AGATCTCACTGACAGGCAAACAG
FoxN2/3	WHL22.607384.3	GATGATGAACTGAAACCTCTTGG	CTTGGGAGAAATCAGTAACCACA
FoxO	WHL22.241099.0	GCCACCTGGACTTTAGTGGATAC	TTGTCCCAATGTTATCAAAGCTG
FoxP	WHL22.624355.0	GCAGAGTTTATTACAGGGTCAGG	TCTACCAGCTCATCACCAATAGC
FoxQ2	SPU_012384.1	TTGCCCAGAGTGACAGTAAGC	TAATACGACTCACTATAGGGAGAATGAAGAGGGTATCATGGATGG
FoxY	WHL22.399521.1	ACGAGGATCGAAGAAGAATGG	TAATACGACTCACTATAGGGAGACTTGGAAGAAGAAGTATGGTTGC
Fra2	WHL22.538597.1	AGCCATGTCACCTCGTTACC	GAACGCACTGCTTGAGAGC
FtzF, Nr5a	WHL22.765708.0	CGGAGACAAGGTGTCAGGTTAC	GGATAAGACAGTTAGGGCAATGG
Fxr	WHL22.622033.0	ACTCAGAAAATGCAGAGAAGTCG	TAATACGACTCACTATAGGGAGACAGAGTGGCTGTTGTTTAGGC
Fxr/Lxr	WHL22.40034.0	TGACCAGGGAAGACCAGATACTA	CATAAGCACAGAATGCAAATGAC
Gabp	WHL22.120896.2	AACACTCGATTCCTGTTTATTTCC	TAATACGACTCACTATAGGGAGAAGCTTCTCATAGTTCATGGATGG
GataC	WHL22.660411.0	CCAACAAGTTCCTACACGTTACC	TAATACGACTCACTATAGGGAGAGAGTGGTGATGAGGATAGTGTGC
GataE	WHL22.78013.0	GACATTGAGAGTCATCGTGAGG	TAATACGACTCACTATAGGGAGATGTTGTATCCATTCATCTTGTGG
Gbx	WHL22.737659.0	ACAGATAAGAGTCCCAGTGATCG	TAATACGACTCACTATAGGGAGAAGGGGAGACTGTAGATTGAAACG
Gcm	WHL22.54333.0	CGTACAATCCTCCTCTCCATTC	TAATACGACTCACTATAGGGAGATCGAGATGTCCACTATGTCCTG
Gcnf1	WHL22.765489.0	TGCCACATACACCCCTACAAG	TAATACGACTCACTATAGGGAGATAGCCACATCAATAATCTCCAGTC
Glass2	WHL22.204563.0	ATCATCATCACACCTGGACTCTC	CTACAAGACAGTAGCGGACGAAC
GliA	WHL22.242427.0	TGAGAAGACATACAGGAGAGAAACC	TAATACGACTCACTATAGGGAGATGATGACCTACCACTGAAATACG
Glis1, Z111	WHL22.242915.0	TACGAAGCTTGCAAAGATATTCC	TAATACGACTCACTATAGGGAGAGTTTGGTAACAGCATACCACTCC
GlisB	WHL22.583631.0	GGATCATCTGCTTCAATATCACC	TAATACGACTCACTATAGGGAGAGTGCATTTGTTAGGTTTCTCTCC

GlisC	WHL22.66475.1	TTCTATTGACTCTCCCCTTTTGG	TAATACGACTCACTATAGGGAGACTGTATGGCTTTTCCTCTAAGTGC
Grf	WHL22.274118.2	TTGATGACTTCCTCTCTCAGCTC	TAATACGACTCACTATAGGGAGACTGCATTTTCTCCATTTTCACTC
Gsc	WHL22.531818.0	ATGACTTCCCCATCGTTTACC	TAATACGACTCACTATAGGGAGATTTGTGTTAGGTGTTGAAAGTGC
Hairy2/4	WHL22.446908.0	CAAAATGCCTGTGGATACTAAACC	TAATACGACTCACTATAGGGAGAGTAGACTGGAATGAAATGACTTGG
Hb9	WHL22.107381.0	TCCGGGTATAGTGTGTCTCGAT	GAGACAGACAGACAGAAATGGACA
Hbn	WHL22.523959.1	CATGAAAACGTCTGGATACTGG	CATGAAAACGTCTGGATACTGG
Hes	WHL22.235339.0	ACTTCTTGCTTGTGCATTGAGAC	TAATACGACTCACTATAGGGAGATCTTCTGGATGAAGTCGTTGTTT
Hex	WHL22.626418.0	CTCCATGTAGCCGATCTATGAAC	GTTAAGGAACGTGAAGTGAATGG
Hey	WHL22.578435.1	LOWEXP	LOWEXP
Hey4	WHL22.235105.0	TTAGTGGTGGAAAATTAGCTGGA	CCCATTGAAGATGAACTCTTTTG
Hlf	WHL22.306206.0	GAAGACGACCGGATTCTCTACTT	TTGACAAGAGTCAGAAATGAACG
Hlx	WHL22.437349.0	GGACATTCAAGAAGTTGTTTTCG	CATTTACGAACACCCTCATTGTT
Hmbox	WHL22.285632.0	ATACACCATCGAGCAGATAGAGC	TAATACGACTCACTATAGGGAGAGCCATTCTCTTTTTACGGTTAGC
Hmg2	SPU_005572.1	AGTAAAGCTGAAGGGCAAGTGA	TAATACGACTCACTATAGGGAGATCTGAGGTGGAGTCGTCTGAA
Hmx	WHL22.152875.1	ACAGTAGCCGTGAACTATCAGGA	TAATACGACTCACTATAGGGAGAGCTGATGATGTAGACGATGTTGA
Hnf1	WHL22.83266.0	NeverOrdered	NeverOrdered
Hnf1aL	WHL22.520179.0	TGATGAAAACCCACCGAAGAGACAGC	TCATCCACAATGCAAGCTCTTCAGC
Hnf4	WHL22.35553.1	TAGCAGCATGCATGAGATGACC	GGGCTGTCCATTGAGGTCAGGT
Hnf6	WHL22.288683.0	CGCTAGAGAAGGCCATGAAC	ACTCTCCCACTCTGCCTTGA
Hox11/13b	WHL22.630181.0	GCCCCTACGCCAATACATTC	TGACCAACTGAGGGATGTGA
Hox7	WHL22.630154.0	TCGGGGCTGTTCAGAGGAG	TGAAGGAGACCAGCGAATAGAG
Nr2dbd	WHL22.755982.0	LOWEXP	LOWEXP
Ia1	WHL22.769122.0	ACCCTACAAGTGCAACTGAAACA	ATGGGCAAGTTGTGCAGTAATAA
Id	WHL22.467043.0	CGTCCTAATTTTAACGTGTTTGG	AATATGTCTTTCGGCGTTGTAGA
Irf4	WHL22.590973.0	AGTAGCAGTGGTGCCTAAATCGT	TGACAATAGAACTCGGGAAAACA
IrxA	WHL22.651130.0	CCACAAGTTATTGTTGTTGCTGA	AAAGTCTCTCAGTCATGGAGTCG

IrxB	WHL22.496846.0	AATGTCCATAACGCTACACATCC	TAATACGACTCACTATAGGGAGAGTCCGCAATAAGATAACAACTCC
Isl	WHL22.143854.0	CTCACTGTGCGTGCTAAACG	CCTCAGGCCACATAACTGCT
Jun	WHL22.318085.0	TGTTTGTATGTCAGTCGAAAACG	TAATACGACTCACTATAGGGAGACAAGTCAATAGTCCGAAGTGACC
Klf11	WHL22.59256.0	CACACATTCTCTATGCCACAAAC	TAATACGACTCACTATAGGGAGATGCTTGGTTAAGTGGTCTGATCT
Klf13	WHL22.198668.0	GTTAGAGATCTGTTGCAGGTTGG	ATGTTTGTGCCAAGTATTCAAGC
Klf15	WHL22.441254.0	TATTTAACGTCGGAAGGAGATGA	GTTGCGAGTTTTACGGTATCAAG
Klf2/4	WHL22.483798.0	GCAGCCTTTTTAGGAGAAACC	TAATACGACTCACTATAGGGAGATGAGATGAGATGACTTGGTGTAGG
Klf3/8/12	WHL22.210154.0	GCCCTTATCAAAATTACGAGAGC	TAATACGACTCACTATAGGGAGAACATGGAAAGACTCCTAAGTGACC
Klf7	WHL22.131060.0	GCTTGAAACCTGGTGAATTTCT	CCAACCAAAGCAGGGAAC
L3MBT	WHL22.447114.0	CGCTGCATCACTCTACATACATC	TAATACGACTCACTATAGGGAGAGGGGGATGACTTTGTTAGAAGAC
L3mbt_1	WHL22.727178.0	NeverOrdered	NeverOrdered
Lass6	WHL22.41285.0	TGGAGGAGTGTGTTTTACATCG	TAATACGACTCACTATAGGGAGATTACCAAGACGATACATCCAACC
Ldb2	WHL22.386521.3	ATTCCATCCTTTATTCCCCTTG	TAATACGACTCACTATAGGGAGAGACTTGCTCTCTGTTGTGAAAGC
Lef1, Tcf	WHL22.106048.0	AGAAGTGTAGAGCGAGGTTTGG	TAATACGACTCACTATAGGGAGATTTTTCTGTAACCCTGTTTCTGG
Lhx2, Limc1	WHL22.91758.0	GTATGAATGACCCCTGTGATCC	TAATACGACTCACTATAGGGAGAAGAGAGTTCAGTTGGTTGTGTCC
Lim1	WHL22.720614.0	AGACTACCTCACAAAGTGCCAAC	TAATACGACTCACTATAGGGAGACGTTCATATCCTTACCCGAGAG
Lmo2t	WHL22.43544.0	AGAAATATCGAGGAGCATCAGC	TAATACGACTCACTATAGGGAGAAGTCACCAGTCTCTGAAGTGTGC
Lmo4	WHL22.440254.0	CCCTACCTCTCGTTCTCTCTACC	TAATACGACTCACTATAGGGAGATCACAATGAAAAATAGTCGTCAGG
Lmpt	WHL22.543989.3	ATGGATGAAAACTTCCACAGC	TAATACGACTCACTATAGGGAGATACTCCTCGTCTAGCCTTGTGC
Lmx1	WHL22.448252.1	ATATTCGCTGCTTCCTTGACC	TAATACGACTCACTATAGGGAGATACTGATGATACCCCCTACATGG
Lox	WHL22.169409.0	GAACAATCCCGCGTACTATCACT	GCATCAATAATTTGTGGCTTCAC
Lztfl	WHL22.224817.0	CTAGGTGAAATTGCTGCTGTTCT	TAATACGACTCACTATAGGGAGACCATAGGGTCCAGTCTCTTGAGT
Mad	WHL22.80541.0	GCATTGACGAGAACTGAAACATC	CAAATGAGTACATGCAGAAACAGG
Maf	WHL22.652540.0	GTGATCAGACCATGTGTGAGGTA	GCAAGGTACACATACCCTTCTTG
Max	WHL22.335395.0	GAGGAGGAAAAATAACTCGCATC	TAATACGACTCACTATAGGGAGAAAAATTAGGGTCTTGTCTCTGTGC
Mbx1	WHL22.357718.1	NeverOrdered	NeverOrdered

