UNIVERSITY OF CALIFORNIA,
IRVINE

Combinatorial regulation by maternal transcription factors during activation of the endoderm gene regulatory network

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Kitt D. Paraiso

Dissertation Committee:
Professor Ken W.Y. Cho, Chair
Associate Professor Olivier Cinquin
Professor Thomas Schilling

2018

DEDICATION

To

the incredibly intelligent

and talented people,

who in one way or another,

helped complete this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATAC-seq	Assay for transposase-accessible chromatin combined with high-throughput sequencing
ChIP-qPCR	Chromatin immunoprecipitation combined with quantitative polymerase chain reaction
ChIP-seq	Chromatin immunoprecipitation combined with high-throughput sequencing
CRM	Cis-regulatory module
DNAse-seq	DNAse I digestion combined with high-throughput sequencing
FAIRE-seq	Formaldehyde-assisted identification of regulatory elements combined with high-throughput sequencing
GRN	Gene regulatory network
Hpf	Hours post-fertilization
MBT	Mid-blastula transition
MethyC-seq	Bisulfite conversion combined with high-throughput sequencing
MNAse-seq	Micrococcal nuclease digestion combined with high-throughput sequencing
MO	Morpholino
RNA-seq	RNA high-throughput sequencing
RT-qPCR	Reverse transcription combined with quantitative polymerase chain reaction
TF	Transcription factor
TPM	Transcripts per million
WT	Wild type
ZGA	Zygotic genome activation

ACKNOWLEDGMENTS

First and foremost, I would like to thank my mentor Dr. Ken W.Y. Cho, who has been supportive and encouraging of my research goals. He allowed me to pursue biological research in his lab, even without much biology background. Through all my setbacks and screw-ups (and there were plenty), he helped me push through and generate scientific works as best as I could. He provided so many opportunities for me to learn a new skill, to collaborate with experts in different fields, and to network with famous scientists in the field that I would not be the biologist that I am today without his guidance.

I thank members of my committee Dr. Thomas Schilling and Dr. Olivier Cinquin. Whether on addressing fundamental biological questions or some miniscule statistical detail, your comments and suggestions greatly improved the quality of my research projects.

Thank you to the Cho lab, both former and recent members. Particularly, I'd like to thank Ira Blitz for his work on almost every aspect of my project and for acting as a second mentor, especially when Ken is out traveling; Margaret Fish for her guidance in working with *Xenopus* embryos; and Rebekah Charney for her help with molecular biology and for the crash course on basic techniques. In addition, I would like to thank the entire lab for their intellectual input and support: William Chiu, Anna Javier, Mui Luong, Soledad Mochel, Elmira Forouzmand, Jin Cho, Sam Koo, Jessica Cheung, Jeff Zhou and Paula Pham. I would like to thank Bio 199 students Masani Coley and Junseok Yong who were exceptional undergraduate researchers and both of whom made significant contributions to my thesis.

I would like to thank the many friends that I have made during my graduate studies in UC Irvine. The ride has been mentally stimulating, interesting and a lot of fun. In particular, I would like to thank my MCB cohort: Shaun Hug, Marissa Macchietto, Jimmy Zhu, Mengfan Tang, Arjun Nair, Alyssa Kent, Seth Figueroa, Rabi Murad and Nancy Drew. I am glad to have met you. Thank you to Selma Alkafeef, Da-Wei Lin, Cristian Aguilar, Simei Yeh, Sarah Carmona, my lab mates whom I have become good friends, and many others. Whether it's surfing, traveling, snowboarding, rock climbing, going to festivals, or whatever activity, you've made the time during my graduate studies amazing and memorable.

Chapter 4 of the dissertation is a reprint of the material as it appears on *Seminars in cell & developmental biology*, vol. 66, pp. 12-24. Academic Press, 2017, DOI: doi.org/10.1016/j.semcdb.2017.03.003. The co-author Ken W.Y. Cho listed in this publication directed and supervised research which forms the basis for the thesis/dissertation. I would also like to acknowledge Rebekah M. Charney and Ira L. Blitz for their work on this manuscript.

CURRICULUM VITAE

Kitt D. Paraiso

EDUCATION

- University of California – Irvine; Irvine, CA — *Developmental and Cell Biology, Ph.D.* — 2018
- University of California – Irvine; Irvine, CA — *Developmental and Cell Biology, M.S.* — 2015
- University of California – Irvine; Irvine, CA — *Physics and Astronomy, M.S.* — 2012
- Rutgers University – Newark; Newark, NJ — *Physics, B.A.* — 2010

RESEARCH TALKS

- *Assembly of maternal transcription factors initiates the emergence of endoderm cis-regulatory regions.* 17th International Xenopus Conference. August 12-16, 2018. Seattle, WA, USA.
- *Multi-type co-regulation of Otx and T-box factors confer transcriptomic specificity.* Developmental and cell biology department retreat. April 13, 2018. Costa Mesa, CA, USA.
- *Ensemble of maternal transcription factors in regulating the endodermal transcriptome.* Tsukuba global science week 2016. September 17-20, 2016. Tsukuba, Ibaraki, Japan.

POSTER PRESENTATIONS

- *Topology of the Xenopus endoderm gene regulatory network.* Kitt D. Paraiso, Rebekah M. Charney, Ira L. Blitz, Ken W.Y. Cho. 2017 Society for developmental biology west coast regional meeting. March 28-31, 2017. Yosemite national park, CA, USA.
- *Control of endoderm gene expression by oocyte-localized transcription factors in Xenopus tropicalis.* Kitt D. Paraiso, Masani Coley, Ira L. Blitz, Margaret B. Fish, Ken W.Y. Cho. Society for developmental biology 75th annual meeting. August 4-7, 2016. Boston, MA, USA.
- *Differentially Expressed Maternal Factors Along the Animal-Vegetal Axis in Xenopus tropicalis.* Kitt Paraiso, Ira Blitz & Ken W.Y. Cho. 15th International Xenopus Conference, August 24-28, 2014. Asilomar, CA, USA.
- *A whole genome approach to explore the gene regulatory network controlling germ layer patterning in the Xenopus tropicalis gastrula*. Kitt Paraiso, William Chiu, Ira Blitz & Ken W.Y. Cho. 17th International Congress of Developmental Biology. June 16-20, 2013. Cancun, Mexico

PUBLICATIONS

- Charney, R.M.*, **Paraiso, K.D.*,** Blitz, I.L., Cho, K.W.Y. (2017). "A gene regulatory program controlling early Xenopus mesendoderm formation: network conservation and motifs." *Seminars in Cell & Developmental Biology*, 66, 12-24. doi.org/10.1016/j.semcdb.2017.03.003 *authorship equally shared
- Charney, R.M., Forouzmand, E., Cho, J.S., Cheung, J., **Paraiso, K.D.**, Yasuoka, Y., Takahashi, S., Taira, M., Blitz, I.L., Xie, X., Cho, K.W.Y. (2017). "Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program." *Developmental Cell,* 40 (6), 595-607. doi.org/10.1016/j.devcel.2017.02.017
- Blitz, I.L., **Paraiso, K.D.**, Patrushev, I., Chiu, W.T.Y., Cho, K.W.Y., Gilchrist, M.J. (2017). "A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo." *Developmental Biology*, 426(2), 409-417. doi.org/10.1016/j.ydbio.2016.07.002
- Forouzmand, E., Owens, N.D.L, Blitz, I.L., **Paraiso, K.D.**, Khokha, M.K., Gilchrist, M.J., Xie, X., Cho, K.W.Y. (2017). "Developmentally regulated long non-coding RNAs in Xenopus tropicalis." *Developmental Biology*, 462(2), 401-408. doi.org/10.1016/j.ydbio.2016.06.016

WORKSHOPS AND SHORT COURSES

- *Gene Regulatory Networks and Development.* October 12-24, 2014. Marine Biological Laboratory. Woods Hole, MA.
- *Computational Genomics Outreach Workshop.* July 29-30, 2013. Center for Cell Circuits, Broad Institute. Cambridge, MA.

FELLOWSHIPS AND TRAVEL AWARDS

- *UCI School of Biological Sciences Graduate Student Travel Award.* 2016.
- *UCI School of Biological Sciences Graduate Student Travel Award.* 2014.
- *NIH T32 HD 50555-5 Developmental Systems Biology Training Grant.* July 1, 2013 – June 30, 2014
- *UCI School of Biological Sciences Graduate Student Travel Award.* 2013.
- *NIH T32 HD 50555-4 Developmental Systems Biology Training Grant.* July 1, 2012 – June 30, 2013

TEACHING EXPERIENCE

Developmental and Cell Biology, UC Irvine, Irvine, CA

• DevBio 111L (Developmental Biology Lab)	Grader	Spring 2016
• DevBio 104 (Developmental Biology)	Discussion TA	Winter 2017/Winter 2016/Winter 2015
• DevBio 104 (Developmental Biology)	Grader	Spring 2014

Physics and Astronomy, UC Irvine, Irvine, CA

• Physics 7E (Waves and Oscillations)	Discussion TA	Summer 2011/Fall 2010
• Physics 3A (Classical Mechanics)	Discussion TA	Summer 2011
• Physics 3LB (Physics lab)	Lab TA	Spring 2011
• Physics 7D (Electromagnetism)	Discussion TA	Winter 2011
• Physics 7LD (Electromagnetism lab)	Lab TA	Winter 2011

MENTORING EXPERIENCE

- *Elizabeth Chamu*; Summer 2015; Minority Sciences Program, Bridges to Baccalaureate Program
- *Masani Coley*; Summer 2015 – Summer 2016; Bio 199 student, UROP funding recipient
- *Joe Giammalva*; Fall 2016 – Summer 2017; Bio 199 student
- *Junseok Yong*; Summer 2017 – Spring 2018; Bio 199 student, UROP funding recipient

OTHERS

- *Guest lecturer*. Bio D104 Developmental Biology. September 2017.
- *Speaker selection committee*. Department of developmental and cell biology. 2016-2017
- *Speaker selection committee*. Department of developmental and cell biology. 2015-2016
- *Judge. Intel International Science and Engineering Fair Qualifying Round*. Irvine, CA. March 14 & 21, 2015
- *Judge. Intel International Science and Engineering Fair Qualifying Round*. Irvine, CA. April 5, 2014
- *Judge. Intel International Science and Engineering Fair Qualifying Round*. Irvine, CA. March 16, 2013
- *Tutor*. National short course in systems biology. Irvine, CA. January 2012

PROFESSIONAL MEMBERSHIPS

- Society for Developmental Biology
- Genetics Society of America

ABSTRACT OF THE DISSERTATION

Combinatorial regulation by maternal transcription factors during activation of the endoderm gene regulatory network

By

Kitt D. Paraiso

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2018

Professor Ken W.Y. Cho

Zygotic gene activation (ZGA) is the process in the earliest stages of animal development when maternal transcription factors activate the expression of the first zygotic genes. In most metazoans, this process is coupled with specification of the three germ layers. In *Xenopus*, these two processes are initiated by the function of maternal transcription factors (TFs), which are localized in the egg and asymmetrically distributed to different cells of the embryo during cleavage stages. In this thesis, I uncover the combinatorial function of vegetally-localized TFs that are involved in the activation of the endodermal fate.

By gain- and loss-of-function approaches, I elucidate a novel function of the TF Otx1 in specification of the endoderm. Together with the well-known endodermal TF Vegt, Otx1 activates endodermal genes, while inhibiting mesodermal genes in the presumptive endoderm. By chromatin immunoprecipitation followed by qPCR or sequencing, I show that Vegt and Otx1 co-bind in the chromatin with the ubiquitously expressed Foxh1 prior to ZGA. These maternal TFs initiate the emergence of cis-regulatory regions by docking to

clusters of consensus DNA binding motifs. The binding of their assembly pre-marks regions of active cis-regulation, which are subsequently decorated with epigenetic enhancer marks and then bound by zygotically-active endodermal TFs.

In addition to co-regulation between maternal TFs, by curating evidence to generate the *Xenopus* endoderm gene regulatory network (GRN), I find evidence that maternal TFs co-regulate downstream target genes with zygotic TFs. Analysis of the *Xenopus* endodermal GRN suggests that feed-forward loops, where a maternal TF activates a primary zygotic gene encoding a TF, and combinatorially these TFs activate a secondary zygotic gene. From ChIP-seq analysis, the chromatin binding of Vegt and the zygotically active Smad2/3 support this finding. While the functional role of this network motif is unknown, computational modeling suggests that this motif could regulate the multiple waves of zygotic transcription during ZGA.

My findings synthesize the roles of maternal TFs in relation to each other, as well as to the roles of zygotic TFs, using cis-regulatory and gene regulatory network perspectives. The use of multiple approaches provides a broader picture of transcriptional regulation and cell specification during this critical process of ZGA.

Chapter 1

Maternal transcription factors during early endoderm formation in *Xenopus*

Transcription factors co-regulate target genes in a cell type-specific manner

Metazoan development from a pluripotent single cell to a functional organism composed of multiple interconnected organ systems involves a trajectory of incremental changes where each successive step depends on the steps prior. A battery of physical, chemical and genetic factors are required for proper completion of each step, among which are genes encoding transcription factors (TFs), which are a class of DNA-binding proteins that regulate the activation or repression of downstream genes. Gain- and loss-of-function approaches in various organisms and cell culture highlight the importance of TFs in development. In addition, TFs have also been shown to play a significant role in various disease pathogenesis. Specifically, mutations in functional regions of a TF can result in diseases. As TFs bind to the DNA to perform their function, mutations in TF binding sites have been implicated in controlling various diseases as well[1]. One such example is the case with the gain of a YY1 consensus binding site in the IL-10 regulatory region which results in down-regulation of IL-10 expression, and this regulatory interaction is linked to asthma disorders[2]. This is supported by genomic analysis, as genome-wide association studies (GWAS) have shown that majority of single nucleotide polymorphisms (SNPs) associated with various diseases reside not in the coding region of the DNA, but rather in intronic and intergenic regions[3]. The majority of these SNPs in non-coding regions were found to be DNAse I hypersensitive and contain a TF binding motif suggesting the role of gene regulation by TFs in these diseases. Therefore, the study of developmental and disease systems is intrinsically linked to the activity of TFs.

TFs do not work in isolation, as cell differentiation does not rely on the activity of a single TF. This is especially highlighted by the presence of the same signaling pathways whose downstream TFs are used within drastically different developmental contexts. For example, one branch of the TGFβ signaling pathway, Nodal signaling, is involved in maintenance of pluripotency in embryonic stem cells, regulates the activation of the endoderm from the early pluripotent state, signals from the endoderm to the overlying tissue to induce the mesoderm, and establish the left-right axis during organogenesis[4,5]. Similarly, recent analysis in single-cell transcriptomic datasets show that TFs are used multiple times in lineages of different tissues within the first 3 days of *Xenopus* development[6]. As individual TFs are used in multiple developmental contexts, this suggests that a single TF cannot specify a lineage and that the combinatorial action of TFs is essential for cell lineage specification.

Even further, not only is the combinatorial action of TFs critical, but coordinated action by TFs as they perform their function is also essential[7]. A survey of chromatin binding of TFs suggests that TF binding site overlap in the same regions of the genome in human cell culture, *Drosophila* and *C. elegans*[8-10]. In many cases, the overlap of TF binding is unclear as many TFs assayed are expressed in different cells or developmental time points. However, a finer analysis of binding sites in these systems along with binding sites in early *Xenopus* and zebrafish embryo formation where the TFs are co-expressed showed that these TFs co-bind in regulatory regions[11-14]. This is especially highlighted in the regulation of the *Xenopus* genes *gsc* and *hhex*[15,16]. In both cases, chromatin binding assays combined with gene reporter assays have shown the importance of coordinated action of TFs to specify

correct spatial expression of these genes. Overall, these findings suggest that the activity of TFs is cell type-specific, and this specificity is defined by the actions of the set of other transcription factors present in this cellular context.