Mef2	WHL22.322012.1	TCGGTCGAAAAGTCTATCAAACA	TAATACGACTCACTATAGGGAGAGGTTGTGGACTGAGTGTACTTCG
Meis	WHL22.2236.1	GCTGTTCATCAGACAGAAACGAG	TCATCGATTGTGAGTCCAAGTCT
Mitf	WHL22.677144.0	ACAAACTCTCAATGCCTAACACG	TAATACGACTCACTATAGGGAGACATCCTGTTTTATTAGCTCAGTGG
Mll3	WHL22.653011.0	AAATAATGGGAAGAGTTGAGAAGG	TAATACGACTCACTATAGGGAGAAATAATAGGTTTGTGGATTGGACAC
Mlx	WHL22.698342.0	TCCCAGATGATTCTAGTGTGAGG	TAATACGACTCACTATAGGGAGAGCTTTGCTGAAGACTAATGATGG
Mlx/IP	WHL22.419504.0	GACCATACATTCGGGTCATTTTA	ACTATTTTGCAAGAGCTCACCAG
Mnt	WHL22.677195.0	AAAGACATCCAACCTTAGCATCC	TAATACGACTCACTATAGGGAGACAGACTTGTAGACCTTGACACTCC
Msx	WHL22.119881.0	TAAGTCATTCCATCCAAAGCAAC	TAATACGACTCACTATAGGGAGAGTCTTGGTCATCTTCAAAAGCTC
MsxL	WHL22.404908.0	CTATTCGGGCTTTTCTTTAGTCG	TAATACGACTCACTATAGGGAGAAGATGAAAGTGAGAGATGATGATGC
Mta1, Mta1_1	WHL22.580176.1	TGGAGGTAGAGAAGAAAAGAAACG	TAATACGACTCACTATAGGGAGACACAAGGCTAATCTGACAAAAGG
Myb	WHL22.684838.1	CTTTCACTTGGTGTCAATTCCAG	TAATACGACTCACTATAGGGAGAGACCTTTGACCTTAGAGGACGAG
Myc	WHL22.687558.0	GCCTTGTCAAACTCTAATCGG	GTGGTCCAAATCCAATACGC
MyoD	WHL22.295720.0	CATCTCCACTTCTTCGTAACTCG	GATGTGTTTGAGCTTGTTTCTGC
MyoD2	SPU_006232	CGTCTGACGGCTACTCACTATCT	GCAACATCAACTGCAATCTTTC
MyoD3	WHL22.531810.0	GTTCCACGCTCAGGTTGATAG	GGCTGAAAGTTGAGCGATTAAA
MyoR2	WHL22.129805.0	AGACCTCTGCAAAACAGCATAAG	AGAAGCAAACGAAGCAGAAGTAA
Myt1-2	WHL22.447009.1	GAGTTCTGATGCTACCCAATCC	TAATACGACTCACTATAGGGAGACATCTTATCTTGCTCACCAGTCC
Ncoa4	WHL22.678301.1	AAAGGAAATTCACCTGCACTG	CTTTGGGTTGATGCTTCCATA
NeuroD1	WHL22.694980.1	GATATTGTCCCGTCTCATCTGTG	TATGGTTTACGGTGTAGGCATTG
NfatDS	WHL22.538507.0	AATGTATCATTGCATACGTGGTG	GCTCTTTGTTTACAATGGAGTGC
Nfe2	WHL22.621478.1	ACACAGCACAAACTGAACTTTCC	TAATACGACTCACTATAGGGAGATGGCATCTATTTTACGTTTACGG
Nfe2l	SPU_008752.1	AACCTGCCTGTCGATTCTTTC	TAATACGACTCACTATAGGGAGAATGGACTGCTCATCACTGTCTC
Nfia	WHL22.579762.0	AAAAGTCAGGCAGGAATGAAAAC	GAGAAGGAAGCTTATGGAACACA
NflL3	WHL22.733532.0	GATTTCACACTCCTCCAAACATC	CGAAGTCTCATTACGTCTTTCGT
NfKBDS(2)	WHL22.672896.0	CTGGTAGGGAAGCACTGTAAGG	TAATACGACTCACTATAGGGAGAGTTGAAGAGGTCATTTGGTTGC
Ngn	WHL22.677570.0	AGCAGTAAGGAGGTGAAAGGAAA	AGTCATCCGCAACACTTTGATTA

Nk1	WHL22.152063.0	CTGGAGAACCTAAATCGTACCG	TAATACGACTCACTATAGGGAGACCTGAGTGCGAGAAATAATGG
Nk2-2	WHL22.739246.0	TCTTTTCTTCTCCTGGTTTCCAC	TAATACGACTCACTATAGGGAGACTGACATACACGCTGATGCTG
Nk7	WHL22.567485.0	GATAGACATCACGCCCACTACC	TAATACGACTCACTATAGGGAGATTCTGGTTTGAGGATACCTTCC
Nkx2.1	WHL22.739581.0	AAGCAGCAGAAGTACCTGTCG	TAATACGACTCACTATAGGGAGAATGACTATTGTGTGGTGCAAGC
Nkx3-2	WHL22.329059.0	ACCTGTAGTCAATCGTCACTCGT	GCTCTCTCATTTCCCTCCATACT
Nkx3.2	WHL22.339351.1	CTAACGGCTTTGCATGATAACAG	CTCGAAATGTCCAAGTCCAAAAT
Nkx6.1	WHL22.567494.0	CCAGGGAAGGTATAATAGCCACT	TGACTGACTGTGGACCAATCATA
Not	WHL22.632281.0	AATCCTTGGAAGACAGCACTTG	TAATACGACTCACTATAGGGAGACTGAACGTACTGGTTGTCGTG
Nr1h6b	WHL22.771234.1	TGAGAAAATGTCGGTGGACTACT	ATGATTCCAGGATCTCTCTTTCC
Nr1h6c	WHL22.609334.0	ACGATTCTAGGGAGCATTACTGG	TAATACGACTCACTATAGGGAGATCCTGTTCCTGTACTTTGACAGC
Nr1m2	WHL22.383507.0	GATCAGGTCCTCACTACGAGTTG	CTGACCTTCATACCGACTCAAAC
Nr1m3	WHL22.21036.0	GTCACTTTGTATTTCCCTCACCA	CAAGACATTGACCATGCACTCTA
Nr1x	WHL22.581390.0	AGACGACAGATGGAATGCTATGA	GTGCACACACCTATTCAGTGACC
Nr2c	WHL22.96335.0	ATCATCTTCATCCTCATCGTCCT	CGCTCTGGATACCTTTACGATCT
NtL	WHL22.259272.0	TCGACATCACCACCTACATACG	TAATACGACTCACTATAGGGAGATAAATCAACCGCTAGAACACTCG
Oct1/2 long	WHL22.67463.1	CTCAACCTGAGCTTCAAGAACAT	CCAGGGAGTGGAGATTGTAAGTA
Otp	WHL22.286934.0	ATCGTGTTATGGTCGAGTTGTTT	GCTAAGCAGAAATGAGCCATGTA
Otx	WHL22.532435.0	AACAGCAGCAACAGCAACAG	AGAGCTGCGTTCAAGGTCAT
Ovo	WHL22.220905.0	TACATGATTTCGACCATCCTAGC	TAATACGACTCACTATAGGGAGAATTTCCTGACTGATGCTATACCG
P3A2	WHL22.405480.0	AGTATGACGGATGATGTCTCTGC	TAATACGACTCACTATAGGGAGAATGTGGTTACTGTTTGCCCTTC
Par1	WHL22.718026.2	AGAGTGTCCTAAGTTCCCCTGAT	TAATACGACTCACTATAGGGAGAAGAGGTGTAGCATTACCATGAGC
Pax2/5/8	WHL22.619292.0	CGGAAAGATAAACGACAAACAAG	TGCTTCTAAACTCGCATGGTATT
Pax4L	WHL22.82981.1	GTGGCTAGCTGAAAACAGTATGG	AGATGGGCAAGTAAGTCCCTAGA
Pax6.1	WHL22.585512.0	CGCAATCAGAGAAGACAGCA	TTAGCCAGCAAGAAGGGAAA
PaxB	WHL22.698752.0	CCTCTCTGCAATCCTTCATTCTA	CCAAGAGTTCCTTTATGGTGTAGG
PaxC	WHL22.535609.0	CGTCAGCAAGATACTTAGCAGGT	TAATACGACTCACTATAGGGAGAAGGTGGGTTTGATACCACTTCTT

Pbx1	WHL22.75202.0	AGTATCAGGGGTGCACAAGAAG	GATTTGGCAGCTGGTAACTGT
Pcbd2	WHL22.529533.0	AATAGGTCAACTCAGGTCATCTGC	TAATACGACTCACTATAGGGAGAGCCAGATAAGATTCATGCAAGG
Pea	WHL22.166439.0	AGAGGGAAGAAAGACATTTTTGG	GTTACAAGGAACTGCCAGAGTTG
Phb1	WHL22.302399.0	TACAGCATTGGAGCAACACC	TCAGGCTTATCGAGCAAGGT
Phb2	WHL22.521451.0	CGGAGACAGCCGAGTTTAGTAG	TAATACGACTCACTATAGGGAGAGGGAGGCATAAGTATTGACACG
Pitx1	WHL22.704986.0	TCGGTTGTTGATTTTCAGTCATC	GGTCACGAGTCGATATACGTTTG
Pitx2	WHL22.11036.0	TTTGTGTAGCTTTCTCCCCTCTA	GAACGGAGATCAAGTGAAGAAGA
Pknox	WHL22.655290.0	TTGATGCCTGACTGACACATAGT	TTGCATGTCTTTTCATGTCTCTG
Pou4f2	WHL22.738139.0	GACATGACTGAACGTCATCAAAA	CATGTAAGGCAAAGGAATAGCTG
Ppar1	WHL22.46852.1	CGAGTGAAGGTGGAGAGGAG	CTTAATGAGTGGCGGAGGAC
Pric	WHL22.636100.0	TAAGGATGTTCAGTAGCCAGAGG	TAATACGACTCACTATAGGGAGAGGTCTCATTCCATTCTGATGC
Pric2	WHL22.636314.1	AGGACTTGACCTTCACTTCTGG	TAATACGACTCACTATAGGGAGAAGTACGGATGGCTGTTCTTAGC
Prox1	WHL22.531966.1	CTGACGGAATTTACTCACACCTC	TTTATCCAGTTACGTCTGGCTTC
Ptf1a	WHL22.476207.0	GTTCTTGAAACTTCTTCGGCATT	AAAGTGTCATTGGGGAGATGAGT
Pu1	WHL22.15960.1	GACGTGTCATCTCTATGAGCTACC	TAATACGACTCACTATAGGGAGACTTTATGGTCACAACCGATGG
Rar	WHL22.595420.0	GTCCCTCCCTACACAGATGC	CCTTCACTGATGAGACAGATAGC
Rara	WHL22.376236.0	Low EXP	Low EXP
Rel	WHL22.499256.0	GCTAGCTCTCACAATGAGTGGTT	GACTTCATGTCAACAATGTCAGC
Reverb	WHL22.595563.0	AAGGGTGTAAGGGTTTCTTTCG	TAATACGACTCACTATAGGGAGATCATAATTGGCATGATTGTTCC
Rfx3/2	WHL22.19679.1	AGCCAGTCTCCTTTTACTCCATC	CATCTGATTGATCTGTGACGTGT
Rhox3	WHL22.277630.0	AAGGAACAGACGACTTCGGTACT	AAGTGCATTTCCATTAGCGTTCT
Riz	WHL22.63053.0	TGGAGTTCAGACAGTCACAGATG	GATTCACTTTGGGCACATTTAAC
Rora	WHL22.499606.0	GGTGAGCAAGCTGGACTTAGTAG	TGTAGTCTTTGAAGGCTCCTCTG
Rreb1	WHL22.421612.0	TTTTGAATGTCCAGAGTCCTTGA	GCTGGAACACAGTGATGTAGGTC
Rrm	WHL22.747798.0	TGATCTAAGCGTGTTTGATTTCC	TAATACGACTCACTATAGGGAGAGTTGTTGTCTCTGCTTCTTCTGG
Runt1	WHL22.425395.0	TACAACCGTGCAATAAAGGTCAC	CAGGGTCTTTGATAACATTGCTG

Rx	WHL22.523971.0	AAGAGCAACGGTGGAATAAAAAC	GCTGATTATACGTTCAGGCAAGA
Rxr	WHL22.717794.2	CTCAAGACAGACGCAAGTAACCT	TGAGGAATCAACAGGAGATGAAG
Sage	WHL22.125033.0	CGTTTTACAACAAAGTCCTCTCG	GTGGTGATGTTGCTGTAGAGGTT
Sage1	WHL22.633149.0	CGTTTCAAGACTTTGGAATAGGA	GGATCATACTCGTTTTCATGCTT
SatB1	WHL22.589656.3	CAAAACGTGTCAAAGATTCAGG	TAATACGACTCACTATAGGGAGAAATGTTCCTCCTTCTGGCTACC
Scl	WHL22.399764.0	TCACATCACCCTCTCTCATCC	TAATACGACTCACTATAGGGAGAGGTTGTCATCGCTAAAATAAACG
Scml1	WHL22.128919.0	AGCGACAACAAGAACGACTTTT	TAATACGACTCACTATAGGGAGAGTAATGTCCGGTGGCTTTACAC
Scratch	WHL22.758741.0	GTACGAAGCATTTTTGATTACGG	TAATACGACTCACTATAGGGAGATGCACTTGACTGATAACACAACC
ScratchX	WHL22.768151.0	CATCATCAGCTTTCAACTCAACC	AGGCACGTATAAAGTCAACACGA
Shr2DOWN	WHL22.302882.1	TGAGCTGTTCACATTAGGATTGG	GGAACTGGTTGAAAGATTGTTCC
Six1/2	WHL22.121485.0	GCAATAACTTCTCACCGCATAAC	TAATACGACTCACTATAGGGAGAGTTCATGTTTTTCTTTCCGACTG
Six3	WHL22.121654.0	CTCATAGACACACCCCAGCA	AGGATGGTGGGATCTTTCTTC
Smad_IP	WHL22.553144.0	AGAGCATATCAAATACCGACACG	TAATACGACTCACTATAGGGAGAGAACTCTCGTCCACAATACTTGC
Smad1/5/8	WHL22.347626.0	GACTGTTGTGATTATGCCTTTGG	TAATACGACTCACTATAGGGAGAACATCTTGGTGAGTTCGTAGACC
Smad2/3	WHL22.242408.1	TCTCAAGAAGGATGAGGTCTGTG	TCAGGAGGAACCATCTGTGTAAG
Smad4	WHL22.57163.0	TGAAGCAGTGTTATGCTCAGATG	ACATCCCCCAAGACTTTTTGAAG
Smad4_1	WHL22.107471.0	TGTCTTGATTCATGGTGTGACAG	CTGTAACCTGCCATCCAATGTC
Smad5	WHL22.72583.1	TCTTCGACAGCTACGGTACACAG	CTCAGTTTCTTCAGTCGATGGTG
Smad6/7	WHL22.579194.0	CTTATTCTCATCCGAGTCTGTGG	TAATACGACTCACTATAGGGAGAGATCAGTTCAATCCAACAAGGAC
Snail	WHL22.131363.0	ATACACTTGTTGAGCGGCAGTA	TAATACGACTCACTATAGGGAGAGTATTCCTTGTCGCAGTATTTGC
SoxB1	WHL22.104606.0	CAGGTATCATCAATCCACAGACC	TAATACGACTCACTATAGGGAGATGCGTAGAGCAACTCAATAGTGTC
SoxB2	WHL22.104525.0	ATCAGAGACTTTCCCCATCATC	TAATACGACTCACTATAGGGAGAGTGTTGCACAGTCCTTGTTGAC
SoxC	WHL22.622787.0	GTCACAAATCGAGAGGAGACG	TAATACGACTCACTATAGGGAGATCTGAGTCTATGAGTTCGCTTACC
SoxD1	WHL22.118185.2	CTCACGAAGCAGAGATAGTCAGC	TAATACGACTCACTATAGGGAGAAATGCTTGTAAGGTGGGTATGG
Sox6L	WHL22.118140.1	GCTTGTCTGATCTACTCCATTCC	TAATACGACTCACTATAGGGAGACTTTAGGGCTACCTGTCATTGG
SoxE	WHL22.466465.0	GAGAGTGCTGAATGGATACGACT	TAATACGACTCACTATAGGGAGAGCATACTTTGGACTGGCGTAG