Otx1 is expressed in a variety of cell lineages

One such TF that regulates developmental fates is the *Xenopus otx1*. The *Xenopus otx1* belongs to a subfamily of the paired-class homeobox TFs composed of the vertebrate *otx1*, *otx2* and *crx* (also known as *otx5*) and invertebrate *otx*/*otd*, whose homologs can be traced as far as ctenophores[17]. These genes have been implicated in anterior structure formation during embryogenesis across bilaterians. The founding member of this bilaterian homeobox sub-type is the *Drosophila otd*, which was identified as an important regulator of anterior structures[18]. Findings in vertebrates followed suit as the first cloned vertebrate OTX genes are expressed in overlapping patterns in the rostral brain of the mouse embryo, and loss-of-function mutations in *Otx1* and *Otx2* result in loss of telencephalic and diencephalic structures, as well as reduction in eye and inner ear structures[19-24]. The zebrafish and *Xenopus* Otx genes are expressed anteriorly in a similar fashion[25-27]. Functional assays through overexpression of *Xenopus otx1* or *otx2* results in preferential increase in anterior over posterior tissues[26,28]. In zebrafish, loss-of-function experiments using morpholino antisense oligonucleotides against *otx1* disrupts inner ear morphogenesis, in addition to defects caused in the embryonic brain[29,30]. Findings in *Drosophila*, mouse, *Xenopus* and zebrafish along with a variety of other organisms have

cemented the role of *otx* genes in anterior structures, particularly, the anterior neuroectoderm.

While the studies of the function of o*tx* genes have shown preferential focus on formation of anterior neuroectoderm structures, *otx* genes appear to have pleiotropic roles in development. In *Xenopus*, previously published single cell RNA-seq shows that *otx1* appears in multiple germ layer lineages, besides its anterior neuroectodermal expression[6]. *otx1* is expressed in the organizer in gastrula stages, possibly performing a similar role to *otx2* during neural induction[26]. In addition, *otx1* expression appears in other mesodermal derivatives during tailbud stages, which include the notochord, lateral plate mesoderm and definitive hemangioblast, where the *otx1* role is unknown. As in the mesoderm, the function of *otx1* in the endoderm in unknown. While the *Xenopus otx1* has been identified as an endodermally-localized maternal TF, the role of *otx1* in the early endoderm and how it functions within an endodermal context is unknown[31].

Maternal otx1 in endoderm development

The differentiation into germ layers is one of the earliest somatic differentiation events in metazoans. Among amniotes such as chick and mice, this decision process occurs after establishment of epiblast cells or pluripotent cells; while in externally developing embryos such as *Xenopus* and zebrafish, differentiation into one of the germ layers is the first somatic cell fate decision. In *Xenopus*, maternally expressed genes in the form of RNA and

protein are packaged and organized in the developing oocyte. This package includes cellular components critical for normal cell function, yolk that provides energy for the early developing embryo and maternally supplied antibodies in the case of a pathogen attack[32,33]. Most importantly, from a cell fate specification standpoint, eggs are packaged with RNAs encoding TFs and signaling molecules. These maternally expressed TFs and signaling molecules control the early decision processes of germ layer formation.

The localization of RNAs in the developing oocyte is regulated along the animal-vegetal axis. RNA binding proteins bind to specific RNAs and act with the cellular transport system, which involves kinesin motors and microtubules, and this process results in localization of these RNAs[32]. This process results in a pool of RNAs that are either ubiquitously expressed, animally-localized or vegetally-localized in the egg. After fertilization, during early cleavage stages, daughter cells from cell division asymmetrically inherit these animally- or vegetally-localized RNAs, which act as cytoplasmic determinants of cell fate. RNAs localized animally or vegetally specify the ectodermal or endodermal cell fates, respectively. In the vegetal blastomeres, the combination of ubiquitously expressed and vegetally-localized factors define the endodermal context in which maternally-expressed *otx1* operates.

Among the vegetally-localized RNAs that encodes a TF is *vegt*[34-37]. This T-box TF sits at the top of the mesendodermal gene regulatory hierarchy[15,38,39]. Gain- and loss-of-function experiments validated the role of *Vegt* in activating the expression of a battery of mesendodermal zygotic TFs including *mixer*, *mix1* , *gata4/5/6* and *sox17*, as well as the

endoderm marker *a2m/endodermin*[40-42]. Additionally, *vegt* activates the expression of nodal signaling ligands including *nodal1*, *nodal5 and nodal6*, whose functions are critical in the formation of the endoderm and the induction of the mesoderm[43-45]. Through a combination of chromatin binding, reporter gene analysis and protein synthesis inhibition experiments, *vegt* has been shown to directly regulate the expression of a subset these genes[46-48]. Due to its importance in zygotic mesendodermal gene expression, large scale approaches have been employed to identify *vegt* targets through microarray analysis[49]. These findings, in combination with similar findings in other vertebrates, suggest that T-box TFs play a prominent role in specification of the endoderm[13,50-52].

Besides *vegt*, other vegetally-localized factors specify the endodermal lineage including *sox7*, *wnt11* and *gdf1*. While previously thought of as a zygotically expressed factor acting as a downstream target of maternal *vegt*, recent RNA-seq and *in situ* hybridization experiments confirmed the maternal expression and vegetal localization of the F-type Sox TF *sox7*[53-55]. Regardless, the function of *sox7* was characterized in the endoderm, and *sox7* activates the expression of a similar set of genes as *vegt* including *a2m*, *mixer*, *nodal5* and *nodal6*, although the function of sox7 is not as well characterized[56]. *Wnt11* encodes a Wnt signaling ligand and is one of the earliest identified vegetally-localized RNAs in the *Xenopus* egg[57]. Signaling by Wnt11 (through the signal transducer Ctnnb1/β-catenin) is required for activation of Spemann organizer genes such as *gsc* and *sia1* in the dorsal mesendodermal region of the embryo[58]. *Gdf1* encodes a TGFβ ligand that activates the Activin/Nodal signaling pathway and is another vegetally-localized factor[59]. Maternally expressed *gdf1* is required for active Nodal signaling through

phosphorylation of Smad2/3, which results in the expression of dorsal mesendodermal genes such as *cer1* and *chrd*[60]. Vegt, Sox7, Ctnnb1/β-catenin (via vegetal *wnt11*) and Smad2/3 (via vegetal *gdf1*) are the set of critical vegetally-active TFs.

In addition to these factors, ubiquitously expressed factors appear to be important for mesendodermal gene regulation. Such is the case with *foxh1*. *foxh1* encodes a forkhead TF and has been shown to co-regulate with vegetally-active TF. Foxh1 was initially identified as a co-factor in the Nodal signaling pathway during the regulation of mesendodermal genes in development through direct interaction with Smad2/3[61,62]. However, recent functional evidence suggests that *foxh1* possesses Nodal-independent function during germ layer formation[12,63]. Our own findings suggest that Foxh1 together with the vegetally-localized Sox7 co-regulate target genes, as Foxh1 ChIP-seq binding sites are enriched for Sox motifs and these regions are identified to be co-bound by Sox7 by ChIP-qPCR[12]. In addition, findings in zebrafish suggest that in addition to co-regulation with Nodal signaling, Foxh1 co-regulates target genes together with the endodermal T-box TF Eomes, which functions similarly as *Xenopus* Vegt. Zebrafish Foxh1 and Eomes combinatorially regulate the expression of zygotic mesendodermal factors *bon*, *gsc* and *sox32*[64,65]. Although the interaction between *Xenopus* Foxh1 and Vegt has not been previously established, Foxh1 and Vegt regulate a similar set of target genes including mesendodermal genes *gsc*, *cer1* and *gdf3*, suggesting a similar interaction, as seen in zebrafish[66]. These findings suggest that the interplay between the ubiquitously expressed Foxh1 and vegetally-active TFs Smad2/3, Sox7 and Vegt is important during endoderm formation.

Through a screen of vegetally-localized factors, I have identified a handful of maternally-expressed putative endodermal regulators, one of which is *otx1* (Fig 2.1). As mentioned earlier, while *otx1* is expressed is a variety of cell lineages, its role in the endoderm is unknown. Chapter 2 addresses the novel role of the maternal and vegetally-expressed *otx1* pursuing the hypothesis that Otx1 is an important regulator of endoderm formation and collaborates with known endodermal transcription factors Foxh1 and Vegt. The chapter addresses how Otx1 functions within the endodermal context as it relates to binding to the chromatin along with the other maternal TFs, and functionally, as to how the co-regulation of target genes by maternal TFs results in the endoderm-specific transcriptome.

Establishment of enhancer activity by maternal TFs

The activity of cis-regulatory regions can be affected by chromatin states such as the chromatin density (euchromatin or heterochromatin) or epigenetic modifications (histone marks and DNA methylation) in addition to TF activity. While the role of TFs in gene regulation has been known for a long time, recent genomic advances have enabled the critical assessment of open chromatin and epigenetic modifications, and their role in gene regulation. Techniques such as DNAse-seq, FAIRE-seq, ATAC-seq and histone H3 ChIP-seq have enabled the genome-wide identification of open chromatin regions; while ChIP-seq on numerous histone marks, MethylC-seq and bisulfite sequencing has enabled the genome-wide study of DNA and histone modifications.

Analyses of these datasets show a striking correlation of the localization in the genome of open chromatin, histone modifications (particularly, enhancer marks), and TF binding[67]. These findings suggest an interplay between these three variables during establishment of cis-regulatory regions. However, what is unclear is which of these variables initiate the activity of cis-regulatory regions. Unlike nucleosomes or histone modifiers, TFs contain DNA binding domains that can provide the DNA sequence specificity in localization of cis-regulatory regions. Thus, these DNA binding domains provide a convenient hypothesis for establishment of active cis-regulatory regions. Multiple models exist as to how TFs associate with the genome[7]. First, TFs could bind to the DNA through collaborative binding. In this model, TFs could dislodge the nucleosomes to establish enhancer regions. Alternatively the action of a single TF, so called pioneer factors, can provide the impetus in establishment of cis-regulatory regions. In this model, the actions of these pioneer factors result in chromatin opening, establishment of enhancer marks, and subsequent binding of other TFs.

During zygotic gene activation, when maternal TFs regulate the embryonic genome, support for both models exists. In the case of *Drosophila*, the dominant model is through the activities of the pioneer factor Zelda. Evidence from DNAse I hypersensitivity and MNAse digestion assays indicate that this non-lineage specific TFs outcompetes nucleosomes for access to DNA resulting in increase in chromatin accessibility[68,69]. This function enables binding of lineage-specific factors involved in anterior-posterior and dorsal-ventral patterning[68-71]. In mammals, the pioneering role has been relegated to the mammalian-specific maternally-expressed TFs, DUX (mouse) and DUX4 (human). DUX TFs

are critical regulators of genes and retrotransposon expression during zygotic gene activation[72-74]. Although *in vivo* experiments are lacking, evidence from 2-cell stage-like induction from mouse and human embryonic stem cells suggests that DUX TFs are important for establishment of open chromatin in the 2-cell stage embryos[74]. In zebrafish, rather than through the function of pioneer factors, it appears that coordinated binding is the preferred mechanism. The genes orthologous to the mammalian pluripotency factors *NANOG*, *OCT4* and *SOX2* have been implicated in the control of the early zygotic genome[75]. The factors Pou5f1 (*OCT4* homolog) and Soxb1 (*SOX2* homolog) pre-mark regions of the genome that are later bound by RNA polymerase II and decorated with the promoter mark H3K4me3[76]. Like Zelda, Pou5f1 has been shown to outcompete nucleosomes for access to the DNA[77]. While binding mechanisms have not been fully elucidated, the preference between coordinated binding or pioneer factor binding could depend on the strength of association of each TF to its consensus binding motif. Strong association could enable independent binding as in the case of pioneer factors, while weaker associations would require coordinated binding by multiple TFs.

In *Xenopus*, the first signs of cis-regulatory region activity appear to be reliant on TFs as well, rather than epigenetic marking. Previously, we showed that the maternally expressed Foxh1 could bind to the chromatin as early as the 32-cell stage[12]. This binding occurs 2-3 cell cycles prior to the establishment of enhancer marks such as H3K4me1 and Ep300, as well as association of RNA polymerase II in nearby promoters of genes. While Foxh1 could act as a pioneer factor, the chromatin binding of other maternal TFs has not been established prior to zygotic gene activation. Possibly, other maternal TFs such as Vegt and

Otx1 can co-bind with Foxh1 through collaborative mechanisms. Chapter 2 addresses the question of how lineage-specific enhancers emerge during zygotic gene activation, particularly in the endoderm. Using analysis of ChIP-seq datasets for maternal TFs and histone modification, the chapter addresses the role of coordinated binding by maternal factors in pre-marking these genomic regions as cis-regulatory regions. In addition, Chapter 3 addresses the question of how maternal TFs discriminate regions of the chromatin to be pre-marked as enhancers.

Uncovering the endodermal gene regulatory network

Combinatorially, TFs operate together in cis-regulatory regions to modulate the expression of their target genes. An alternative perspective to view combinatorial activity of TFs is at the level of networks. TFs form an interconnected network of gene regulation whose design has crucial implications in the differentiation process that cells must undergo to follow their correct lineages. These gene regulatory networks (GRNs) are helpful in visualizing and understanding the function of TFs as they relate to the network itself. Two examples can help demonstrate the utility of GRNs: boundary sharpening between the dorsal and ventral regions of the *Xenopus* mesoderm by by *gsc* and and *tbxt*, and maintenance of pluripotency in mammals by *SOX2* and *OCT4*.

The first example illustrates a small GRN that enables formation of distinct lineages of the dorsal and ventral regions of the *Xenopus* mesoderm. The TFs *tbxt* and *gsc* are

mesodermally expressed genes, where *tbxt* expression is broad and *gsc* expression is dorsally-localized[78,79]. These genes form an indirect negative feedback loop whereby Gsc directly represses *tbxt* expression, while Tbxt can induce a TF (possibly, *ventx2*) that represses *gsc* expression[80-82]. This small cross inhibitory network results in sharpening of the boundary differentiating the ventral (*tbxt*) and dorsal organizer (*gsc*) mesodermal regions, which then generates distinct mesodermal derivatives[82].

The second example illustrates a small GRN in embryonic stem cells that enables maintenance of pluripotency in mammals. This process involves the intricate balance generated by the TFs SOX2 and OCT4 to inhibit germ layer differentiation. SOX2 and OCT4 form a positive feedback loop to maintain each other's expression[83]. Meanwhile, SOX2 is capable of inducing ectodermal genes, however, this is inhibited by OCT4[84,85]. Similarly, OCT4 can activate the expression of mesendodermal genes, but this is inhibited by SOX2. The combination of the positive feedback regulation between OCT4 and SOX2 coupled with their ability to either activate or inhibit specific germ layer lineages enable maintenance of the pluripotent state of these stem cells. Interestingly, destabilization of this balance provides a mechanism to exit from pluripotency and to specify either the mesendodermal or the ectodermal cell fates.

While these smaller GRNs are useful, well studied developmental networks, particularly in the early development of *Xenopus*, sea urchin, and *Drosophila*, show that dozens of key TFs are involved during cell differentiation[15,38,86,87]. In *Xenopus*, previous efforts have

uncovered dozens of players during the first stages of development from fertilization through the early stages of gastrulation. However, recent genomic experiments generated much larger datasets, which are useful in generating these networks. RNA-seq datasets helped identify genome-wide regulatory targets of each TF while ChIP-seq provided evidence for directness of regulatory interactions. These datasets can provide the full expanse of regulatory targets of key factors such as the dorsal specifying genes Ctnnb1/β-catenin and Gsc, as wells as mesendodermal regulators Foxh1 and Smad2/3[63,88-90]. Through literature search and analysis of these genomic datasets, we have updated the *Xenopus* mesendoderm GRN. While we focused on a similar set of genes, we have dramatically expanded the network connectivity. In addition, I discuss systems biological approaches to shed light on the functionality of the network through network motif analysis. Chapter 4 reports on the establishment of the most up-to-date *Xenopus* mesendoderm GRN. It focuses on understanding early *Xenopus* development from a systems biological point of view.

Zygotic genome activation and temporal control of gene expression

Zygotic genome activation (ZGA) is a critical developmental process in which maternal genetic control of development is transferred to zygotic genes. The exact timing of this embryonic milestone differs across species. Amniotes such as humans and mice undergo ZGA at or after 24 hours post-fertilization (hpf), within the first few cell cycles. In contrast, externally developing organisms operate on much faster developmental time scales, possibly due to greater environmental pressure. *Xenopus tropicalis*, zebrafish and

Drosophila ZGA occur 4.5 hpf, 4.3 hpf and 2.5 hpf, respectively. How these time-scales are regulated in a species-specific manner is unclear, although various models exist to explain ZGA timing in fast developing embryos.