SoxF	WHL22.57106.0	TCTACGACCCCCTCTCAAGTAAC	TGATCCAGTAATGAGGAACATGG
Sp2	WHL22.380228.0	GTTAGTGTAAGTCAAGCCCAAGC	TAATACGACTCACTATAGGGAGACATTGATCTGTTGTATGCTCTGC
Sp5/Z199	WHL22.380187.0	CTAGCCCCCTTGCAATGTTAG	GTTCTTCGCCTTGTGGATTTAAC
Spz12	WHL22.446293.0	AAGCTCTTCGGTCAGAGTTCC	TAATACGACTCACTATAGGGAGATTGTTGCACACTTCACACTGG
Srebp	WHL22.192617.0	TTGCCTTACTTCTCAACAACCAA	TTCATCCAGAGAGATGCTCAGTC
Srf	WHL22.60780.0	CAAGAAAACGAAAGGAAGAGTGA	TAATACGACTCACTATAGGGAGATCTGATGTAGGTGTGTAGCAGCA
Stat	WHL22.738007.1	CGGTACCTGGACCTATCTTGAAT	CTTGGAAGACAAGTTCACCACTC
Su(h)	SPU_021566	TATACTCCTCGTCTGCCATCG	CTCATCCTCATGGTGTTGTACG
Tbr	WHL22.503644.0	TTGACATGAAGAAAGAACTTGAGC	TAATACGACTCACTATAGGGAGAGGTATGGGTGCTGATAGTGACC
Tbx2/3	WHL22.457020.0	TCACAAAAGAGGAACAGAAATGG	TAATACGACTCACTATAGGGAGAGGGATGGGTGTCTAAATAACTCG
Tbx20	WHL22.730224.0	GGGGTTAGATCGCTACAATAACTTC	TAATACGACTCACTATAGGGAGAGTCCGTACAACTATCACAAACTGG
Tbx6/16	WHL22.504191.1	GAAGGAGAAGTCGGATAACTGC	TAATACGACTCACTATAGGGAGAGTAACAATAATCCCCGAAATGC
Tead4	WHL22.355310.1	CACTGCCTTACAGAACAAAGCTG	CACTGCTCATCTCAATGGGTTAG
Tel	WHL22.589178.0	CAGCGACTCAGGTCATAGTTCAC	TATGTTAGCTTCTGCCCCTGTTC
Tgif	WHL22.614286.0	TGGTGGTTGAAGAAATAGTTTGG	TCTGTCTTTCCTTACGCTGTACC
Thr&B	WHL22.211956.0/1	TCATGCTCAGTGGTGAAAGTAAA	CAAGGAATACAAGTCCCTCCATT
Tle1/Groucho	WHL22.510742.0	TGTCCTCTGGTCTCTTACAGCTC	CTCTTTGGACTGGAATATGCTTG
Tll	WHL22.282560.0	GTGGGTAGCTGCAGTGGTTT	AACCAAGTTGTGGTGGCTGT
Trx2	WHL22.568858.2	GCAAGAAGAGTTTGGTTCCAGTC	TAATACGACTCACTATAGGGAGACCTCTCTATGCTCTGTGACTGTC
Tsc22D2	WHL22.509140.2	CAAGCAATGGACTTGGTTAAGAG	GCAGGATATCATTAAGGCAGATG
Tulp4L	WHL22.608861.0	GATGTCCAAGATGTCGTGAAGAG	CGACCAGAATGCAAAATGAATAC
Unc4.1	WHL22.186660.2	AGGACGACCAAACTAAGAAAAGG	ACCTCCACTCCACACTTGTAAGA
Usf	WHL22.467508.3	TAACATATCGCGTGGTACAGGTC	CGCACACACACAACAAAAATATC
Whsc1	WHL22.117840.0	AACTGATTCTGATCTCGATTTGC	TAATACGACTCACTATAGGGAGATCTTCTTTCTGCTTCTTGTACGC
Xbp1	WHL22.490709.0	CCTATTCGTGAAAACTCCATCAT	CACATGTCAAGGAAGTGTCTGAG
Yyi	WHL22.224218.0	GATGATGTGCCAAGAACAATAGC	TAATACGACTCACTATAGGGAGACTGCTGAAAAGAACAATAACAAAGG

Z121, Osr	WHL22.496011.1	TACACCTCCTCTCACACCAACAA	TGAACTCCTTCTTGGCTCTCATC
z141	WHL22.75688.0	ATATCTGCGTCAAATGCGACTACT	CTTCCATAGCTTCCATGTCCTCTA
z166	WHL22.717588.0	ATTCCCGCATCCCTTTATTC	CACACAATGGTGATGCAATTT
z204	WHL22.247138.0	CAAATGGACATACGGGTACACAC	ATTTACTTCCTCTGGCTGCATGA
z54/spalt	WHL22.150131.0	AATTCCTCCACCTCTCCCTTATC	TCTGGGCTAATTCTCTCTTGAGG
z55	WHL22.639210.0	CATTACACAGCACGTTACAGCAG	CATGACAGGTGAACTCGATCTTG
z67	WHL22.684989.1	TCCTAGCAAGAATCAGGAAGACC	CAGGAACATCTGGCTTGTTTATG
z92	WHL22.368434.0	ACAATGAGAATGAAGAGAGGAACG	CTTCTTGAGATCATGGTAGCGATT
Zic	WHL22.331651.0	CAATCGCGTTTCAGTTGACTAC	ACGTACCATTCACTCAAGTTCGT

Table S2.2: Regulatory state domains and their associated morphological structure.

Regulatory State Domain	Morphological Association	Larval Territory
APE1	Oral Medial AP	Apical Plate Ectoderm
APE2	Oral Lateral AP	Apical Plate Ectoderm
APE3	Photoreceptors	Apical Plate Ectoderm
APE4	Central Medial AP	Apical Plate Ectoderm
APE5A	Central Ring AP	Apical Plate Ectoderm
APE5B	Central Aboral Single Cells	Apical Plate Ectoderm
APE6A	Central Lateral AP	Apical Plate Ectoderm
APE6B	Central Distal AP	Apical Plate Ectoderm
APE7	Aboral Central AP	Apical Plate Ectoderm
APE8A	Aboral Lateral AP	Apical Plate Ectoderm
APE8B	Aboral Distal AP	Apical Plate Ectoderm
CBE1M*/APE4	Central Medial CB	Ciliated Band Ectoderm
CBE1R*/APE5A	Central Ring CB	Ciliated Band Ectoderm
CBE1S*/APE5B	Central Aboral Single Cells	Ciliated Band Ectoderm
CBE2*/APE6A	Central Lateral CB	Ciliated Band Ectoderm
CBE3*/APE6B	Central Distal CB	Ciliated Band Ectoderm
CBE4	Upper Vertical CB	Ciliated Band Ectoderm
CBE5	Lower Vertical CB	Ciliated Band Ectoderm
CBE6	Inner Arm CB	Ciliated Band Ectoderm
CBE7	Arm CB	Ciliated Band Ectoderm
CBE8	Sensory Neurons	Ciliated Band Ectoderm
OE1*/APE1	Central Oral AP/Near Apical Central OE	Oral Ectoderm
OE2*/APE2	Oral Lateral AP/Near Apical Lateral OE	Oral Ectoderm
OE3*/APE3	Photoreceptors	Oral Ectoderm
OE4	Upper Vertical Band OE	Oral Ectoderm
OE5	Upper Stomodeum OE	Oral Ectoderm
OE6	Cheeks OE	Oral Ectoderm
OE7	Upper Mouth/Edge OE	Oral Ectoderm
OE8	Lower Mouth/Edge OE	Oral Ectoderm
OE9	Lower Stomodeum OE	Oral Ectoderm
OE10	Post OE	Oral Ectoderm

OE11	Lower Vertical Band OE	Oral Ectoderm
OE12	Arm OE	Oral Ectoderm
OE13	Inner Arm OE	Oral Ectoderm
OE14	Post Oral Neurons	Oral Ectoderm
ABO1	Anterior AE	Aboral Ectoderm
ABO2	Posterior AE	Aboral Ectoderm
ABO3	Apex AE	Aboral Ectoderm
ABO4	Anal Distal AE	Aboral Ectoderm
ABO5	Anal Medial AE	Aboral Ectoderm
SKM1A	Aboral Apex	Skeletal Mesoderm & Small Micromeres
SKM1B	Aboral Vegetal Cluster	Skeletal Mesoderm & Small Micromeres
SKM2	Vegetal Lateral Body Rods	Skeletal Mesoderm & Small Micromeres
SKM3	Oral Vegetal Clusters	Skeletal Mesoderm & Small Micromeres
SKM4	Lower Arm Rod Extensions	Skeletal Mesoderm & Small Micromeres
SKM5	Oral Vegetal Horizontal Rod/Ventral Transverse Rod	Skeletal Mesoderm & Small Micromeres
SKM6	Oral Vertical Rods/AnteroLateral Rods	Skeletal Mesoderm & Small Micromeres
SKM7*/MES7	Upper Arm Rod Extensions	Skeletal Mesoderm & Small Micromeres
SMIC1	Small Micromeres	Skeletal Mesoderm & Small Micromeres
MES1A	Anterior Left Coelomic Pouch	Mesoderm
MES1C	Central Left Coelomic Pouch	Mesoderm
MES1P	Posterior Left Coelomic Pouch	Mesoderm
MES2	Hydropore Canal	Mesoderm
MES3A	Anterior Right Coelomic Pouch	Mesoderm
MES3C	Central Right Coelomic Pouch	Mesoderm
MES3P	Posterior Right Coelomic Pouch	Mesoderm
MES4	Circumesophageal Muscles	Mesoderm
MES5A	Blastocoelar Cells 1	Mesoderm
MES5B	Blastocoelar Cells 2	Mesoderm
MES6	Pigment Cells	Mesoderm
MES7	Non-Skeletogenic Mesoderm Skeleton	Mesoderm
FG1	Anterior Oral Foregut	Endoderm
FG2	Anterior Aboral Foregut	Endoderm
FG3	Posterior Oral Foregut	Endoderm

FG4	Posterior Aboral Foregut	Endoderm
CSP1	Oral Cardiac Sphincter	Endoderm
CSP2	Aboral Cardiac Sphincter	Endoderm
MG1	Oral Sub Sphincter 1	Endoderm
MG2	Aboral Sub Sphincter 1	Endoderm
MG3	Oral Sub Sphincter 2	Endoderm
MG4	Aboral Sub Sphincter 2	Endoderm
MG5	Anterior Oral Midgut	Endoderm
MG6	Anterior Aboral Midgut	Endoderm
MG7	Posterior Oral Midgut	Endoderm
MG8	Posterior Aboral Midgut	Endoderm
PSP1	Oral Pyloric Sphincter	Endoderm
PSP2	Aboral Pyloric Sphincter	Endoderm
HG1	Anterior Oral Hindgut	Endoderm
HG2	Anterior Aboral Hindgut	Endoderm
HG3	Posterior Oral Hindgut	Endoderm
HG4	Posterior Aboral Hindgut	Endoderm
AN1	Oral Anus	Endoderm
AN2	Aboral Anus	Endoderm

A p p e n d i x A

DEVELOPMENTAL SPECIFICATION SCHEMATICS OF SEA URCHIN TERRITORIES

Index of Territories

Apical Plate Ectoderm

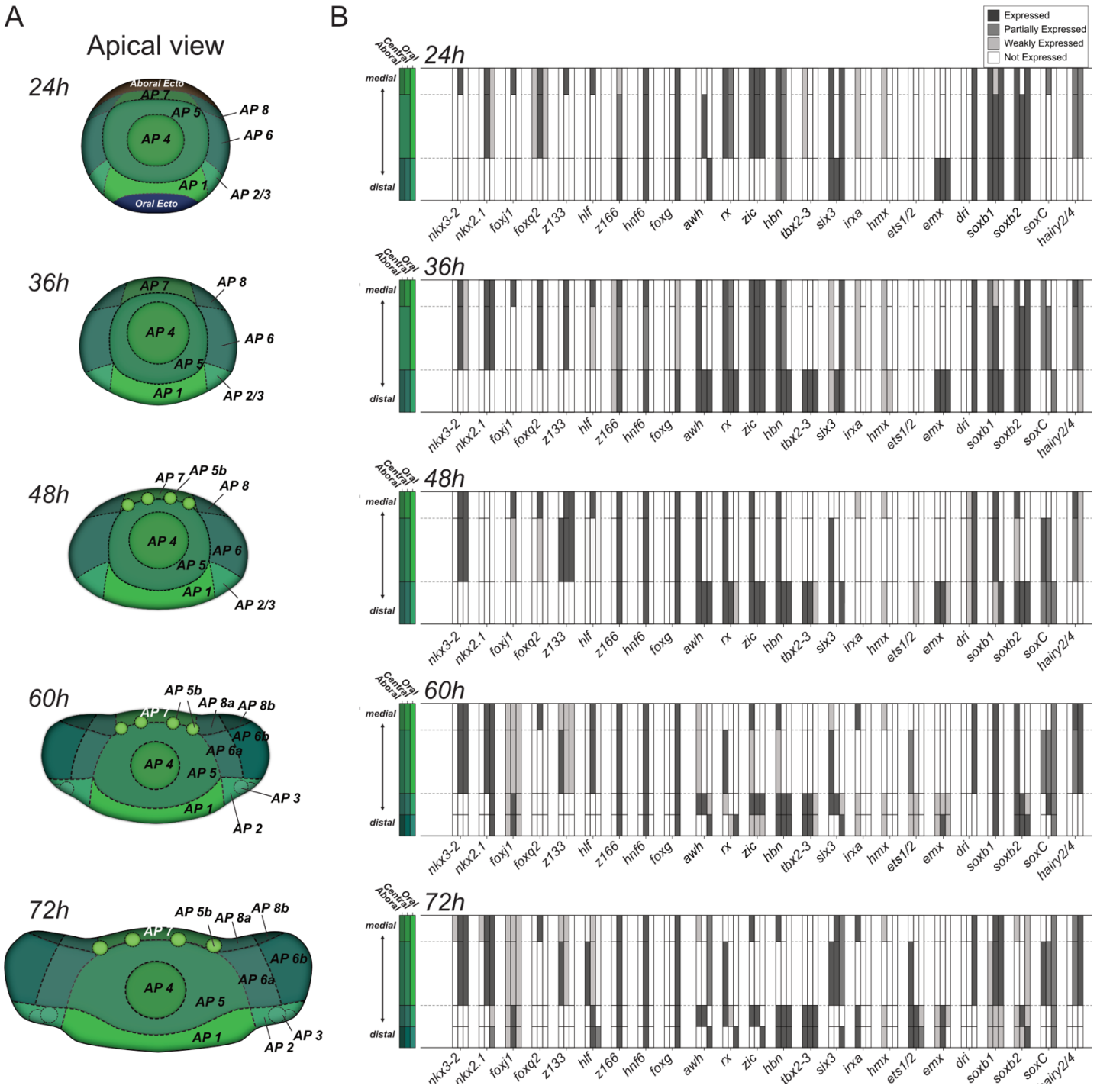


Figure A.1: Schematic of apical plate specification. (A) Embryonic staging schematics represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Embryos are drawn in an apical view and are oriented with the oral ectoderm facing down. Identified regulatory state domains within the apical plate ectoderm are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of regulatory genes expressed within the apical plate ectodermal regulatory states at corresponding developmental time-points. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the apical plate ectoderm and represent the composition of regulatory genes responsible for their specification. As a whole, these schematics (A-B) demonstrate the specification of domains within the apical plate ectoderm along developmental time. The emergence of a novel domain in (A) can be attributed to the establishment of a new regulatory state in (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

Ciliated Band Ectoderm

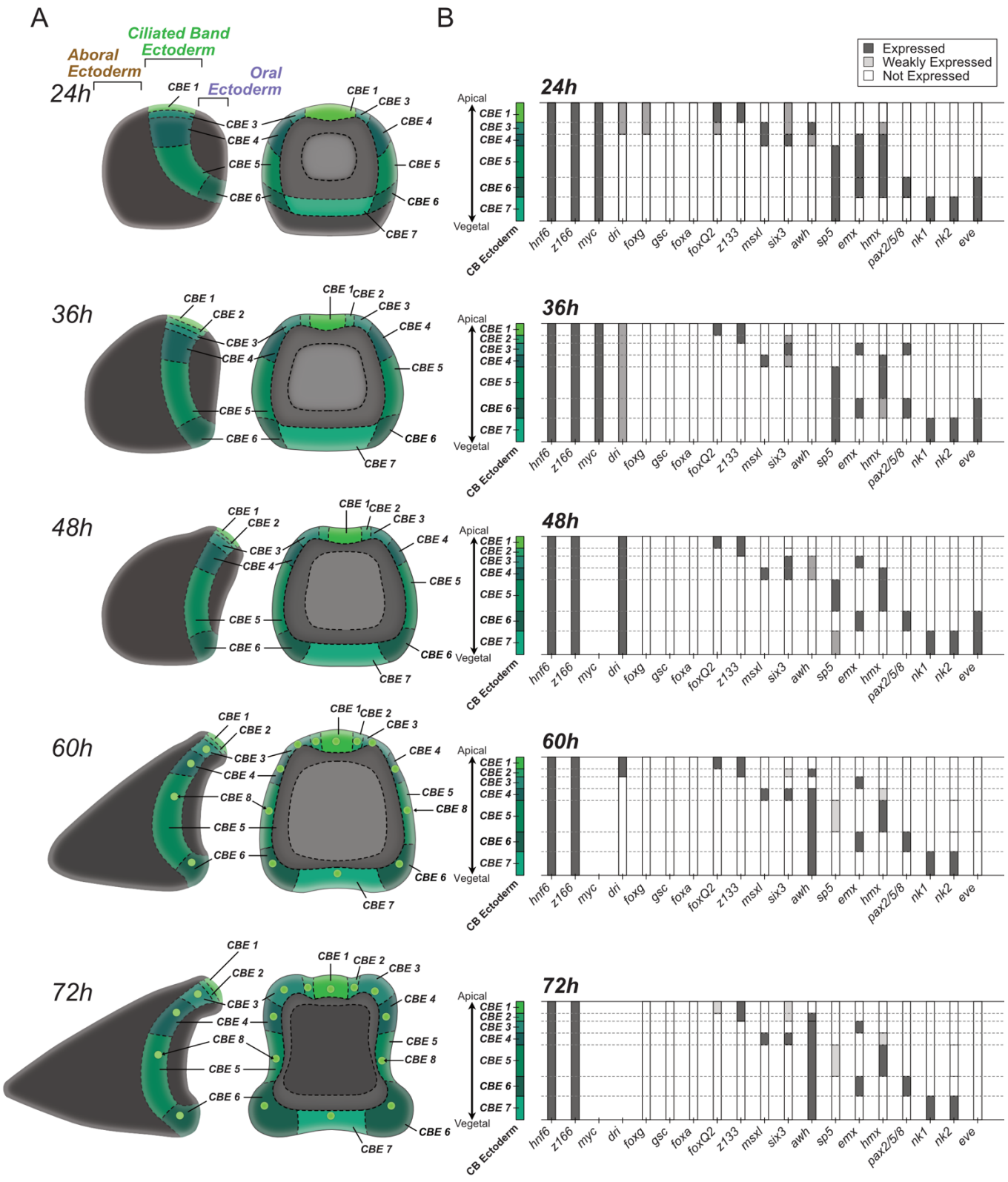


Figure A.2: Schematic of Ciliated Band Ectoderm Specification. (A) Embryonic staging schematics represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Embryos are drawn in lateral and oral views and are oriented with the oral ectoderm facing right and out respectively. Identified regulatory state domains within the ciliated band ectoderm are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of regulatory genes expressed within the ciliated band ectodermal regulatory states at corresponding developmental time-points. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the ciliated band ectoderm and represent the composition of regulatory genes responsible for their specification. As a whole, these schematics (A-B) demonstrate the specification of domains within the ciliated band ectoderm along developmental time. The emergence of a novel domain in (A) can be attributed to the establishment of a new regulatory state in (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