Multiple models exist to explain the timing of the onset of ZGA, and many molecules have been implicated in this control. Among these molecules are (a) TFs, (b) nucleosomal histones and (c) cell-cycle regulators. In the case of TFs, while RNA is maternally deposited, protein products of these genes might not be present in abundance as to activate the expression of zygotic genes. In *Xenopus*, this is supported by RNA-seq experiments showing maternal RNA is generally not poly-adenylated[54,91]. In addition, the protein expression of TFs such as Vegt and Foxh1, and the TATA-box binding protein Tbp are relatively low compared to protein levels around the time of ZGA[12,92,93]. In the case of nucleosomal histones, the models suggest that nucleosomal histones act as repressors of transcription. Nucleosomal histones compete with TF binding to DNA, which effectively prevents activation of the TF target genes. In *Xenopus* and zebrafish, titration experiments through increase or decrease in amount of nucleosomal histones H3 and H4 show that the amount of histones negatively correlates with the onset of ZGA[77,94]. Lastly, in models implicating cell cycle regulators, replication factors act as limiting factors, which result in the timing of cell cycles during cleavage stages[95]. Although it is not clear if these experiments affected ZGA, the change in timing of ZGA likely occurred earlier as well. Overall, a variety of cellular components are involved the timing of the onset of ZGA, including nucleosomal histones, TFs and cell cycle regulators.

Recent observations in gene expression during early development show that ZGA occurs in multiple waves[96,97]. Particularly, prior to the timing of the onset of ZGA, some genes are transcribed representing an early wave of zygotic transcription. In *Xenopus*, early waves of zygotic transcription have been identified in critical endodermal genes such as *nodal5* and *nodal6* prior to the whole genome activation[98]. These findings have been supported by more recent transcriptomic datasets showing multiple genes are transcribed prior to whole genome activation[54,91]. While multiple models exist to explain the control of the onset of ZGA, these models are incomplete in that they cannot explain the multiple waves of gene activation. This suggests that there must be additional levels of control to specify the different waves.

Our systems biological analysis seems to suggest a hypothesis. During our analysis of the *Xenopus* mesendodermal GRN, we observed that the early germ layer network shows an abundance of regulatory motifs[66]. Specifically, we find that feedforward loops which involve triple activation (gene A activates gene B; and genes A and B regulate gene C) are particularly abundant compared to other motifs such as positive feedback and auto-regulation. In our network, the feedforward loops consist of a maternal TF, which activates a gene encoding primary zygotic TF. Together the maternal TF and the primary zygotic TF co-activate a secondary zygotic gene. Possibly, if both maternal and primary zygotic TFs were required for activation of the secondary zygotic gene, then this would explain the delayed timing of secondary zygotic gene expression. Genomic datasets corroborate these systems biological findings. The maternal factor Foxh1 binds to the chromatin during the early and late waves of zygotic transcription[12]. In addition, during the later waves of

zygotic transcription, Foxh1 co-binds with Smad2/3 and these transcription factors co-regulate dorsal mesendodermal genes. This suggests that feedforward loops could provide a second tier control of ZGA onset to generate multiple waves of transcription.

Together, the spatial and temporal controls of gene expression are critical in cell fate specification and differentiation. However, the study of gene regulation has placed a particular attention to the spatial aspect of expression. For example, in the gastrula stage embryo, many genes have been identified as endodermal (*sox17*[99], *mix*[100], *mixer*[101]), ventral (*ventx1*[102], *ventx2*[103]) organizer (*gsc*[78]), and ectodermal (*foxi1*[104]) markers. In addition, the control of their expression localization has been elucidated as described by GRNs[15,38,66]. However, similar analogous markers have not been identified in describing various developmental stages besides the early and late wave genes. Even further, how the expression timing of these genes is generated is unclear. Chapter 4 reports on the findings of feedforward loops as it relates to the *Xenopus* mesendodermal network. Chapter 5 elaborates on the second tier of control of the onset of ZGA and pursues the hypothesis that feedforward loops are important components of this temporal regulatory control. Particularly, the chapter focuses on the function of maternal Vegt and zygotic Nodal signaling.

The role of maternal transcription factors in early development

For my thesis, I focused on the function of maternal TFs during the process of ZGA. I used molecular biology, genomic and systems biological approaches to understand the function of maternal factors. In Chapter 2, I talk about my findings on the novel role of Otx1 during early endoderm development. Otx1 initiates the emergence of endodermal enhancers, and the activation of the endodermal gene regulatory network in collaboration with Vegt and Foxh1. Chapter 3 addresses the question of how maternal TFs discriminate specific regions in the genome for docking, and report on key features of endodermal cis-regulatory regions. Chapter 4 steps back and broadly looks at the role of maternal TFs, within the context of the GRN from egg through gastrula stages. This chapter introduces the function of network motifs containing maternal transcription factors during endoderm differentiation. Chapter 5 elaborates on the function of network motifs and their possible function in regulating the timing of gene expression. I use early *Xenopus* embryos to generate these findings, although the basic mechanisms uncovered in this work related to gene regulatory networks, network motifs, spatial and temporal control of gene expression and epigenetics could be applicable to other developmental systems.

References

1.Deplancke, B., Alpern, D. & Gardeux, V. The Genetics of Transcription Factor DNA Binding Variation. *Cell* **166,** 538–554 (2016).

2.Hobbs, K., Negri, J., Klinnert, M., Rosenwasser, L. J. & Borish, L. Interleukin-10 and Transforming Growth Factor- β Promoter Polymorphisms in Allergies and Asthma. *American Journal of Respiratory and Critical Care Medicine* **158,** 1958–1962 (1998).

3.Maurano, M. T. *et al.*. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337,** 1190–1195 (2012).

4.Schier, A. F. Nodal Signaling in Vertebrate Development. *Annual Review of Cell and Developmental Biology* **19,** 589–621 (2003).

5.James, D., Levine, A. J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells.. *Development* **132,** 1273–82 (2005).

6.Briggs, J. A. *et al.*. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution.. *Science* **360,** (2018).

7.Reiter, F., Wienerroither, S. & Stark, A. Combinatorial function of transcription factors and cofactors.. *Curr Opin Genet Dev* **43,** 73–81 (2017).

8.The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

9.Gerstein, M. B. *et al.*. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project.. *Science* **330,** 1775–87 (2010).

10.Roy, S. *et al.*. Identification of functional elements and regulatory circuits by Drosophila modENCODE.. *Science* **330,** 1787–97 (2010).

11.Yasuoka, Y. *et al.*. Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification. *Nature Communications* **5,** (2014).

12.Charney, R. M. *et al.*. Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program.. *Dev Cell* **40,** 595–607.e4 (2017).

13.Nelson, A. C. *et al.*. In Vivo Regulation of the Zebrafish Endoderm Progenitor Niche by T-Box Transcription Factors.. *Cell Rep* **19,** 2782–2795 (2017).

14.Gentsch, G. E., Monteiro, R. S. & Smith, J. C. Cooperation Between T-Box Factors Regulates the Continuous Segregation of Germ Layers During Vertebrate Embryogenesis.. *Curr Top Dev Biol* **122,** 117–159 (2017).

15.Koide, T., Hayata, T. & Cho, K. W. Xenopus as a model system to study transcriptional regulatory networks.. *Proc Natl Acad Sci U S A* **102,** 4943–8 (2005).

16.Rankin, S. A., Kormish, J., Kofron, M., Jegga, A. & Zorn, A. M. A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer.. *Dev Biol* **351,** 297–310 (2011).

17.Moroz, L. L. *et al.*. The ctenophore genome and the evolutionary origins of neural systems.. *Nature* **510,** 109–14 (2014).

18.Finkelstein, R., Smouse, D., Capaci, T. M., Spradling, A. C. & Perrimon, N. The orthodenticle gene encodes a novel homeo domain protein involved in the development of the Drosophila nervous system and ocellar visual structures.. *Genes Dev* **4,** 1516–27 (1990).

19.Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. & Boncinelli, E. Nested expression domains of four homeobox genes in developing rostral brain.. *Nature* **358,** 687–90 (1992).

20.Matsuo, I., Kuratani, S., Kimura, C., Takeda, N. & Aizawa, S. Mouse Otx2 functions in the formation and patterning of rostral head.. *Genes Dev* **9,** 2646–58 (1995).

21.Acampora, D. *et al.*. Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation.. *Development* **121,** 3279–90 (1995).

22.Ang, S. L. *et al.*. A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain.. *Development* **122,** 243–52 (1996).

23.Suda, Y., Matsuo, I., Kuratani, S. & Aizawa, S. Otx1 function overlaps with Otx2 in development of mouse forebrain and midbrain.. *Genes Cells* **1,** 1031–44 (1996).

24.Acampora, D. *et al.*. Epilepsy and brain abnormalities in mice lacking the Otx1 gene.. *Nat Genet* **14,** 218–22 (1996).

25.Mercier, P., Simeone, A., Cotelli, F. & Boncinelli, E. Expression pattern of two otx genes suggests a role in specifying anterior body structures in zebrafish.. *Int J Dev Biol* **39,** 559–73 (1995).

26.Blitz, I. L. & Cho, K. W. Anterior neurectoderm is progressively induced during gastrulation: the role of the Xenopus homeobox gene orthodenticle.. *Development* **121,** 993–1004 (1995).

27.Kablar, B. *et al.*. Xotx genes in the developing brain of Xenopus laevis.. *Mech Dev* **55,** 145–58 (1996).

28.Andreazzoli, M., Pannese, M. & Boncinelli, E. Activating and repressing signals in head development: the role of Xotx1 and Xotx2.. *Development* **124,** 1733–43 (1997).

29.Hammond, K. L. & Whitfield, T. T. The developing lamprey ear closely resembles the zebrafish otic vesicle: otx1 expression can account for all major patterning differences.. *Development* **133,** 1347–57 (2006).

30.Foucher, I. *et al.*. Differentiation of cerebellar cell identities in absence of Fgf signalling in zebrafish Otx morphants.. *Development* **133,** 1891–900 (2006).

31.Pannese, M., Cagliani, R., Pardini, C. L. & Boncinelli, E. Xotx1 maternal transcripts are vegetally localized in Xenopus laevis oocytes.. *Mech Dev* **90,** 111–4 (2000).

32.King, M. L., Messitt, T. J. & Mowry, K. L. Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. *Biology of the Cell* **97,** 19–33 (2005).

33.Poorten, T. J. & Kuhn, R. E. Maternal transfer of antibodies to eggs in Xenopus laevis.. *Dev Comp Immunol* **33,** 171–5 (2009).

34.Stennard, F., Carnac, G. & Gurdon, J. B. The Xenopus T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation.. *Development* **122,** 4179–88 (1996).

35.Zhang, J. & King, M. L. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning.. *Development* **122,** 4119–29 (1996).

36.Lustig, K. D., Kroll, K. L., Sun, E. E. & Kirschner, M. W. Expression cloning of a Xenopus T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation.. *Development* **122,** 4001–12 (1996).

37.Horb, M. E. & Thomsen, G. H. A vegetally localized T-box transcription factor in Xenopus eggs specifies mesoderm and endoderm and is essential for embryonic mesoderm formation.. *Development* **124,** 1689–98 (1997).

38.Loose, M. & Patient, R. A genetic regulatory network for Xenopus mesendoderm formation.. *Dev Biol* **271,** 467–78 (2004).

39.Heasman, J. Patterning the early Xenopus embryo.. *Development* **133,** 1205–17 (2006).

40.Zhang, J. *et al.*. The role of maternal VegT in establishing the primary germ layers in Xenopus embryos.. *Cell* **94,** 515–24 (1998).

41.Clements, D., Friday, R. V. & Woodland, H. R. Mode of action of VegT in mesoderm and endoderm formation.. *Development* **126,** 4903–11 (1999).

42.Xanthos, J. B., Kofron, M., Wylie, C. & Heasman, J. Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis.. *Development* **128,** 167–80 (2001).

43.Takahashi, S. *et al.*. Two novel nodal-related genes initiate early inductive events in Xenopus Nieuwkoop center.. *Development* **127,** 5319–29 (2000).

44.Houston, D. W. *et al.*. Repression of organizer genes in dorsal and ventral Xenopus cells mediated by maternal XTcf3.. *Development* **129,** 4015–25 (2002).

45.Houston, D. W. & Wylie, C. Maternal Xenopus Zic2 negatively regulates Nodal-related gene expression during anteroposterior patterning.. *Development* **132,** 4845–55 (2005).

46.Kofron, M. *et al.*. Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors.. *Development* **126,** 5759–70 (1999).

47.Hilton, E., Rex, M. & Old, R. VegT activation of the early zygotic gene Xnr5 requires lifting of Tcf-mediated repression in the Xenopus blastula.. *Mech Dev* **120,** 1127–38 (2003).

48.Clements, D. & Woodland, H. R. VegT induces endoderm by a self-limiting mechanism and by changing the competence of cells to respond to TGF-beta signals.. *Dev Biol* **258,** 454–63 (2003).

49.Taverner, N. V. *et al.*. Microarray-based identification of VegT targets in Xenopus.. *Mech Dev* **122,** 333–54 (2005).

50.Bjornson, C. R. *et al.*. Eomesodermin is a localized maternal determinant required for endoderm induction in zebrafish.. *Dev Cell* **9,** 523–33 (2005).

51.Teo, A. K. *et al.*. Pluripotency factors regulate definitive endoderm specification through eomesodermin.. *Genes Dev* **25,** 238–50 (2011).

52.Xu, P. *et al.*. Maternal Eomesodermin regulates zygotic nodal gene expression for mesendoderm induction in zebrafish embryos.. *J Mol Cell Biol* **6,** 272–85 (2014).

53.De Domenico, E., Owens, N. D., Grant, I. M., Gomes-Faria, R. & Gilchrist, M. J. Molecular asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere transcript sequencing.. *Dev Biol* **408,** 252–68 (2015).

54.Owens, N. D. L. *et al.*. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development.. *Cell Rep* **14,** 632–647 (2016).

55.Owens, D. A. *et al.*. High-throughput analysis reveals novel maternal germline RNAs crucial for primordial germ cell preservation and proper migration.. *Development* **144,** 292–304 (2017).

56.Zhang, C., Basta, T., Fawcett, S. R. & Klymkowsky, M. W. SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in Xenopus.. *Dev Biol* **278,** 526–41 (2005).

57.Ku, M. & Melton, D. A. Xwnt-11: a maternally expressed Xenopus wnt gene.. *Development* **119,** 1161–73 (1993).

58.Tao, Q. *et al.*. Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos.. *Cell* **120,** 857–71 (2005).

59.Rebagliati, M. R., Weeks, D. L., Harvey, R. P. & Melton, D. A. Identification and cloning of localized maternal RNAs from Xenopus eggs.. *Cell* **42,** 769–77 (1985).

60.Birsoy, B., Kofron, M., Schaible, K., Wylie, C. & Heasman, J. Vg 1 is an essential signaling molecule in Xenopus development.. *Development* **133,** 15–20 (2006).

61.Chen, X., Rubock, M. J. & Whitman, M. A transcriptional partner for MAD proteins in TGF-beta signalling.. *Nature* **383,** 691–6 (1996).

62.Chen, X. *et al.*. Smad4 and FAST-1 in the assembly of activin-responsive factor.. *Nature* **389,** 85–9 (1997).

63.Chiu, W. T. *et al.*. Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program.. *Development* **141,** 4537–47 (2014).

64.Slagle, C. E., Aoki, T. & Burdine, R. D. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and Eomesodermin.. *PLoS Genet* **7,** e1002072 (2011).

65.Nelson, A. C. *et al.*. Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression.. *BMC Biol* **12,** 81 (2014).

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

67.Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why?. *Mol Cell* **49,** 825–37 (2013).

68.Foo, S. M. *et al.*. Zelda potentiates morphogen activity by increasing chromatin accessibility.. *Curr Biol* **24,** 1341–1346 (2014).

69.Sun, Y. *et al.*. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.. *Genome Res* **25,** 1703–14 (2015).

70.Yáñez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D. & Stark, A. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding.. *Genome Res* **22,** 2018–30 (2012).

71.Xu, Z. *et al.*. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila.. *Genes Dev* **28,** 608–21 (2014).

72.Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network.. *Nat Genet* **49,** 935–940 (2017).

73.De, I. A. *et al.*. DUX-family transcription factors regulate zygotic genome activation in placental mammals.. *Nat Genet* **49,** 941–945 (2017).