Oral Ectoderm

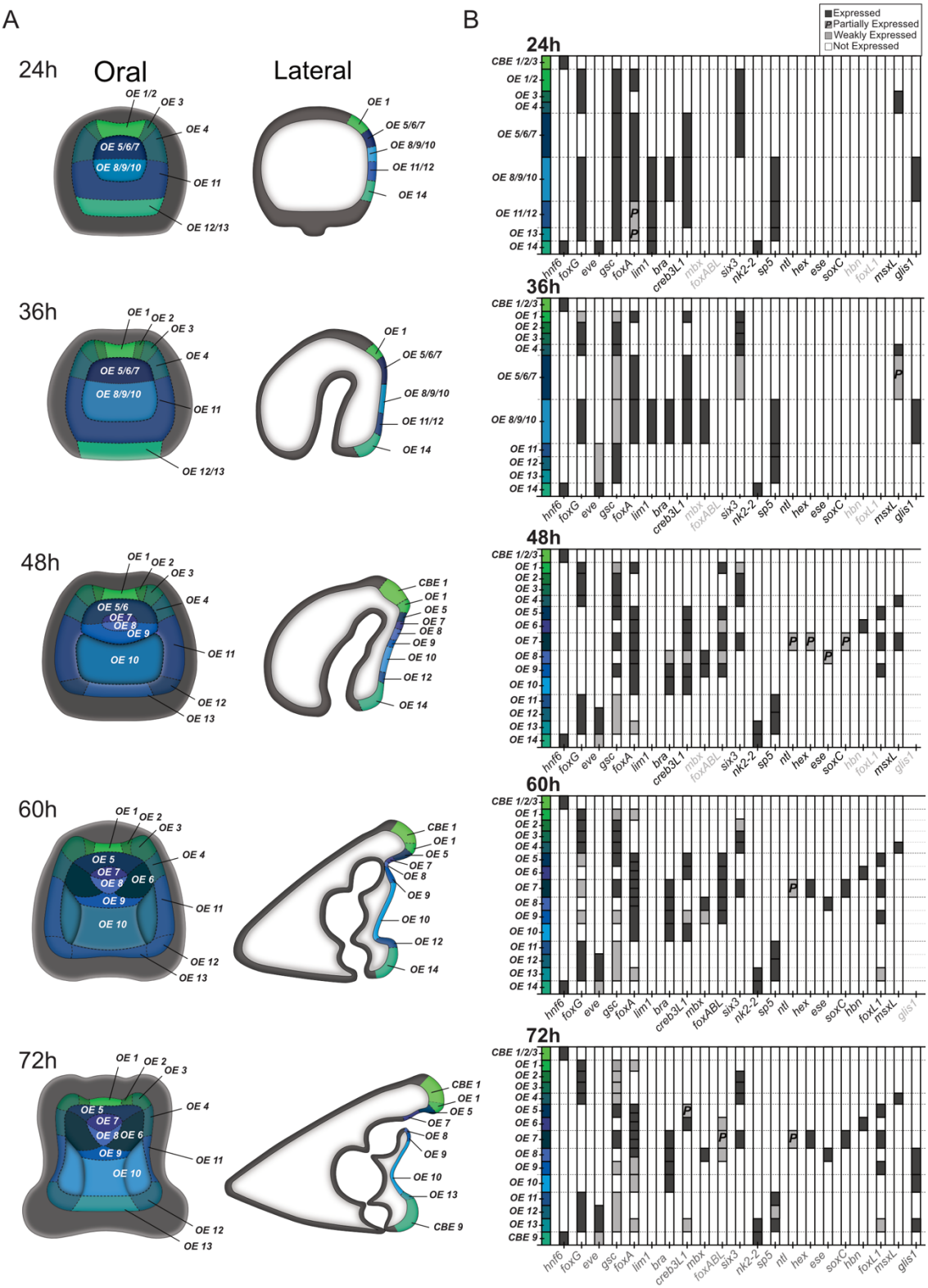


Figure A.3: Schematic of Oral Ectoderm Specification. (A) Embryonic staging schematics represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Embryos are drawn in oral and lateral views and are oriented with the oral ectoderm facing out and to the right respectively. Identified regulatory state domains within the oral ectoderm were are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of regulatory genes expressed within the oral ectodermal regulatory states at corresponding developmental time-points. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the oral ectoderm and represent the composition of regulatory genes responsible for their specification. As a whole, these schematics (A-B) demonstrate the specification of domains within the oral ectoderm along developmental time. The emergence of a novel domain in (A) can be attributed to the establishment of a new regulatory state in (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

Aboral Ectoderm

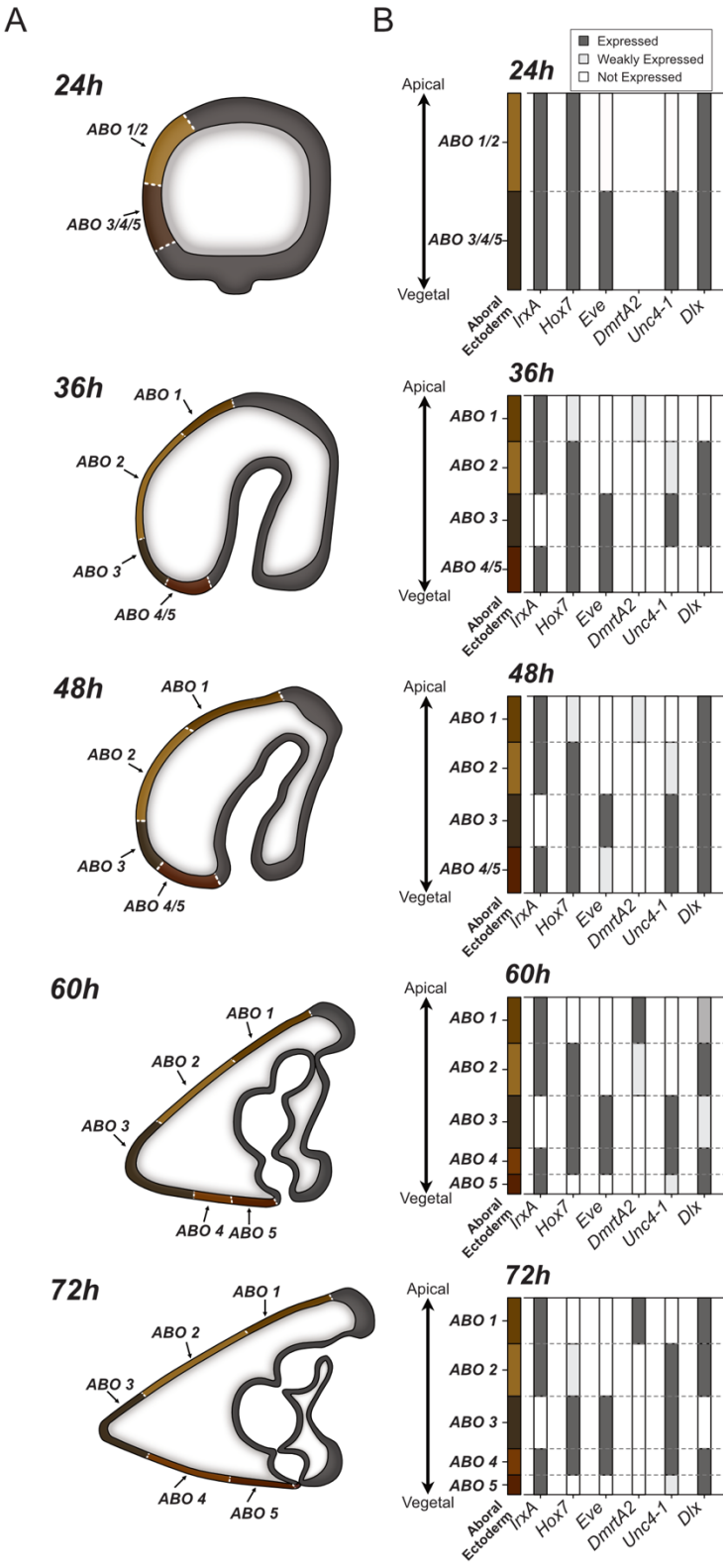


Figure A.4: Schematic of Aboral Ectoderm Specification. (A) Embryonic staging schematics represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Embryos are drawn in a lateral view and are oriented with the oral ectoderm facing right and the vegetal pole lying down. Identified regulatory state domains within the aboral ectoderm are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of regulatory genes expressed within the aboral ectodermal regulatory states at corresponding developmental time-points. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the aboral ectoderm and represent the composition of regulatory genes responsible for their specification. As a whole, these schematics (A-B) demonstrate the specification of domains within the aboral ectoderm along developmental time. The emergence of a novel domains in (A) can be attributed to the establishment of a new regulatory state (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

Skeletal Mesoderm

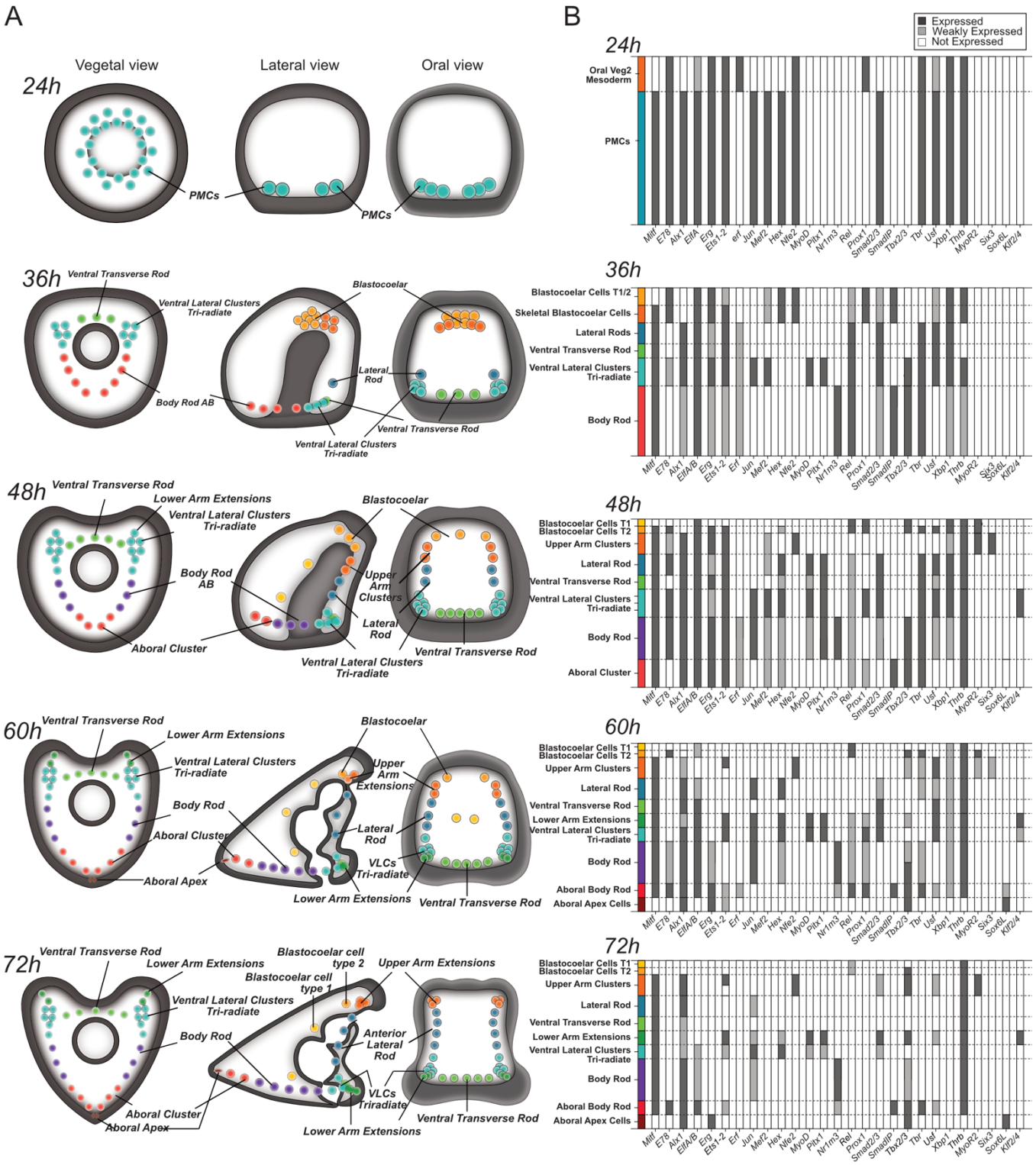
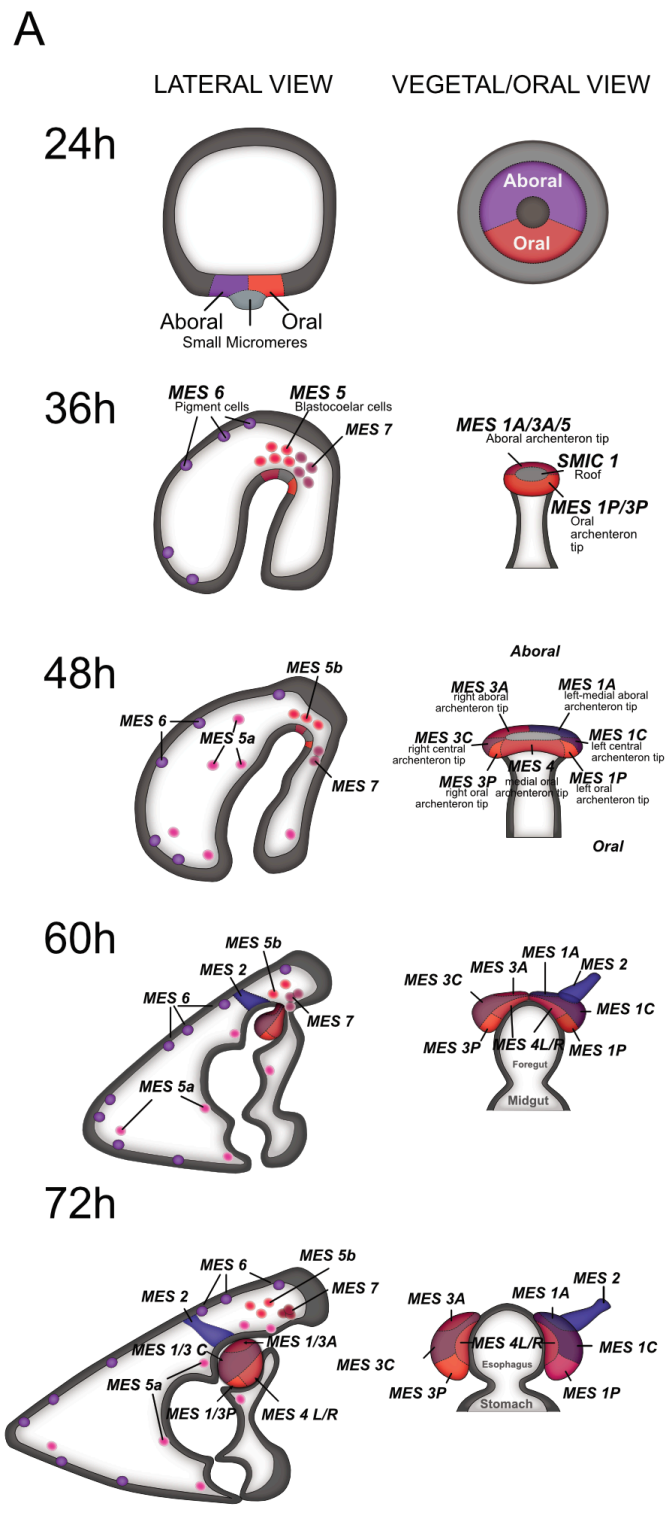


Figure A.5: Schematic of Skeletal Mesoderm Specification. (A) Embryonic staging schematics represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Embryos are drawn in a vegetal, lateral, and oral view and are oriented with the oral ectoderm facing up, right, and out respectively. Identified regulatory state domains within the skeletal mesoderm were determined as single cells and are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of regulatory genes expressed within the skeletal mesodermal regulatory states at corresponding developmental time-points. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the skeletal mesoderm and represent the composition of regulatory genes responsible for their specification. As a whole, these schematics (A-B) demonstrate the specification of domains within the skeletal mesoderm along developmental time. The emergence of a novel domain in (A) can be attributed to the establishment of a new regulatory state in (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

Mesoderm


A
LATERAL VIEW
VEGETAL/ORAL VIEW
24h
Aboral
Oral
Aboral
Oral
Small Micromeres
36h
MES 6
Pigment cells
MES 5
Blastocoelar cells
MES 7
MES 1A/3A/5
Aboral archenteron tip
SMIC 1
Roof
MES 1P/3P
Oral archenteron tip
48h
MES 5b
MES 6
MES 5a
MES 7
Aboral
MES 3A
right aboral archenteron tip
MES 1A
left-medial aboral archenteron tip
MES 3C
right central archenteron tip
MES 1C
left central archenteron tip
MES 4
MES 3P
right oral archenteron tip
medial oral archenteron tip
MES 1P
left oral archenteron tip
Oral
60h
MES 5b
MES 2
MES 6
MES 7
MES 5a
MES 1A
MES 3A
MES 2
MES 3C
MES 1C
MES 3P
MES 4L/R
MES 1P
Foregut
Midgut
72h
MES 6
MES 5b
MES 2
MES 7
MES 1/3 C
MES 1/3A
MES 5a
MES 1/3P
MES 4 L/R
MES 2
MES 3A
MES 1A
MES 4L/R
MES 1C
MES 3C
Esophagus
MES 3P
MES 1P
Stomach