74.Hendrickson, P. G. *et al.*. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons.. *Nat Genet* **49,** 925–934 (2017).

75.Lee, M. T. *et al.*. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition.. *Nature* **503,** 360–4 (2013).

76.Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates.. *Science* **341,** 1005–9 (2013).

77.Joseph, S. R. *et al.*. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos.. *Elife* **6,** (2017).

78.Cho, K. W. Y., Blumberg, B., Steinbeisser, H. & Robertis, E. M. D. Molecular nature of Spemanns organizer: the role of the Xenopus homeobox gene goosecoid. *Cell* **67,** 1111–1120 (1991).

79.Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Herrmann, B. G. Expression of a xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67,** 79–87 (1991).

80.Artinger, M., Blitz, I., Inoue, K., Tran, U. & Cho, K. W. Y. Interaction of goosecoid and brachyury in Xenopus mesoderm patterning. *Mechanisms of Development* **65,** 187–196 (1997).

81.Trindade, M., Tada, M. & Smith, J. C. DNA-Binding Specificity and Embryological Function of Xom (Xvent-2). *Developmental Biology* **216,** 442–456 (1999).

82.Saka, Y. & Smith, J. C. A mechanism for the sharp transition of morphogen gradient interpretation in Xenopus. *BMC Developmental Biology* **7,** 47 (2007).

83.Boyer, L. A. *et al.*. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122,** 947–956 (2005).

84.Thomson, M. *et al.*. Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell* **145,** 875–889 (2011).

85.Shu, J. *et al.*. Induction of pluripotency in mouse somatic cells with lineage specifiers.. *Cell* **153,** 963–75 (2013).

86.Levine, M. & Davidson, E. H. Gene regulatory networks for development.. *Proc Natl Acad Sci U S A* **102,** 4936–42 (2005).

87.Peter, I. S. & Davidson, E. H. Assessing regulatory information in developmental gene regulatory networks.. *Proc Natl Acad Sci U S A* **114,** 5862–5869 (2017).

88.Gupta, R., Wills, A., Ucar, D. & Baker, J. Developmental enhancers are marked independently of zygotic Nodal signals in Xenopus.. *Dev Biol* **395,** 38–49 (2014).

89.Nakamura, Y., de, P. A. E., Veenstra, G. J. & Hoppler, S. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules.. *Development* **143,** 1914–25 (2016).

90.Kjolby, R. A. S. & Harland, R. M. Genome-wide identification of Wnt/β-catenin transcriptional targets during Xenopus gastrulation.. *Dev Biol* **426,** 165–175 (2017).

91.Collart, C. *et al.*. High-resolution analysis of gene activity during the Xenopus mid-blastula transition.. *Development* **141,** 1927–39 (2014).

92.Veenstra, G. J. C., Destrée, O. H. J. & Wolffe, A. P. Translation of Maternal TATA-Binding Protein mRNA Potentiates Basal but Not Activated Transcription in Xenopus Embryos at the Midblastula Transition. *Molecular and Cellular Biology* **19,** 7972–7982 (1999).

93.Stennard, F., Zorn, A. M., Ryan, K., Garrett, N. & Gurdon, J. B. Differential expression of VegT and Antipodean protein isoforms in Xenopus. *Mechanisms of Development* **86,** 87–98 (1999).

94.Amodeo, A. A., Jukam, D., Straight, A. F. & Skotheim, J. M. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition.. *Proc Natl Acad Sci U S A* **112,** E1086–95 (2015).

95.Collart, C., Allen, G. E., Bradshaw, C. R., Smith, J. C. & Zegerman, P. Titration of four replication factors is essential for the Xenopus laevis midblastula transition.. *Science* **341,** 893–6 (2013).

96.Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition.. *Annu Rev Cell Dev Biol* **30,** 581–613 (2014).

97.Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates.. *Dev Cell* **42,** 316–332 (2017).

98.Yang, J., Tan, C., Darken, R. S., Wilson, P. A. & Klein, P. S. Beta-catenin/Tcf-regulated transcription prior to the midblastula transition.. *Development* **129,** 5743–52 (2002).

99.Hudson, C., Clements, D., Friday, R. V., Stott, D. & Woodland, H. R. Xsox17α and -β Mediate Endoderm Formation in Xenopus. *Cell* **91,** 397–405 (1997).

100.Rosa, F. M. Mix.1 a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of Xenopus embryos. *Cell* **57,** 965–974 (1989).

101.Henry, G. L. Mixer a Homeobox Gene Required for Endoderm Development. *Science* **281,** 91–96 (1998).

102.Ault, K. T., Dirksen, M. L. & Jamrich, M. A novel homeobox gene PV.1 mediates induction of ventral mesoderm in Xenopus embryos.. *Proceedings of the National Academy of Sciences* **93,** 6415–6420 (1996).

103.Melby, A. E., Clements, W. K. & Kimelman, D. Regulation of Dorsal Gene Expression in Xenopus by the Ventralizing Homeodomain Gene Vox. *Developmental Biology* **211,** 293–305 (1999).

104.Suri, C. Xema a foxi-class gene expressed in the gastrula stage Xenopus ectoderm, is required for the suppression of mesendoderm. *Development* **132,** 2733–2742 (2005).

Chapter 2

Assembly of maternal transcription factors initiates the emergence of tissue-specific zygotic cis-regulatory regions

* The text of this chapter of the dissertation is part of an unpublished manuscript. The co-author Ken W.Y. Cho listed in this publication directed and supervised research which forms the basis for the thesis/dissertation. I would also like to acknowledge Ira L. Blitz, Masani Coley and Jessica Cheung from the University of California, Irvine; as well as Masanori Taira from Chuo University, Tokyo, Japan and Norihiro Sudou from the Women's Medical University, Tokyo, Japan for their work on this manuscript.

Abstract

Gene transcription relies on the dynamic interaction between chromatin states and transcription factors. Elucidation of the sequence of events underlying these interactions is essential for understanding tissue specification, differentiation, cellular reprogramming and cancer. Early embryogenesis provides a useful experimental system to study these events. Maternal transcription factors function at the onset of zygotic genome activation to specify the primary germ layers. We show that the genome-wide interactions of *Xenopus* Otx1, Vegt and Foxh1 occur prior to the onset of zygotic genome activation. These transcription factors co-occupy mesendodermal cis-regulatory modules before the deposition of enhancer epigenetic marks, thus pre-marking regulatory regions of the genome. Functionally, endoderm-specific factors Otx1 and Vegt co-regulate genes and act as dual function transcription factors inducing endodermal targets, while also functioning as repressors to exclude mesodermal gene expression in the endodermal germ layer. Overall, our datasets highlight the importance of transcription factor assemblies during gene regulation.

Introduction

Metazoan development begins with a single totipotent cell that gives rise to numerous cell types, each expressing lineage-restricted sets of genes. The activation of gene transcription in the embryo relies on maternal transcription factors (TFs), which sit high in the regulatory hierarchy to coordinate the gene regulatory cascades that lead to stereotypical development of embryos. Zygotic gene activation (ZGA) is a major regulatory event where control of the zygotic genome is transferred from maternal to zygotic TFs[1]. While the timing of the transition differs between species, all undergo ZGA, which is controlled by maternal inputs. With the exception of amniotes, the processes of germ layer specification are coupled to ZGA and depend on unequally distributed maternal determinants present in the egg before fertilization[2]. These maternal inputs are the first inputs to specify the differentiation of germ layer cell types. Here, we examine the earliest lineage restriction events of germ layer specification (the delineation of ectoderm, mesoderm, endoderm) using *Xenopus* and address how maternal TFs regulate embryonic genome programming.

Maternal TFs presumably coordinate the actions of enhancers, which are cis-regulatory modules (CRMs) that dock numerous TFs, to regulate the activity of core gene promoters. Since TFs drive lineage-specific transcription programs by binding CRMs dispersed throughout the genome, a major question that remains to be addressed is how maternal TFs bind specific regions in the chromatin to endow the transcriptional responses that initiate germ layer specification. We address this fundamentally important biological question by going back to the earliest stages of embryonic development when transcription

from the embryonic genome has not yet begun, the number of different cell types is small, and the genome appears relatively naïve.

Genomic studies of human and mouse embryonic stem cells have suggested a role for epigenetic priming of CRMs and promoter regions[3,4]. Priming of these sites is thought to have important roles in the deployment of gene regulatory networks (GRNs). However, examination of the chromatin states of early *Xenopus*, zebrafish and *Drosophila* embryos have not been able to find epigenetic signatures that are frequently associated with active or repressive enhancer states. In *Xenopus*, histone H3R8me2 activating marks appear on at least two genes, *sia1* and *nodal3*, expressed before the major wave of ZGA[5], but the majority of enhancer chromatin marks appear to arise during the major ZGA phase[6-9]. Therefore, prior to ZGA, cleavage and early blastula stage chromatin is free of both the activating enhancer marks H3K4me1 and H3K27ac, and permissive H3K4me3 as well as repressive H3K27me3 histone marks. Similarly, in zebrafish, neither H3K4me3 nor H3K27me3 were detected before ZGA; and a study of a larger cohort of histone modifications in *Drosophila* which includes H3K4me1, H3K4me3, and H3K27me3, showed that epigenetic states are largely established post-ZGA[10,11]. Furthermore, we previously demonstrated that the maternally-expressed forkhead-domain TF Foxh1 binds to the genome during cleavage stages prior to ZGA, before the appearance of activating enhancer marks, suggesting that Foxh1 selectively pre-marks CRMs on mesendodermal genes[9]. These findings raise an important question as to how active CRMs are chosen and activated during germ layer specification by maternal TFs.

In *Xenopus*, maternal RNAs are specifically enriched animally (future ectoderm) or vegetally (future endoderm) in the egg, and these factors are subsequently asymmetrically inherited by different blastomeres. Specification of the endodermal germ layer is regulated by the maternal T-box TF, Vegt, which sits at the top of the hierarchy of both endodermal and mesodermal gene regulatory cascades[12-18]. These experiments show that Vegt is capable of activating both mesodermal and endodermal genes, implying that other TFs are involved in restricting Vegt's function to endoderm gene activation in the vegetal pole cells fated to this lineage.

We therefore searched for other maternal TFs that are localized in the prospective endoderm region, and report here an unexpected role of Otx1 in endoderm specification. Using a combination of ChIP-seq and transcriptomic data generated from both gain- and loss-of-function analyses of Vegt and Otx1, we show that a combinatorial code of maternal TFs binds to a set of CRMs in the genome before the onset of zygotic transcription to begin endodermal differentiation programs. Our finding is consistent with the view that early lineage specification is driven by the binding of critical maternal TFs to selective CRMs that regulate cell-type-specific gene expression, and recruit epigenetic regulators that modify the chromatin template to further facilitate the assembly of transcriptional machinery as development proceeds.

Results

Identification of maternal vegetally-localized transcription factors

We wished to identify maternal TFs that might play roles in endodermal specification in the early embryo. Therefore, we examined transcript localization in cleavage stage *Xenopus tropicalis* embryos before the onset of ZGA. Eight-cell stage embryos were dissected into animal and the vegetal blastomere tiers. RT-qPCR and RNA-seq analysis on these samples showed the expected vegetal localization of *vegt* and *gdf1/vg1* RNAs, as well as animal localization of *foxi2* RNA (Fig. 2.1a; Fig. 2.2a-d). RNA-seq analysis on each tier, in biological triplicates, (Fig 2.2e) identified 309 coding genes that are differentially expressed between the animal (Table 2.1) and vegetal cells (Table 2.2). While hundreds of TFs are expressed in the egg[19], only a small fraction is expressed in a regionalized manner. We found 9 TFs differentially localized vegetally, while 4 are enriched animally (Fig 2.1b), using the criteria where the posterior probability > 0.95 and the fold-difference > 1.5. The *vegt* and *otx1* genes are among the most highly expressed vegetally, and are the most differentially expressed (>8 fold) between vegetal and animal blastomeres (Fig 2.1c). Our findings are consistent with previous RNA-seq based screens for localized factors which showed the vegetal enrichment of *otx1*[20,21]. While Vegt is a well known T-box TF that regulates endoderm formation via the regulation of zygotic endodermal TF and Nodal genes[16-18,22], the role of Otx1 in the endoderm is still unknown.

Previously, *otx1* transcripts were detected in the vegetal pole of oocytes[23], however its expression pattern in early embryos has not been well described. By whole mount *in situ* hybridization we found that *otx1* remains expressed in the vegetal mass (presumptive endoderm) and is excluded from the presumptive mesoderm prior to ZGA (Fig 2.1d). At this stage, mesoderm is localized above the equator[24] and epiboly moves this tissue into the equator by the beginning of gastrulation. By stage 9 this mesoderm lacking *otx1* can be seen in the marginal zone (Fig 2.1e) and *otx1* expression persists in the endoderm through early gastrulation (stage 10.5)(Fig 2.1f). *otx1* is also seen in gastrula stage neuroectoderm in a pattern reminiscent of *otx1* orthologs, *otx2* and *crx*[25-27]. The exclusion of *otx1* RNA in the putative mesoderm is in contrast to *vegt*, which is expressed vegetally as maternal RNA and in the gastrula marginal zone after the onset of zygotic transcription[12-15]. The expression pattern of *otx1* suggests that it plays a role specifically in endodermal development.

While Otx1 has not yet been implicated in vertebrate germ layer formation or early patterning, echinoderm maternal Otx TFs are known to function as important regulators of endodermal specification[28,29]. By mining the literature and transcriptomic datasets[23,28-46] we find that *otx1* orthologs are expressed maternally and vegetally across metazoan evolution, with the exception of amniotes (Fig 2.3), suggesting a conserved role in germ layer patterning. We therefore hypothesized that Otx1 is an important regulator of endoderm formation in *Xenopus*, perhaps functioning together with maternal Vegt.

Vegt and Otx1 combinatorially regulate the endodermal transcriptome

To gain insight into the role of Otx1 in the endoderm, we first expressed Otx1 in animal cap cells (future ectoderm) and assayed for induction of endodermal genes by RT-qPCR at early gastrula stage 10.5 (Fig 2.4a). Otx1 is sufficient to activate endodermal markers such as *nodal*, *mixer* and *darmin*, in addition to *foxa2,* which is expressed in both endodermal and mesodermal layers.

Next, we sought to determine whether Otx1 and Vegt collaborate to regulate endodermal gene expression. Since Vegt is capable of activating endodermal genes in the animal cap[15,17], we considered the possibility that Vegt and Otx1 can collaboratively induce the expression of endodermal genes. In addition, we also considered the possibility that Otx1 can inhibit the expression of mesodermal genes as (1) Otx1 has been shown to act as a repressor[47,48], (2) zygotic *vegt* is a mesoderm gene regulator and (3) maternal *otx1* is exclusively expressed in endodermal cells. We therefore microinjected a single dose of *vegt* or *otx1* mRNA or a cocktail of these two mRNAs into the animal pole at 1-cell stage, and assayed for endodermal and mesodermal gene expression in dissected animal caps. As expected, Otx1 and Vegt alone act as activators of the endodermal genes *nodal* and *mixer* (Fig 2.4b). Combinatorially, Otx1 and Vegt additively activate the expression of *mixer*. Interestingly, we find a strong synergistic induction of *nodal* in the presence of both mRNAs. Conversely, Otx1 strongly downregulates the induction of the mesodermally-expressed gene *fgf20* caused by Vegt[49]. To further support these findings, we then performed dose response experiments, using the same animal cap paradigm, where a sub-

threshold concentration of *vegt* mRNA was kept constant, and increasing doses of *otx1* were added (Fig 2.4c). Assaying the expression of the same genes, increasing doses of Otx1 induced *nodal* and *mixer* expression, with *nodal* being more sensitive to Otx1 dose than *mixer*. On the contrary, *fgf20* was strongly repressed even at the lowest Otx1 dose. This finding also proposes a dual function for maternal Otx1 - promoting endoderm development vegetally, while repressing mesodermal development in the endoderm by blocking the expression of *fgf20*.

We next examined the genome-wide interaction between Otx1 and Vegt in regulating mesendodermal genes by employing the same *vegt* and *otx1* mRNA titration conditions followed by RNA-seq analysis. To identify similarly regulated genes, we searched our RNA-seq expression profiles for the dose-response patterns of *nodal, mixer* and *fgf20*. Pearson correlation metric was used to identify genes with the most similar regulation to the patterns of *nodal, mixer* and *fgf20*. We ranked the Pearson correlation coefficient and found the 50 most highly correlated genes for the *nodal*-type, *mixer*-type and *fgf20*-type regulation (Table 2.3). Among genes that are regulated similarly to the *nodal*- and *mixer*-type, are *tbx3, lefty, lhx1, snai1, cer1*, and *mix1*, which show enriched expression in the endoderm of the early gastrula. Among the *fgf20*-type regulated genes are the mesodermally-expressed genes *fgf4* and *wnt11b*. Consistently, when we plotted the spatial expression during gastrula stage of *nodal*-, *mixer*- and *fgf20*-types of co-regulation, we find *nodal*- and *mixer*-type genes to be expressed generally in the endoderm, while *fgf20*-type genes were expressed in the mesoderm (Fig 2.4d).

Our analysis so far has been based on ectopic expression assays. Therefore, to assess the combinatorial roles of endogenous Otx1 and Vegt, we blocked their activity by injecting translation blocking antisense morpholino oligonucleotides (MOs) into embryos, either independently or in combination. We used a previously published *vegt* morpholino[50], and we designed and tested a translation blocking morpholino to block Otx1 function, which we showed to be specific through rescue experiments (Fig 2.5). Since Vegt and Otx1 are expressed zygotically in other regions of the embryo, vegetal masses were dissected at gastrula stage 10.5 to isolate these TFs functions in the endoderm. RNA-seq was then performed on these explants in biological duplicates to assess Vegt's and Otx1's specific roles on the endodermal transcriptome. As the period around zygotic gene activation involves drastic changes in gene expression, we found difficulty in using traditional differential expression callers in identifying activated or inhibited genes as these rely on gene expression in read counts or transcripts per million[51]. Rather, we relied on the fold changes in both biological replicates and defined differentially expressed genes as 2-fold activated or inhibited. In addition, a gene expression cutoff at 10 transcripts per million was imposed to avoid genes with extremely high fold changes due to low expression, which is notoriously noisy[52]. Single knockdowns of Vegt or Otx1 affected the expression of 301 and 225 genes, respectively (Fig 2.4e,f). However, a much larger cohort of 450 genes were affected in expression in the presence of both MOs. Furthermore, when examining the expression of genes that are only down-regulated in the double MO but not the individual MOs, we noticed that these genes are generally reduced in expression in both the Vegt MO and the Otx1 MOs, although not as strongly as by the double MO (Fig 2.4g). Similarly, genes that are only activated in the double MO are generally activated in both Vegt and Otx1 MOs

(Fig 2.4h). These expression patterns suggest that Vegt and Otx1 are combinatorially required to regulate the expression of these genes. Among these genes that are affected are mesodermally-expressed genes *fgf20* and *mespa*, which are both inhibited by Vegt and Otx1 (Fig 2.4i). Endodermally-expressed genes *foxa1* and *hnf1b*, on the other hand, are activated by both Vegt and Otx1. While we do not see as stark of a contrast as observed in animal cap ectopic expression assays (Fig 2.4b-d), the spatial expression analysis of Vegt and Otx1 co-regulated genes shows that negatively co-regulated genes are enriched in mesodermal expression, whereas positively co-regulated genes are enriched in endodermal expression (Fig 2.4j). Overall, our ectopic expression and knockdown experiments support the notion that maternal Vegt and Otx1 function in a combinatorial manner to drive proper expression of endodermal genes while minimizing vegetal expression of mesodermal genes.

Vegt and Otx1 co-bind to endodermal cis-regulatory modules before the onset of ZGA

Previously, we showed that the ubiquitously expressed Foxh1 binds to the genome prior to ZGA, and before the appearance of promoter-bound RNA pol II and H3K4me1 and Ep300 enhancer marks[9]. We therefore asked whether endoderm-specific TFs Vegt and Otx1 also interact with the genome before the onset of ZGA. Chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq) was used to investigate Vegt and Otx1 binding at early blastula stage 8. Vegt antibody has been used for ChIP-qPCR previously[53] and the quality of our Otx1 antibody has been demonstrated using ChIP-qPCR analysis (Fig 2.6). Biological replicate datasets were used to identify reproducible and high confidence TF-bound regions and a total of 21,711 and 5,151 bound regions were identified for Vegt and Otx1,

respectively. *De novo* motif analysis of the bound sequences from the ChIP-seq data identified a variation of the bicoid-type homeodomain (TAATCCCY)[54] and T-box half-site motifs (TCACACCT)[55] for Otx1 and Vegt respectively (Fig 2.7). Both Vegt and Otx1 bind to mesendodermal genes such as *nodal6, sox17b, mix1, mixer, fgf20, gsc* and *hhex,* and their binding regions (peaks) are highly overlapping (Fig 2.8a,b). The set of Vegt and Otx1 overlapping peaks represents 64% of Otx1 peaks and 18% of Vegt peaks. Simulation of a set of peaks similarly-sized to the Otx1 peak set across the genome show an average overlap of 56 (1.1%) peaks over 1,000 trials with the 21,711 Vegt peaks ($p < 2.2e-16$), suggesting that the overlap between Vegt and Otx1 chromatin binding is statistically significant.

We examined the distribution of peaks relative to genomic features. Approximately 1.1% of the entire *X. tropicalis* genome represents promoter proximal sequence (within 1kb upstream of TSSs), ~25% is intronic and ~70% is intergenic (Fig 2.8c). Vegt and Otx1 binding is enriched 2-3 fold in promoter proximal regions and ~1.3 fold in intronic regions at the expense of intergenic regions. Despite this enrichment, > 90% of Vegt and Otx1 binding reside in the intronic and intergenic regions, suggesting binding to enhancers over promoters. By assigning these binding regions to the nearest genes, we find that Vegt and Otx1 tend to bind to endodermal and mesodermal genes, rather than ectodermal genes (Fig 2.8d).

Previously, we showed that the maternal TF, Foxh1, is bound to CRMs as early as cleavage stages, and prior to the establishment of enhancer marks Ep300 and H3K4me1[9]. We asked

whether endoderm-specific TFs could engage the chromatin during these early stages of development. We tested whether Vegt and Otx1 also interact with the genome by performing ChIP-qPCR on 32-cell stage embryos, 3-4 cell cycles prior to ZGA (Fig 2.8e,f). Indeed, strong binding of endogenous Vegt and Otx1 was detected during pre-ZGA stages on the regulatory regions of mesodermal and endodermal genes. This suggests that maternal TFs in general engage the chromatin during the transcriptionally quiescent cleavage stages, prior to establishment of enhancer marks.

Identification of Vegt and Otx1 direct target genes defines these TFs as dual function activators and repressors

By combining these ChIP-seq peaks with the findings from the knockdown experiments, we identified a set of Vegt-regulated (Table 2.4), Otx1-regulated (Table 2.5) and co-regulated (Table 2.6) genes during zygotic gene activation. Vegt-activated genes include *nodal5.2* and *cer1*, both of which have previously been identified as Vegt direct targets[18,53,56,57]. Among the mostly newly identified 45 direct targets (Table 2.4) of Vegt is *nodal6*, which harbors multiple ChIP-seq peaks upstream of the gene. Surprisingly, as Vegt is not a known repressor, we identified a total of 34 genes directly repressed by Vegt, which includes mesodermal genes *fgf20*, *mespa* and *mespb*. This suggests that Vegt could have a previously unidentified repressive function. On the other hand, Otx1 directly inhibits 8 genes, which include the mesodermal gene *fgf20* (Table 2.5). Otx1 directly activates 15 genes, which includes the endodermal genes *gata4* and *foxa1*. This is interesting as both homologs of

gata4 (*gatae*) and *foxa1* (*foxa*) are direct targets of the sea urchin *otx* gene[29]. This suggests a functional conservation between the *Xenopus* and sea urchin endoderm GRNs.

In order to identify genes that require the combinatorial regulation of Vegt and Otx1, we imposed the three conditions (1) Vegt binding; (2) Otx1 binding; and (3) either the gene is affected by all three conditions (Vegt MO, Otx1 MO and double MO) or the gene is affected by double MO but not by single MOs. This list may be limited since, as previously mentioned, co-regulated genes can be affected by single MOs, although not as strongly as by double MOs (Fig 2g,h). Nonetheless, based on these conditions, 8 genes are inhibited and 22 genes are activated combinatorially and directly by Vegt and Otx1 (Table 2.6). Overall, our chromatin binding analysis and perturbation experiments both support the model that co-regulation by Vegt and Otx1 is a key component of early endoderm formation.

Maternal Otx1 and Vegt assemble on cis-regulatory regions together with Foxh1

De novo motif analysis of Otx1 and Vegt ChIP-seq bound regions identified their respective motifs. In this analysis, we also identified a set of maternal TFs that are likely candidate TFs that co-regulate target genes together with Vegt and Otx1 (Fig 2.7). Prominently, we found the Fox motif in the Otx1 ChIP-seq peak regions and Vegt-Otx1 co-bound regions, suggesting an interplay between Foxh1, Vegt and Otx1. When we compared Otx1 and Vegt peaks to Foxh1 peaks, all derived from embryos of the same developmental stage (blastula stage 8), we find that both Vegt and Otx1 binding overlaps with Foxh1 (Fig 2.9a,b). Greater than 70% of Otx1 peaks and greater than 25% of Vegt peaks overlap with Foxh1 peaks.

This peak overlap might indicate co-occupancy of Foxh1, Vegt and Otx1 on CRMs within the same vegetal cells. To test this hypothesis, we performed sequential ChIP-qPCR. First, we performed ChIP using Otx1 antibody, followed by dissociation, and a second round of ChIP using Vegt, Foxh1 or Flag (as a negative control) antibodies. Regions bound by Otx1 are co-occupied by Foxh1 and Vegt within the same endodermal cells (Fig 2.9b,c). This suggests that these maternal TFs assemble on cis-regulatory regions during cleavage through blastula stages.

Possibly, co-occupancy favors the binding of TFs on these CRMs over other regions of the genome. To test this model, we knocked down Foxh1 expression using a previously published, translation-blocking morpholino[58] and assessed Otx1 and Vegt binding by ChIP-qPCR at blastula stage 8. In the absence of Foxh1, association of both Otx1 and Vegt in multiple regions is decreased (Fig 2.9d). These observations suggest that these three maternal TFs co-occupy CRMs in the genome as an assembly.

Pre-marking of endodermal CRMs by maternal Otx1/Vegt/Foxh1 assembly

We wished to determine whether Otx1, Vegt and Foxh1 co-bound regions correspond to functional cis-regulatory modules. Since CRMs are functional gene regulatory elements that tend to dock multiple TFs, ChIP-seq binding of various *Xenopus tropicalis* zygotic TFs were analyzed. This set of zygotic TFs includes endodermally active TFs Ctnnb1/β-catenin, Foxa, Gsc, Otx2, and Smad2/3[9,58-60]. Regions near endodermally-expressed genes such as *foxa1, pnhd* and *gata4*, which are co-bound by maternal TFs, show extensive co-binding of zygotic

TFs as well (Fig 2.10a). The promoter region of *gata4* shows co-binding of all maternal and zygotic TFs analyzed. Genome-wide analysis support this finding, as regions that are co-bound by all three maternal transcription factors display ChIP-seq signal of zygotic TFs (Fig 2.10a,b). Interestingly, co-binding of Smad2/3 signal is enriched at sites triply bound with Otx1, Vegt and Foxh1, suggesting an interplay between Nodal signaling and all three of these maternal TFs. These observations suggest that the Otx1, Vegt and Foxh1 co-bound regions are sites for subsequent assembly of zygotically-active TF recruitment.

While CRMs that act as enhancers typically display epigenetic marks, in early frog, fish and *Drosophila* embryos enhancer marks are not present until after ZGA[7,8,10,11]. We therefore wished to determine whether Otx1, Vegt and Foxh1 co-bound regions are subsequently decorated with enhancer marks. We interrogated staged histone mark datasets for the presence of H3K4me1, H3K27ac and Ep300, well known enhancer marks [7,8]. The Ep300 active enhancer mark is not readily detectable until blastula stage 9, after ZGA has begun and well after the stage when our current ChIP-seq datasets were generated (stage 8). Post-ZGA Ep300 displays strong signal at the regions of maternal TF co-binding (Fig 2.10a,c). Similarly, the H3K27ac active enhancer mark is weakly present at stage 8 and is largely established during early gastrula (Fig 2.10d). The general enhancer mark H3K4me1 is absent at the stage of our ChIP-seq binding and this mark is largely established at stage 9 (Fig 2.10e). Interestingly, this mark appears to persist through later developmental stages. These datasets, taken together, implicate Otx1, Vegt and Foxh1 co-bound regions as CRMs that likely function as enhancer elements that recruit other zygotically expressed TFs at later stages.

Discussion

In anamniote embryos, the first cell fate decisions are specification of the primary germ layers by maternal factors. Here, we identify a role for the maternal TF Otx1 in endodermal specification. The relationship between Otx1 and the maternal TFs Vegt and Foxh1 was explored. Interaction of these TFs with the genome occurs during cleavage stages and in the blastula, prior to the onset of zygotic transcription. By interrogation of genome-wide binding, we find numerous regions where all three TFs are present (Fig 2.9a,b) and sequential ChIP studies show that they co-occupy CRMs (Fig 2.9c), suggesting these TFs collaborate to regulate gene behavior. We also show that these regions of maternal TF co-binding are sites for docking numerous zygotically-active mesendodermal TFs (Fig 2.10a,b). We also demonstrate that these CRMs, while not significantly decorated with enhancer-associated histone marks or Ep300 before ZGA, become marked after ZGA (Fig 2.10a,c-e), supporting the notion that they are functionally active CRMs. We conclude that all three of these maternal TFs are akin to pioneering factors in that they pre-mark CRMs, genome-wide, before the acquisition of histone modifications associated with active enhancers.

The sequence of events leading to the appearance of open chromatin, epigenetic marks and TF binding is under intensive investigation. The early metazoan embryonic (pre-ZGA) genome lacks enhancer marks, and both permissive and repressive histone modifications (H3K4me3 and H3K27me3, respectively) are established during late blastula and gastrula stages[6-8,10,11]. Current evidence for the presence of open chromatin in the early *Xenopus*

embryo is lacking, however, data from Drosophila and zebrafish suggest that early embryonic chromatin is relatively closed prior to ZGA[11,61]. Indirect evidence in *Xenopus* and zebrafish supports this finding, as high nucleosome concentration prevents transcription, and transcription factors must compete with nucleosomes to regulate gene expression[62-64]. We find that combinatorial binding of maternal TFs Otx1, Foxh1 and Vegt occurs pre-ZGA, prior to the recruitment of H3K4me1 and Ep300. This suggests that chromatin binding by maternal TFs is not controlled by epigenetics, but rather by TF binding preferences for DNA motifs influenced by protein-protein interactions (Fig 2.9d) on CRMs.

These same maternal TFs may even be involved in recruitment of the writers of epigenetic marks. An example of epigenetic writer recruitment by TFs during *Xenopus* dorsal-ventral patterning is Ctnnb1/β-catenin's facilitation of Prmt2 binding to the *sia1* regulatory region[5]. Prmt2 then deposits H3R8me2 marks resulting in transcriptional activation of *sia1*. Interestingly, inhibition of zygotic transcription through a-amanitin treatment showed that maternal factors are involved in a majority of H3K4me3 promoter marks and H3K27me3 polycomb repressive marks; and a subset of Ep300 recruitment through late gastrula stages[8]. Recent evidence in human embryos shows that maternal factors, perhaps including OTX2, also control post-ZGA chromatin accessibility at the 8-cell stage[42]. Our current and previous data[9] demonstrate that maternal TF binding to CRMs precedes both the H3K4me3 and H3K27me3, and enhancer marks (Fig 2.10a,c-e), suggesting that TFs are the major drivers of site-selective chromatin modifications in the early embryo and not the reverse.