Mesoderm

B

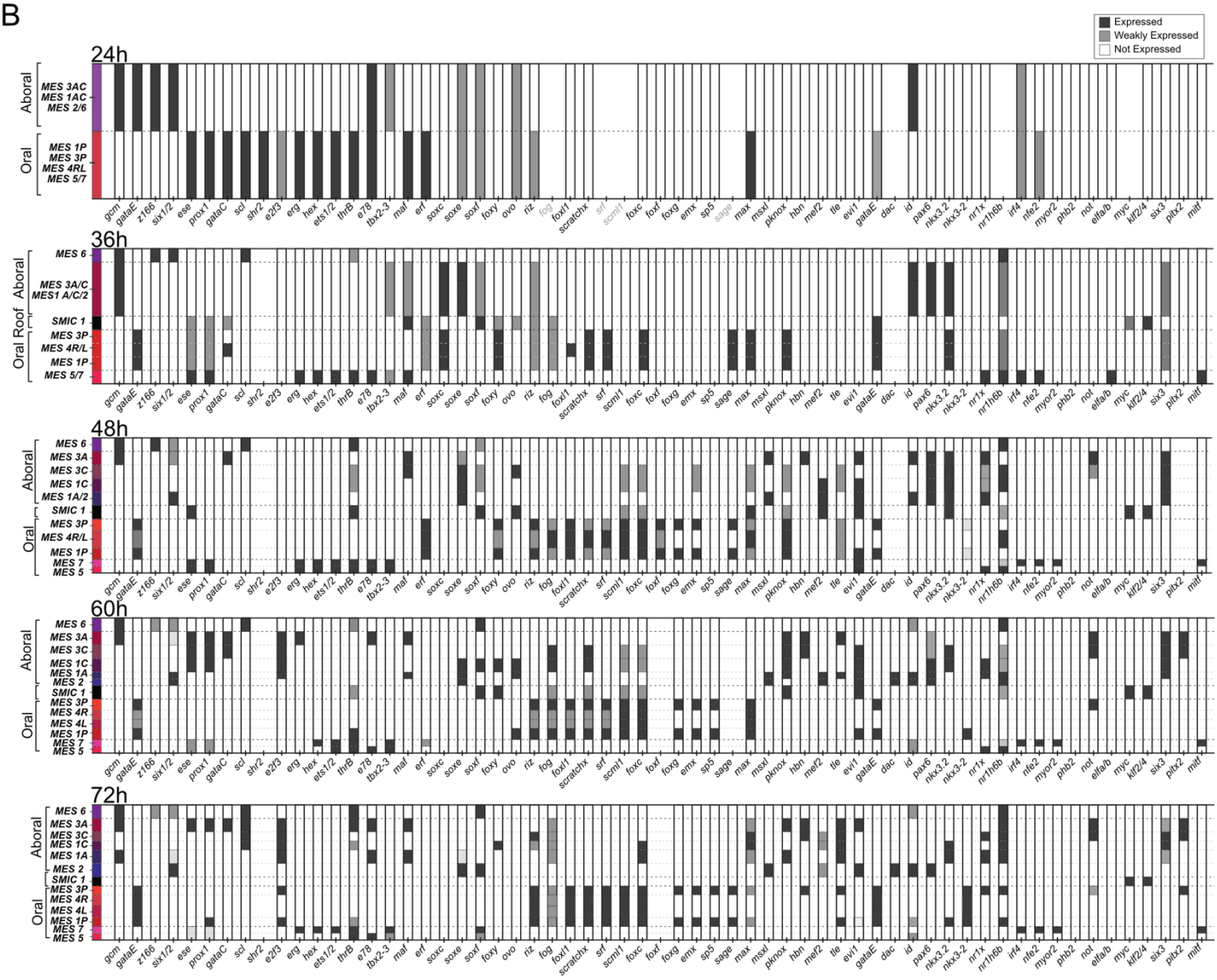


Figure A.6: Schematic of Mesoderm Specification. (A) Embryonic staging schematics represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Embryos are drawn in a lateral view and are oriented with the oral ectoderm facing right and the vegetal pole lying down. Archenteron or coelomic pouches are drawn in oral facing view and oriented with oral ectoderm facing out and vegetal pole lying down. Identified regulatory state domains within the mesoderm are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of regulatory genes expressed within the mesodermal regulatory states at corresponding developmental time-points. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range or distribution of

regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the mesoderm and represent the composition of regulatory genes responsible for their specification. As a whole, these schematics (A-B) demonstrate the specification of domains within the mesoderm along developmental time. The emergence of a novel domain in (A) can be attributed to the establishment of a new regulatory state in (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.

Endoderm

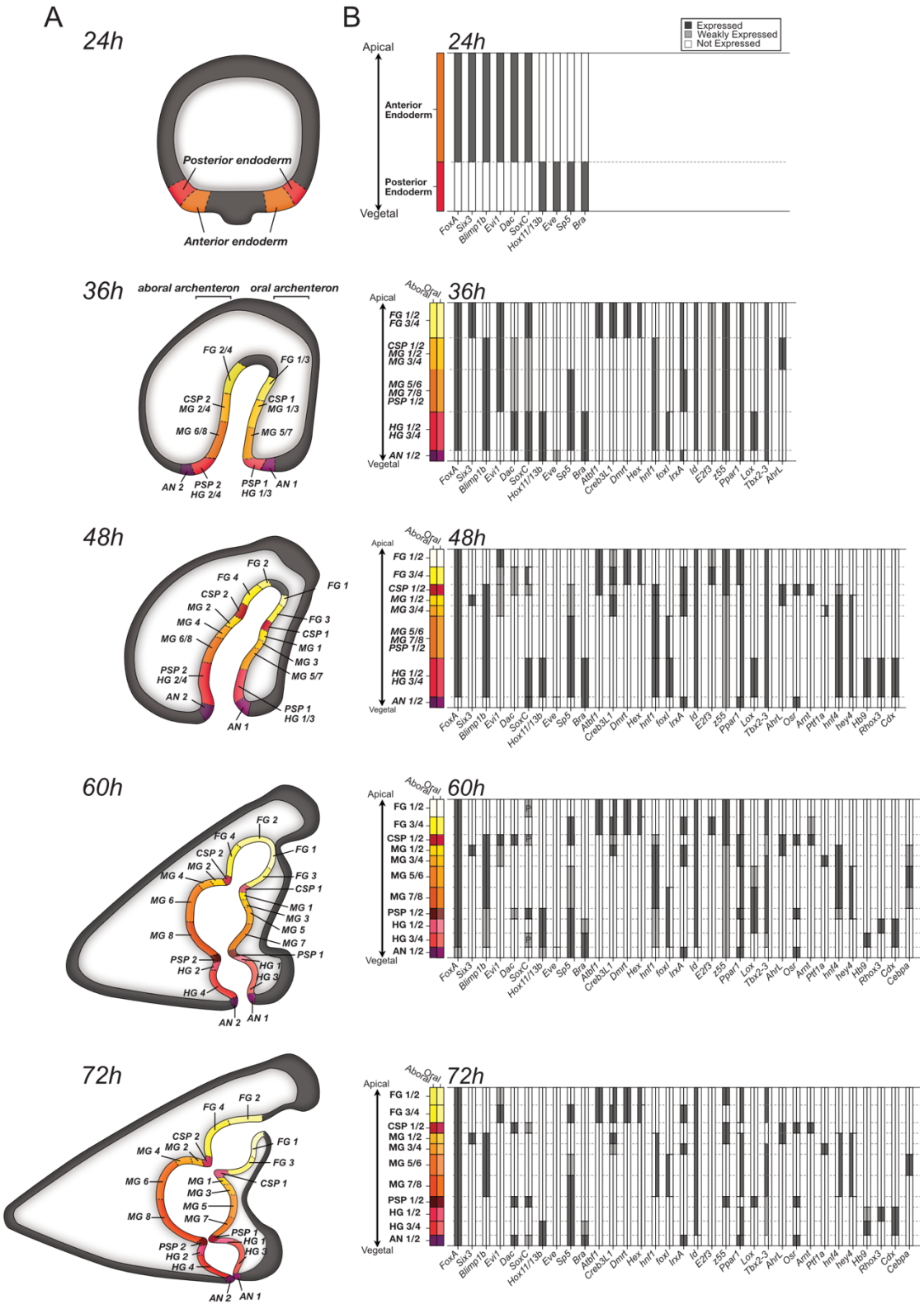

A
B
24h
Posterior endoderm
Anterior endoderm
Expressed
Weakly Expressed
Not Expressed
Apical
Vegetal
Anterior Endoderm
Posterior Endoderm
FoxA
Six3
Blimp1b
Evi1
Dac
SoxC
Hox11/13b
Eve
Sp5
Bra
36h
aboral archenteron
oral archenteron
FG 2/4
FG 1/3
CSP 2
MG 2/4
CSP 1
MG 1/3
MG 6/8
MG 5/7
AN 2
PSP 2
HG 2/4
PSP 1
HG 1/3
AN 1
Abo.
Oral
FG 1/2
FG 3/4
CSP 1/2
MG 1/2
MG 3/4
MG 5/6
MG 7/8
PSP 1/2
HG 1/2
HG 3/4
AN 1/2
Atbf1
Creb3L1
Dmrt
Hex
hnf1
foxl
Irx
Id
E2f3
z5f5
Ppar1
Lox
Tbx2-3
AhrL
48h
FG 2
FG 4
FG 1
CSP 2
MG 2
FG 3
MG 4
CSP 1
MG 6/8
MG 1
MG 3
PSP 2
HG 2/4
MG 5/7
AN 2
PSP 1
HG 1/3
AN 1
Osr
Arnt
Ptf1a
hnf4
hey4
Hb9
Rhox3
Cdx
60h
FG 2
FG 4
FG 1
CSP 2
MG 2
FG 3
MG 4
CSP 1
MG 6
MG 1
MG 3
MG 5
MG 8
MG 7
PSP 1
PSP 2
HG 2
HG 1
HG 3
HG 4
AN 2
AN 1
Cebpa
72h
FG 2
FG 4
FG 1
CSP 2
MG 2
FG 3
MG 4
MG 6
MG 1
CSP 1
MG 3
MG 5
MG 8
MG 7
PSP 1
PSP 2
HG 1
HG 2
HG 3
HG 4
AN 2
AN 1

Figure A.7: Schematic of Endoderm Specification. (A) Embryonic staging schematics on the left represent the developmental stages of mesenchyme blastula (24h), early-mid gastrula (36h), late gastrula (48h), prism (60h), and early pluteus larva (72h). Schematics are drawn in a lateral view and are oriented with the oral ectoderm facing right and the vegetal pole lying down. Identified regulatory state domains within the endoderm are highlighted in color. For emphasis and simplicity, all other identified domains from remaining territories are not shown unless otherwise noted to indicate orientation and/or boundary. (B) Graphic representation of expression tables of the regulatory genes expressed within the endodermal regulatory states at corresponding developmental time-point. This expression table includes only the regulatory states expressed in the corresponding regulatory state domains (e.g., identical color code and ID) of the adjacent embryonic schematics. The expression range of distribution of regulatory genes is viewed along the vertical axis and shown in columns. Regulatory states are shown in rows for each domain of the endoderm and represent the composition of regulatory genes responsible for their specification. For the developmental stages ranging between 36-72hpf, the endoderm is divided into oral and aboral sub-territories along the archenteron/gut. This subdivision is represented in the table as two adjacent columns for every gene, oral on the right and aboral on the left. Those regulatory states can be read in rows accordingly. As a whole, these schematics demonstrate the specification of domains within the endoderm along developmental time. The emergence of a novel domain in (A) can be attributed to the establishment of a new regulatory state in (B), via the combinatorial expression of regulatory genes, including new, different, and/or existing genes. The specification of all domains can be viewed for any developmental stage as well as their progression in the following developmental stage or their origin in the preceding development stage. For description on ID of regulatory state domains, refer to Table S3.2.