The function of maternal Vegt has been extensively characterized as an activator of endodermal and mesodermal genes[12-18]. However, when we analyzed RNA expression in the vegetal mass of Vegt morphants, mesodermal genes were induced, corroborating the previous finding that the putative mesoderm shifts to the vegetal mass in Vegt knockdowns[16]. Combined with the chromatin binding, our data supports a model whereby maternal Vegt acts as a direct repressor of mesodermal genes vegetally. This finding poses a novel inhibitory role for Vegt and suggests that Vegt acts as a dual function transcription factor. This would add Vegt to the list of T-box transcription factors that can act as repressors. This list includes mouse Tbx20, *Xenopus* Tbx3, zebrafish Tbx24, *C. elegans* Tbx2, and *Drosophila* Midline[65-69]. Mechanistically, the repressive roles of Tbx20, Tbx24 and Midline have been attributed to the interaction with the general co-repressor Tle/Groucho[67,69,70]. Alternatively, the switch from the activator to the repressive role of Tbx2 has been attributed to the SUMOylation of this protein[68]. These mechanisms from other T-box family members present some attractive hypotheses in studying the possible mechanism for Vegt's repressive function.

The role of maternal Otx1 in germ layer specification has not previously been investigated. In the current study we identified *otx1* mRNA as vegetally localized, consistent with previous studies on *Xenopus* oocytes and early embryos[20,21,23], and sought to understand its role in endoderm specification. Using both gain- and loss-of-function experiments we provide support for the notion that Otx1 collaborates with Vegt to activate endodermal gene expression and also to repress mesoderm formation (Fig 2.4). Repression of mesodermal gene expression appears to occur at least in part through Otx1-mediated

down-regulation of both zygotic *fgf20* and *fgf8* expression. Fgf signaling inhibits endodermal gene expression in both *Xenopus* and zebrafish[71-73] and is absent in the vegetal pole of normal blastulae, whereas it is active in the mesoderm[49,74]. The mechanism for toggling Otx1's dual functions on different genes remains unclear. A genome-wide study on the function of Otx2 in dorsal-ventral patterning similarly identified CRMs responding to activation and repression functions for this paralogous TF, whereby Otx2 together with Lhx1 functions as an activator, while Otx2 together with Gsc functions as a repressor[59]. Phosphorylation of Otx2 appears to play a role in its repressor function[75] and since these phosphorylation sites are also conserved in Otx1, it is possible that the dual role of Otx1 might be similarly regulated.

How does the role of Otx1 in early frog embryos apply to endoderm in other organisms? Otx genes are most well known for their roles in embryonic anterior specification, a function that pre-dates the evolutionary split between deuterostomes and protostomes. The only functional study on *otx* genes in the initial specification steps of endoderm comes from echinoderms, the sea urchin and starfish, where vegetally-localized *otx* was shown to activate expression of multiple endodermal genes[28,29]. Therefore, we mined publically available *in situ* hybridization and transcriptomic datasets to determine whether *otx1* orthologs are expressed maternally and vegetally (Fig 2.3). Our analysis suggests that Otx1 orthologs seem to play a role in endoderm formation across deuterostomes, with the exception of amniote embryos. Interestingly, our analysis shows that Otx1 activates *gata4* and *foxa1* in *Xenopus*, similar to the activation of *gatae* and *foxa* orthologs by Otx in sea urchin[29] suggesting conservation of GRNs. While much is known about the expression

pattern of Otx1 homologs, further functional analysis is needed to uncover their roles in diverse endodermal gene regulatory programs.

Our current study and previous work[9] are in line with the notion that assembly of maternal transcription factors initiates the emergence of tissue-specific zygotic cis-regulatory regions. While we have so far identified the assembly of maternal TFs on the genome, other maternal TFs undoubtedly are involved in this process. Some candidate maternally-expressed TFs include vegetally-localized Sox7 and ubiquitously expressed TFs such as Zic2, Sox3 and Pou5f3/Oct60, motifs of which are enriched under Vegt and Otx1 peaks. Important future questions that need to be addressed are how do a larger set of maternal TFs coordinate ZGA, both temporally and with appropriate spatial expression, and how do these TF assemblies differentially recruit chromatin modifiers that establish germ layer and cell type-specific epigenetic states.

Materials and Methods

Animal husbandry and embryo manipulation

Xenopus tropicalis females were injected with 10 units of Chorulon (Merck and Co.) 1-3 nights before embryo collection and 100 units of Chorulon on the day of embryo collection. Eggs were collected in a dish coated with 0.1% BSA in 1/9x MMR. The eggs are in vitro fertilized with sperm suspension in 0.1% BSA in 1/9x MMR. The embryos are dejellied with

3% cysteine in 1/9x MMR, pH 7.8, 10 minutes after fertilization and are then ready for manipulation. Embryos were staged using the Nieuwkoop-Faber developmental table[76,77].

Ectopic expression, morpholino knockdown and morpholino rescue

Ectopic expression: The *Xenopus tropicalis vegt* and *otx1* open reading frames are cloned into the BamH1 site of the pCS2+ vector using Gibson cloning[78]. Similarly, the Otx1 open reading frame was cloned into the BamH1 site of pCS2+3xFLAG vector to generate an epitope-tagged Otx1. The plasmids are then digested with HpaI and mRNA was generated using i*n vitro* using SP6 mMessage mMachine transcription kit (Thermo Fisher Scientific). For the ectopic expression experiments, *otx1* or *vegt* mRNA was injected into one-cell stage embryos on two opposite sides of the animal pole. Animal caps were dissected at stage 9 (6 hpf) and RNA was harvested from the animal caps at 7 hpf (approximately early gastrula at stage 10.25-10.5).

Knockdown: Otx1 translation blocking morpholino was designed by GeneTools, Inc., against *X. tropicalis* Otx1 (5'-ATGACATCATGCTCAAGGCTGGACA-3') and the Vegt morpholino was previously tested (5'-TGTGTTCCTGACAGCAGTTTCTCAT-3')[50]. The morpholinos or the Otx1 morpholino with the rescue construct (FLAG-otx1 mRNA) directly into the vegetal pole of one-cell stage embryos. Vegetal masses were dissected at 6 hpf and RNA was harvested from vegetal masses at 7 hpf for RNA analysis.

RNA assays and analysis

RNA isolation and RT-qPCR: RNA was collected from embryo and embryo fragments as previously described[79]. RNA samples were reverse transcribed using the MMLV reverse transcriptase. qPCR was performed using Roche Lightcycler 480 II using the Roche SYBR green master with the default SYBR green protocol. Fold change in gene expression is calculated using the DDCp approach and the error among technical replicates is calculated using the first approximation of the Taylor expansion for the DDCp value. **RNAseq and differential expression:** RNA-seq libraries were generated using Smart-seq2 cDNA synthesis followed by tagmentation[80], quality-tested using an Agilent Bioanalyzer 2100, quantified using KAPA qPCR and sequenced using Illumina sequencers at the UC Irvine Genomics High Throughput Facility. Reads were aligned using RSEM v.1.2.12[81] and Bowtie v2.2.7[82] to the *Xenopus tropicalis* genome version 9.0[83,84]. Differential expression between samples was performed using the EBseq v1.8.0[85] function *EBTest* on R v3.1.0[86]. **Gastrula stage RNA expression analysis:** Gastrula expression in transcripts per million of various tissue fragments were downloaded from NCBI GEO accession number GSE81458[19]. The expression pattern of each gene is z-score normalized and plotted in R v3.1.0 using the *heatmap* function[86]. **Pattern matching for similarly regulated genes:** In the combinatorial ectopic expression of Vegt and Otx1 experiment followed by RNA-seq, we pattern matched the whole genome expression in TPM with the TPM expression of *nodal*, *mixer* and *fgf20* with the *cor* function in R using the Pearson method. We then ranked the correlation metric of the entire genome to each of the patters and selected the 50 most correlated genes[86].

Whole mount *in situ* hybridization: We used the protocol of Harland[87] with modifications as previously described[25]. Template for *otx1* riboprobes was obtained following PCR amplification from cDNA prepared by Smart-seq2[80] reverse transcription. PCR product contained bacteriophage T3 and T7 promoters for the synthesis of sense and antisense probe, respectively. The forward and reverse primers for amplification were: 5'-GCAGCaattaaccctcactaaaggTTCAGCGGGGTGGATTGCAG-3' and 5'-GCAGCTAATACGACTCACTATAGGacacagggacaaacagagccaa-3'.

ChIP assays and analysis

ChIP-qPCR and ChIP-seq: ChIP on *X. tropicalis* embryos was performed as previously described[58]. ChIP DNA using 1 embryo equivalent, and ChIP input DNA at 0.1, 0.01 and 0.001 embryo equivalents assayed in triplicates by qPCR using the Roche Lightcycler 480 II using the Roche SYBR green master with the default SYBR green protocol. Percent input of ChIP DNA was calculated by generating a linear model of input DNA embryo equivalents and qPCR Cp values. ChIP-seq libraries were generated using Nextflex ChIP-seq kit (Bioo Scientific), analyzed using an Agilent Bioanalyzer 2100, quantified using KAPA qPCR and sequenced using Illumina instruments at the UC Irvine Genomics High Throughput Facility. **Peak calling and IDR:** Reads were aligned to the *X. tropicalis* genome v9.0[83,84] using Bowtie v2.2.7[82] with default options, and peaks were called against stage 8 input DNA[9] using Macs v2.0.10[88] with the option –p 0.001 but otherwise default options. ENCODE based irreproducibility discovery rate (IDR) was performed using the following p-value

thresholds for the following comparisons: 0.01 for original biological replicates, 0.02 for pseudoreplicates of each biological replicate and 0.0025 for pseudoreplicates generated from pooled reads of biological replicates[89]. **Motif analysis:** Motif analysis was performed using DREME[90] and matching identified motifs to databases was performed using TOMTOM[91]. **IGV track:** Using Samtools v0.1.19, aligned reads in SAM format were converted to BAM format (*samtools view –bS*) and duplicates were removed (*samtools rmdup*)[92]. Bedtools v2.19.1[93] was used to convert BAM files to BED format. IGVtools functions *sort* (default options) and *count* (-w 25 -e 250) were used to generate the WIG files loaded into IGV v2.3.20[94]. **ChIP signal heatmaps:** The ChIP signal of histone modifications, Ep300, zygotic TFs near Vegt, Otx1 and Foxh1 peaks were generated as follows. Using the TF summit file generated by MACS2 after peak calling, 2500 bp was added to the 5' and 3' sides of the summit to generate a 5000 bp peak window using Python v2.6.6. Datasets were downloaded from NCBI GEO using the accession numbers: GSE85273 for RNA polymerase II and stage 10 Foxa[9]; GSE67974 for H3K4me1 and Ep300[8]; GSE56000 for H3K4me1 and H3K27ac[7]; GSE72657 for stage 10 Ctnnb1/β-catenin[60]; and GSE53654 for stage 10 Smad2/3[58,]. The stage 10 Gsc and Otx2 ChIP-seq, and their respective input DNA datasets were obtained from the DDBJ Sequence Read Archive using the accession numbers DRA000508, DRA000510, DRA000576 and DRA000577[59]. For the datasets downloaded from NCBI GEO, Sratoolkit v2.8.1 was used to convert SRA files to FASTQ using the command *fastq-dump*. The ChIP datasets were processed as in IGV track all the way through the IGVtools *sort* function (see above). After sorting, the BED files from this analysis were overlaid into the TF ChIP 5000 bp peak window using Bedtools v2.19.1[93] function *coverageBed* . The *coverageBed* output is pre-processed to generate ChIP signal in

matrix form, and the signal matrix is plotted in R v3.1.0 using the *heatmap* function[86]. **Sequential ChIP**: The initial steps of ChIP are performed as usual. After incubation with the first antibody, elution is performed using 1x TE, pH 8.0, with 10mM DTT, 500 mM NaCl and 0.1% SDS at 37oC for 30 minutes. After the incubation, the eluate is diluted 10x with RIPA. The diluted eluate is incubated with the second antibody and the rest of the ChIP is performed as usual[95]. **Foxh1 knockdown:** Foxh1 protein was knocked-down using a previously tested translation blocking morpholino (Sequence: TCATCCTGAGGCTCCGCCCTCTCTA)[58]. The morpholino was injected directly into the zygote vegetal pole. Embryos were fixed ~4 hours post fertilization around stage 8 and assayed for either Vegt or Otx1 binding by ChIP-qPCR (see above).

Maternal gene expression of non-Xenopus Otx genes

The early embryonic RNA-seq datasets were obtained from NCBI GEO using the accession numbers GSE22830 for zebrafish[38] and GSE86592 for chick[41] and were aligned to their respective genomes[96,97] using RSEM v.1.2.12[81] and Bowtie v2.2.7[82]. The processed RNA expression in read counts of early embryonic *A. queenslandica* and *D. melanogaster* were obtained from NCBI GEO using the accession number GSE70185[45]. The axolotl[40], *N. vectensis*[46] and *C. intestinalis*[34] RNA expression were obtained as supplementary information from their respective publications.

Data Availability

Raw and processed RNA-seq and ChIP-seq datasets generated for this study are available at NCBI Gene Expression Omnibus using the accession GSE118024.

Acknowledgements

This work was made possible, in part, through access to the Genomic High-Throughput Facility Shared Resource of the Cancer Center Support Grant (P30CA-062203) at the University of California, Irvine, and NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and S10OD021718-01. We thank Xenbase for genomic and community resources (http://www.xenbase.org/, RRID: SCR_003280) in addition to their bioinformatic assistance; and the UC Irvine High Performance Computing Cluster (https://hpc.oit.uci.edu/) for their valuable resources and helpful staff. We would like to thank Margaret Fish, Rebekah Charney, Elmira Forouzmand, Elizabeth Chamu, Yuuri Yasuoka as well as previous and current members of the Ken Cho lab for experimental assistance, valuable advice and critical comments. This research was funded by the following grants awarded K.W.Y.C.: R01GM126395, R01HD073179 and NSF 1562176. K.D.P. was a recipient of T32-HD60555.

Author contributions

K.D.P., K.W.Y.C., and I.L.B. designed the reported experiments and wrote the manuscript. K.D.P., M.C., J.C. and I.L.B. performed wet bench experiments; and K.D.P. performed bioinformatics analyses. N.S. and M.T. generated the Vegt antibody.

Chapter 2 Figures and Tables

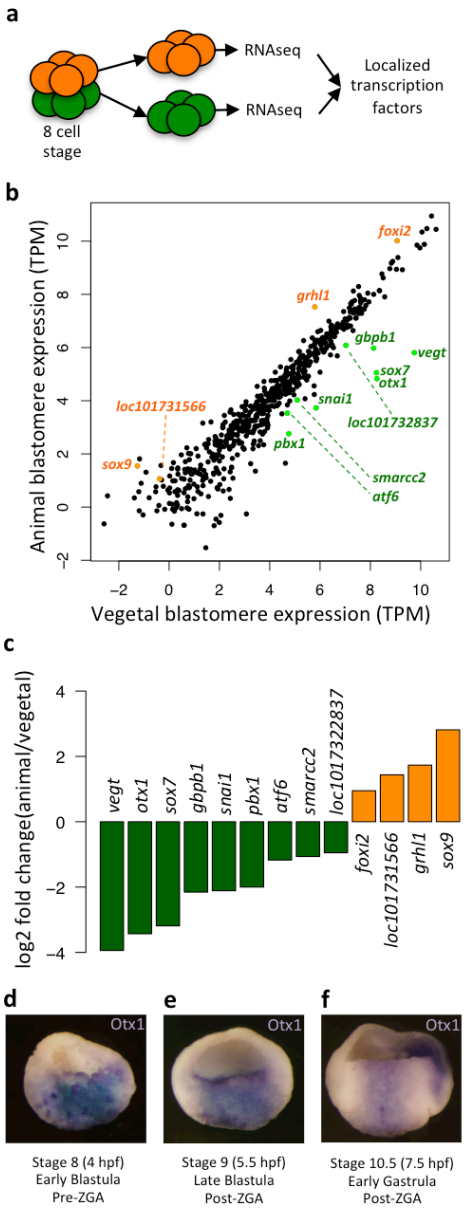

Figure 2.1. *Screen for core maternal endodermal transcription factors.* (a) Experimental design to screen for core endodermal TFs. (b) Log2-transformed transcripts per million (TPM) expression of TFs in animal and vegetal blastomeres. Differentially expressed TFs are highlighted in green (vegetal) or orange (animal). (c) Log2-transformed fold change of expression levels of localized TFs. *In situ* hybridization showing *otx1* expression pattern pre- and post-ZGA during early blastula (d), late blastula (e) and early gastrula (f).

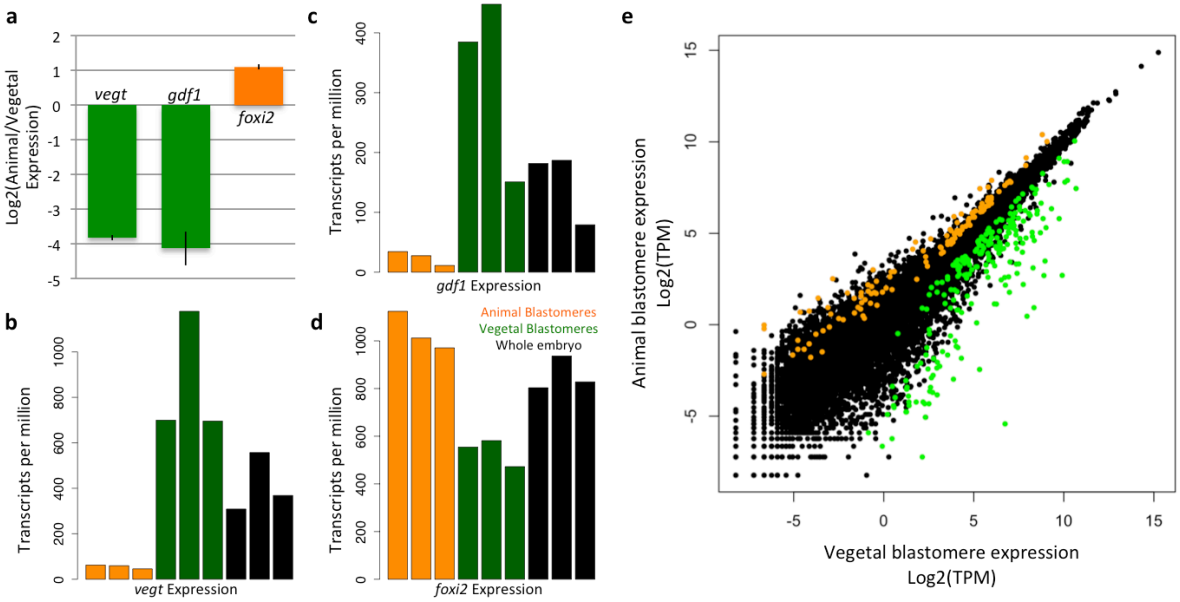

Figure 2.2. Validation of screen for core maternal endodermal transcription factors. (a) Validation of 8-cell dissections by RT-qPCR of animally- and vegetally-localized RNAs. Expression is in log2-transformed fold change of animal/vegetal. (b-d) RNA-seq expression in transcripts per million (TPM) of known localized RNAs. Experiments were performed in biological triplicates. *vegt* (b) and *gdf1* (c) RNAs are vegetally-localized, while *foxi2* (d) is animally-localized. (e) Average expression of genes in animal and vegetal blastomeres in log2-transformed transcripts per million. Differentially expressed genes are highlighted in green (vegetal) or orange (animal).

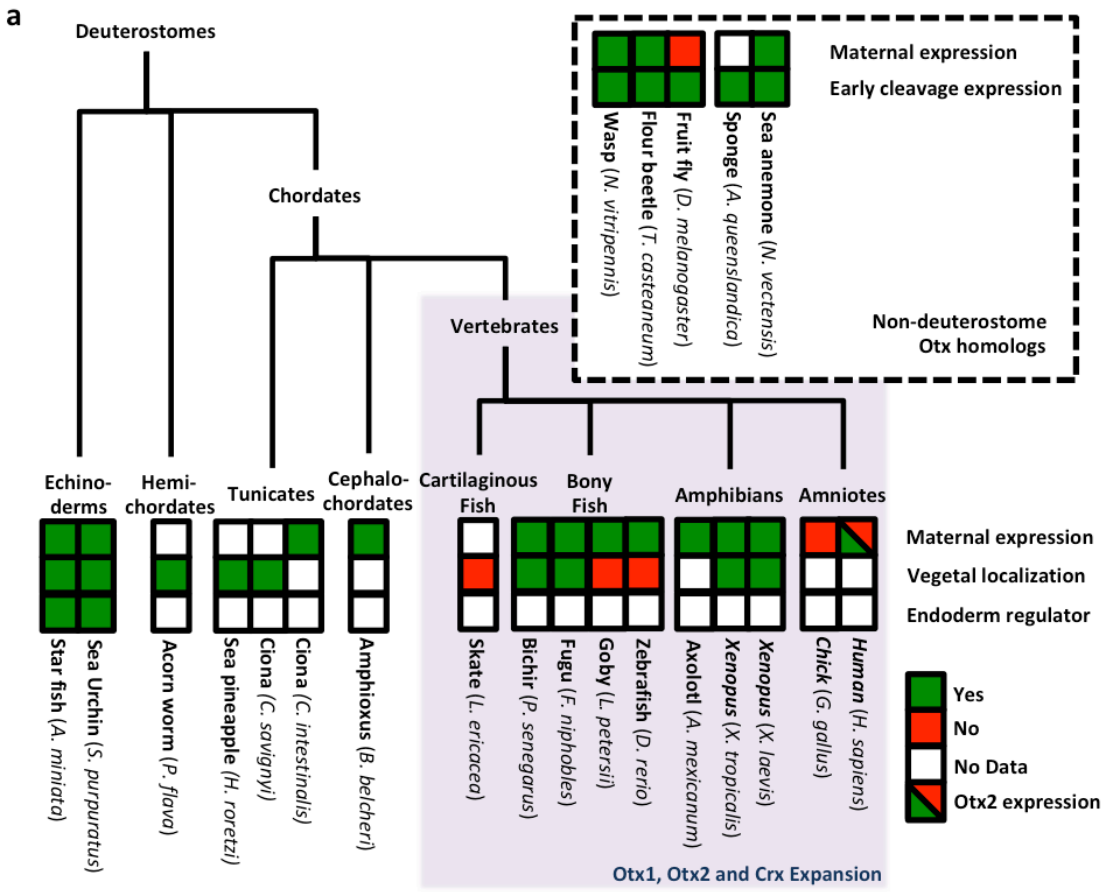

Figure 2.3. Metazoan expression of otx1 and otx orthologs. (a) Maternal and vegetal expression of vertebrate *otx1* and invertebrate *otx1* genes across deuterostome embryos from published RT-qPCR, RNA-seq and *in situ* hybridization datasets. Inset shows maternal and early embryonic expression of *otx* orthologs in protostomes, cnidarians and sponges from RNA-seq and *in situ* hybridization datasets. GEO datasets and citations used to generate this analysis can be found in the Materials and Methods and main text, respectively.

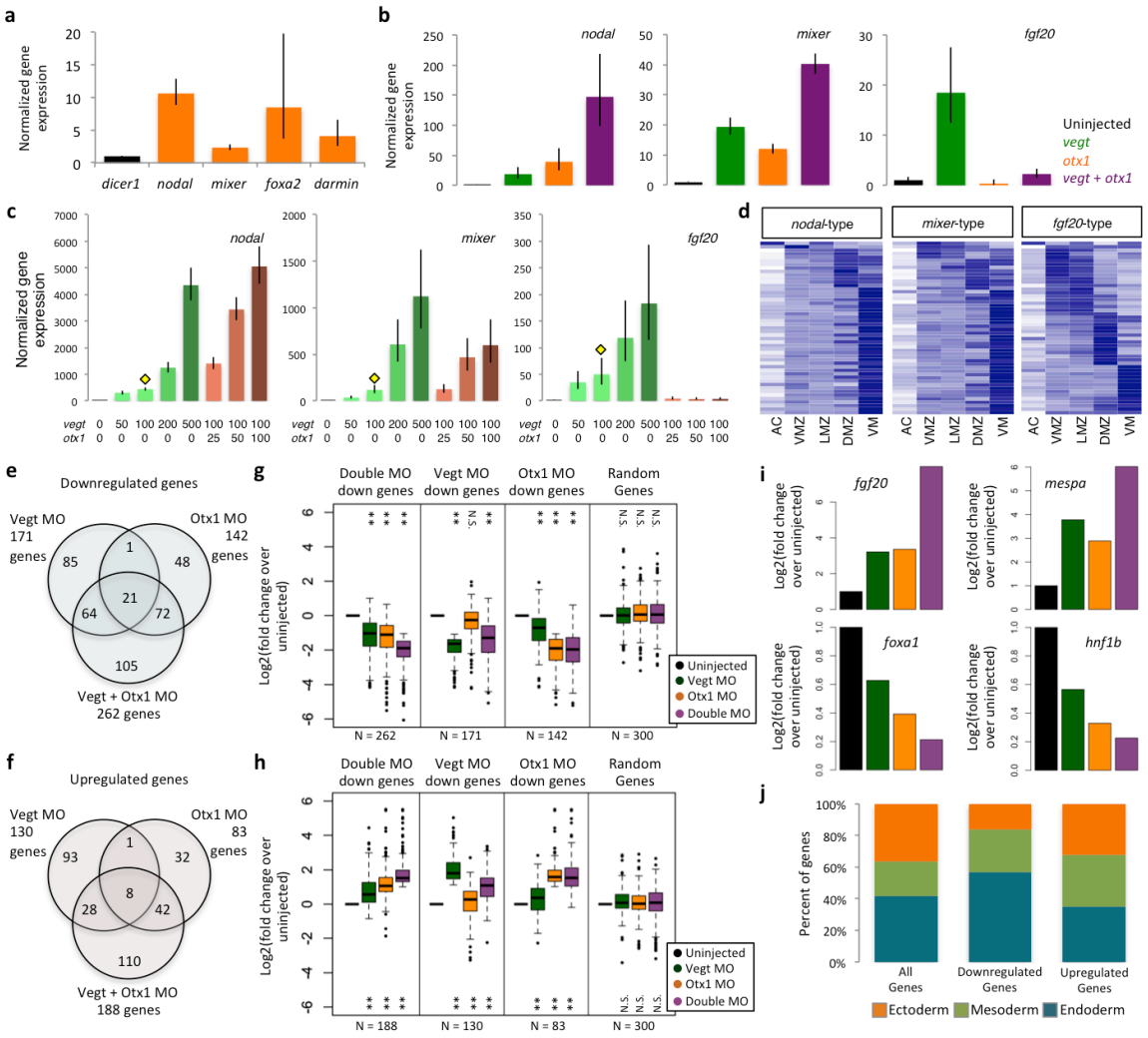

Figure 2.4. *Vegt and Otx1 are a dual function transcription factors in the endoderm.* (a) Ectopic expression of *otx1* in animal caps showing induction of endodermal genes by RT-qPCR. (b) Single and combined ectopic expression of *otx1* and *vegt* showing synergistic and antagonistic co-regulation by RT-qPCR. (c) Combinatorial ectopic expression of *vegt* and *otx1* in the animal cap using dosage titration to assay for similarly co-regulated genes by RT-qPCR and subsequently, RNA-seq. Yellow diamond indicates the sub-threshold concentration of *vegt*, with which titrating doses of *otx1* mRNA is co-injected. (d) Spatial expression pattern in the gastrula stage of genes that are positively co-regulated (*nodal*-type and *mixer*-type) or negatively co-regulated (*fgf20*-type) by *vegt* and *otx1*. Venn diagram of genes that are downregulated (e) or upregulated (f) in morpholino knock-down of *otx1, vegt* or both. Expression pattern of downregulated (g) or upregulated (h) genes in the RNA-seq datasets in transcripts per million normalized to uninjected control. (i) Expression of mesodermal (*fgf20*, *mespa*) genes and endodermal (*foxa1*, *hnf1b*) genes in the morpholino RNA-seq experiment. (j) Spatial expression of genes that are downregulated and upregulated in the double morpholino experiment. Abbreviations: AC = animal cap, VMZ = ventral marginal zone, LMZ = lateral marginal zone, DMZ = dorsal marginal zone and VM = vegetal mass; ** p-value < 10^{-3}, N.S. = not significant

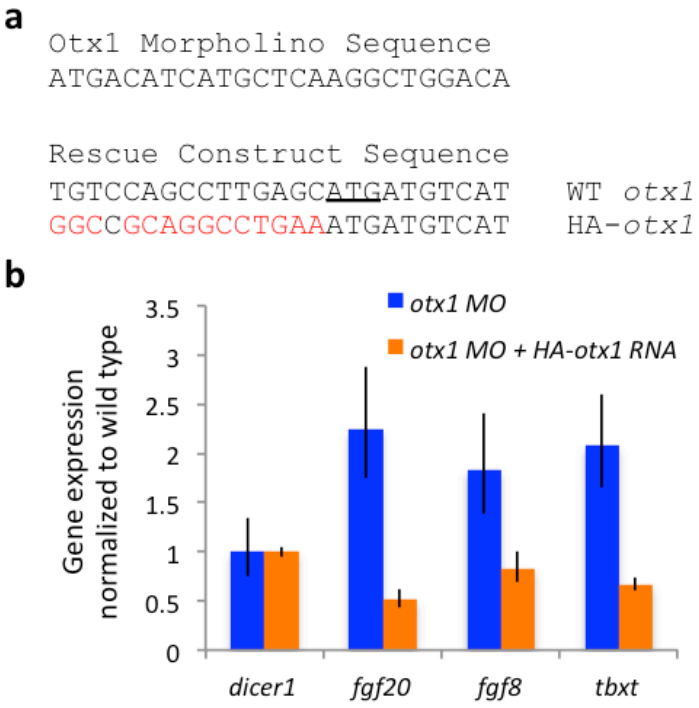

Figure 2.5. Morpholino knock-down of Otx1. (a) Sequence of the *X. tropicalis otx1* initiator ATG region, along with the translation blocking morpholino and the HA-*otx1* rescue construct. Underlined on the WT *otx1* sequence is the translation start site. In red in the HA-*otx1* sequence are 14/25 mismatching bases between WT and rescue sequence that is targeted by the morpholino. (b) RT-qPCR at mid-gastrula of morpholino injected compared to uninjected vegetal masses to test morpholino specificity.

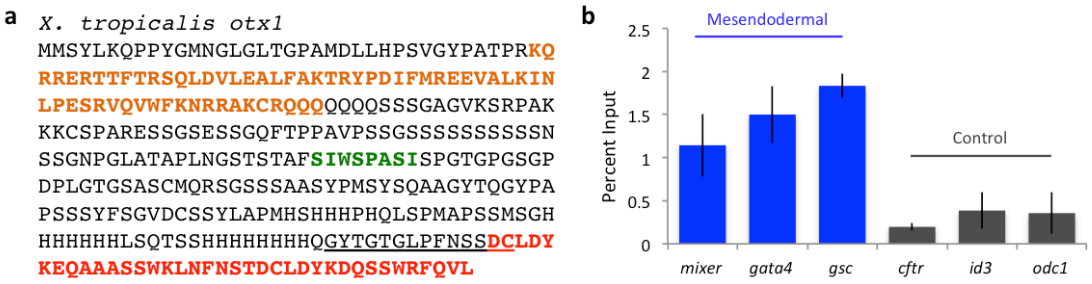

Figure 2.6. ChIP-qPCR validation of Otx1 peptide antibody. (a) *X. tropicalis* Otx1 protein sequence where the underlined sequence is the target of the Otx1 polyclonal peptide antibody. In colors are conserved regions of the protein: the homeodomain (orange), the WSP domain (green) and the Otx-tail (red). (b) ChIP-qPCR using the antibody assaying for Otx1 binding in the promoter regions of mesendodermal versus control genes.

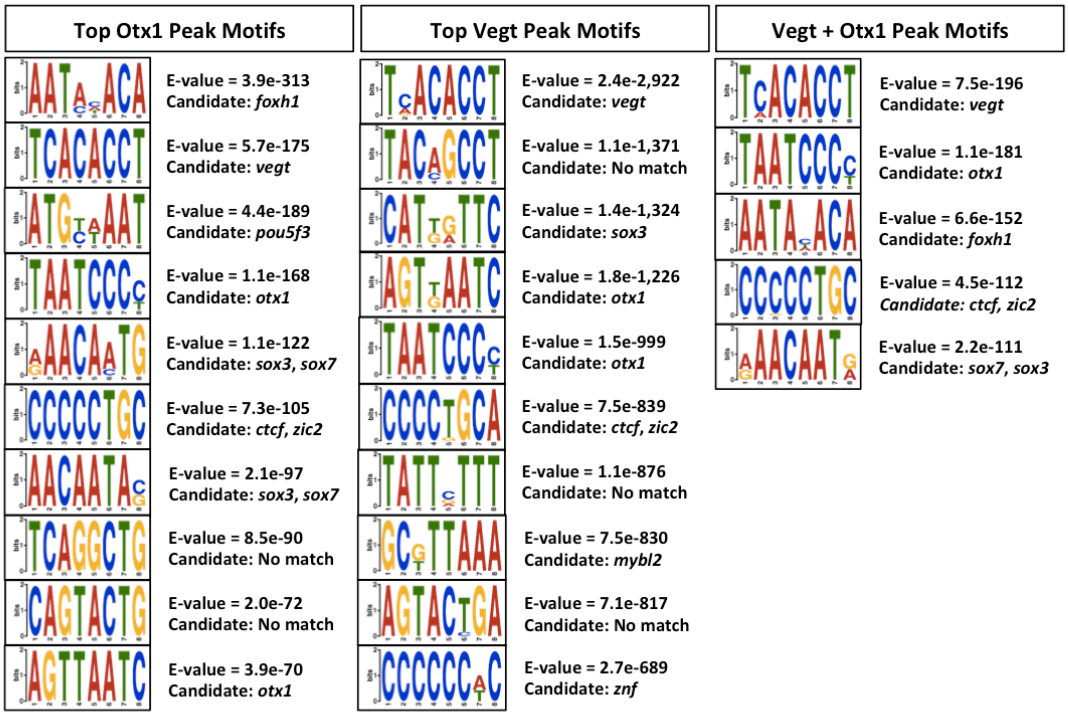

Figure 2.7. Motif analysis of Vegt and Otx1 ChIP-seq datasets. Top motifs ranked by enrichment found in the Otx1 ChIP-seq, Vegt ChIP-seq and Vegt-Otx1 co-bound regions. Candidates were identified based on whether the transcription factor (1) binds to a DNA

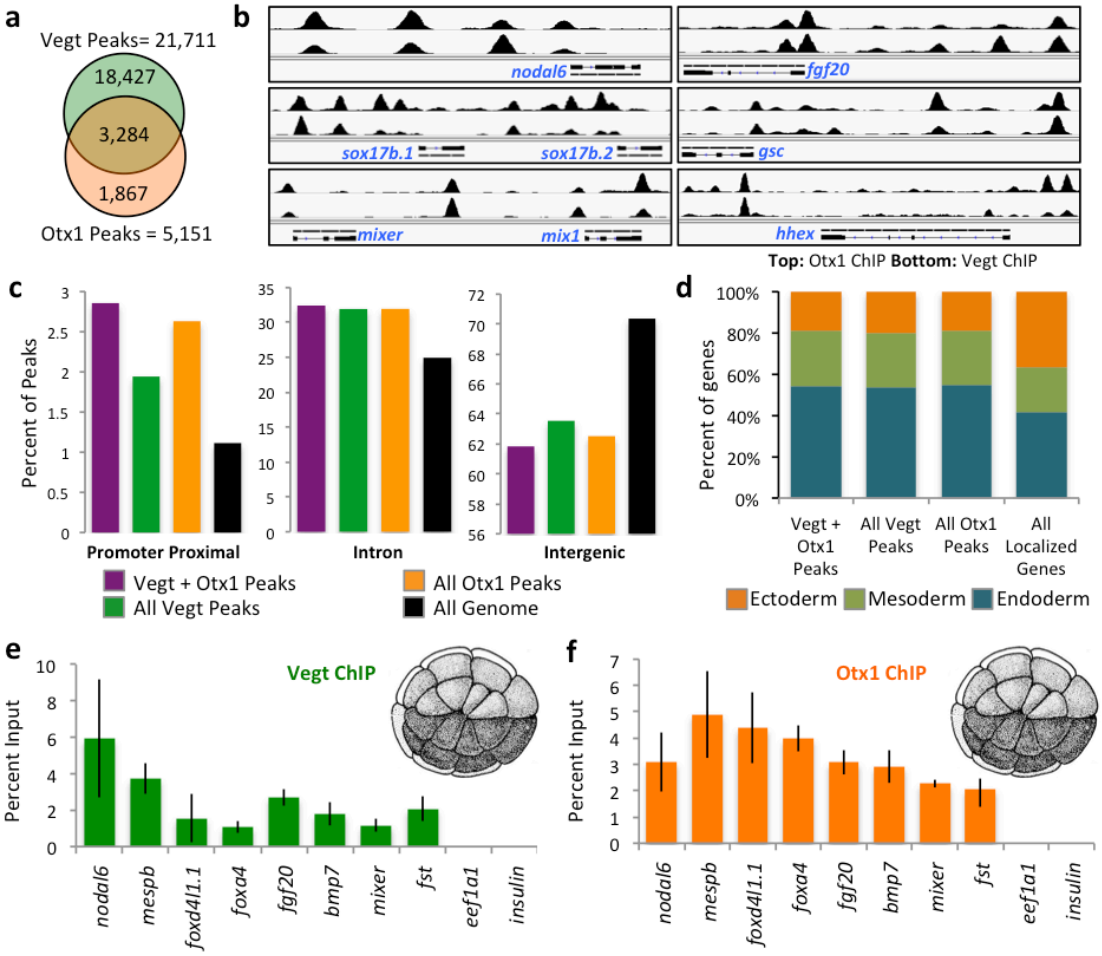

Figure 2.8. *Vegt and Otx1 bind to the chromatin near mesendodermal genes prior to ZGA.* (a) Venn diagram of Otx1 and Vegt peak overlaps. (b) IGV track of Otx1 and Vegt ChIP-seq near endodermal (*nodal6*, *sox17b.1*, *sox17b.2*, *mixer*, *mix1*), mesodermal (*fgf20*) and mesendodermal (*gsc*, *hhex*) genes. (c) Location of Vegt and Otx1 binding. (d) Expression of genes bound by Vegt and Otx1. Chromatin binding of Vegt (e) and Otx1 (f) near genes involved in germ layer formation identified using ChIP-qPCR at 32-cell stage embryos. *eef1a1* and *insulin* promoters were used as negative control.
motif that matches the enriched motif and (2) is expressed maternally along with Vegt and Otx1.

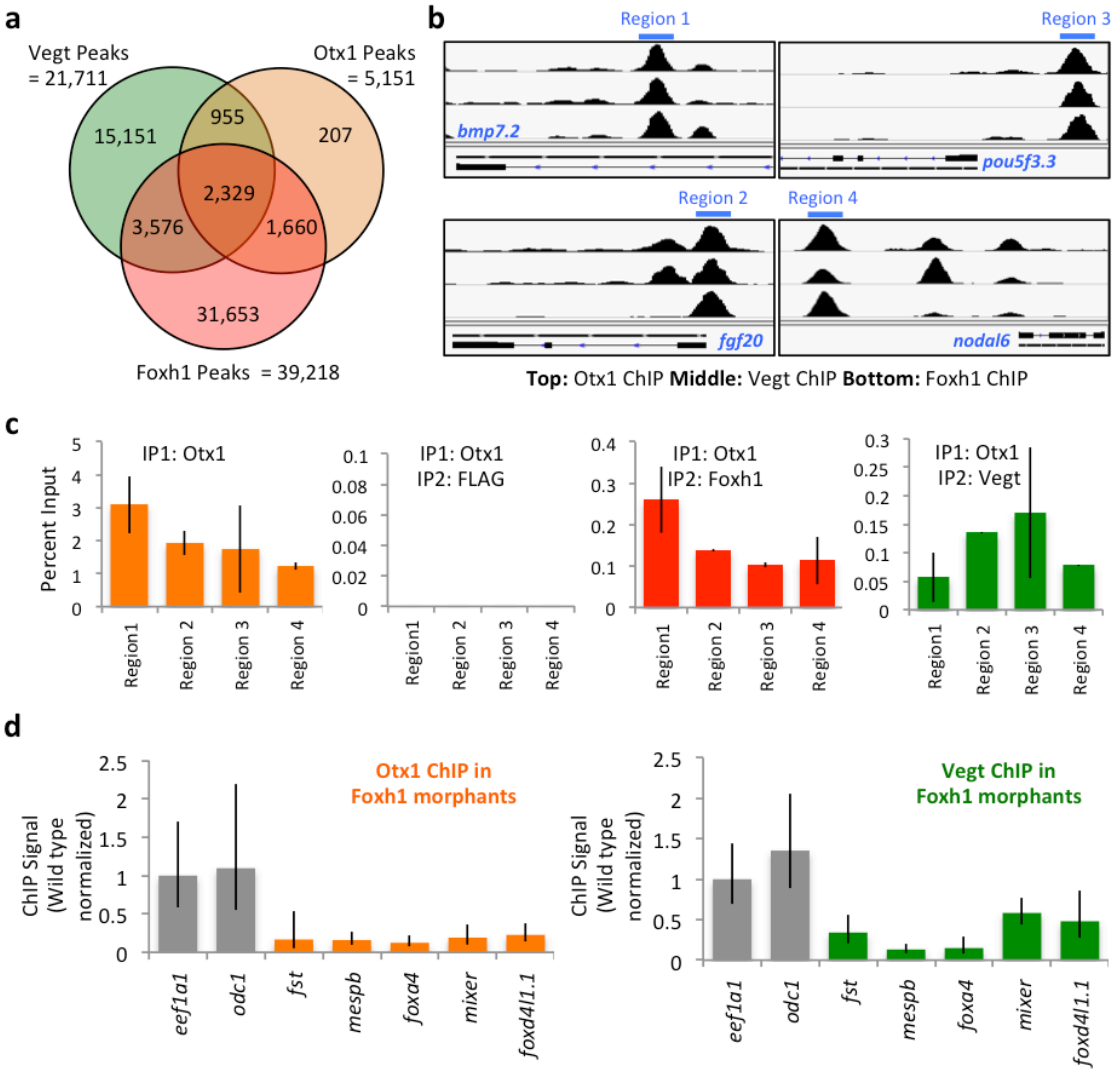

Figure 2.9. *Ubiquitous and endoderm-specific factors form an assembly in the chromatin.* (a) Overlap of Vegt, Otx1 and Foxh1 peaks. (b) Genome browser view of Otx1, Vegt and Foxh1 binding, highlighting four regions of peak overlaps. (c) Sequential chromatin immunoprecipitation using anti-otx1 followed by anti-FLAG, anti-vegt or anti-Foxh1 antibody in regions of peak overlap. (d) Foxh1 MO knock-down followed by Otx1 or Vegt ChIP-qPCR performed on stage 8 embryos. *eef1a1* and *odc1* were used as negative control.

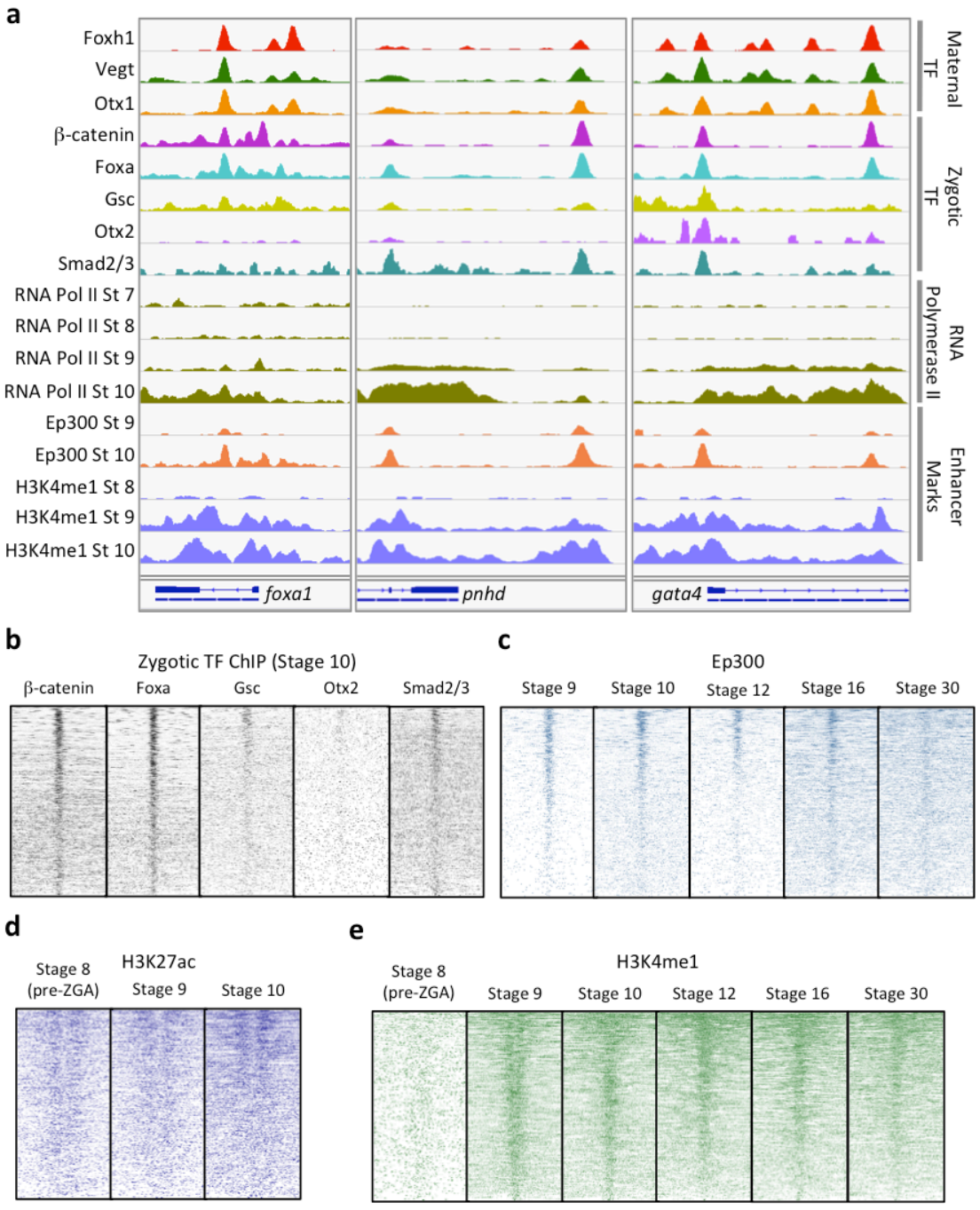

Figure 2.10. *Combinatorial Foxh1, Vegt and Otx1 binding pre-marks zygotic cis-regulatory modules.* (a) IGV browser of maternal Foxh1, Vegt and Otx1 binding overlaid with mesendodermal zygotic transcription factor binding (β-catenin, Foxa, Gsc, Otx2 and Smad2/3), RNA polymerase II and enhancer marks (H3K4me1 and p300) near the genes *foxa1*, *pnhd* and *gata4*. (b) Later gastrula stage zygotic TF binding of Ctnnb1/β-catenin, Foxa, Gsc, Otx2 and Smad2/3 in combinatorially bound regions by maternal TFs. Heatmap of active enhancer marks Ep300 (c) and H3K27ac (d), and general enhancer mark H3K4me1 (e) in combinatorially bound regions by maternal TFs.

Animally-localized maternal genes			
abca3	*herc2*	*mprip*	*taok2*
acsl4	*kdm4a*	*msi1*	*tbc1d14*
adprh.2	*kiaa0232*	*mvk*	*tbxa2r*
apba3	*kiaa0922*	*nt5m*	*tmem164*
appl2	*kif13a-like*	*pgrmc1*	*tmem55a*
ash1l	*lipe*	*pnkd*	*tnfaip1*
atxn7	*loc100036640*	*prr11*	*tomm40l*
b3galt4	*loc100216088*	*prrg4*	*tp53i11*
c2orf49	*loc100487499*	*psap*	*trim32*
cadm1	*loc100487666*	*ptk2*	*ttbk2*
casc5	*loc100489276*	*ptprq*	*ubxn11*
cbfa2t2	*loc100489717*	*rab6a*	*usp54*
cd74	*loc100490228*	*rbm46*	*wdr45b*
dand5	*loc100490651*	*rmdn3*	*Xetrov90000828m*
dfnb31	*loc100490846*	*scaf4*	*Xetrov90004801m*
dfnb59	*loc100494768*	*sfxn1*	*Xetrov90015306m*
dgkq	*loc100494894*	*sh3rf1*	*Xetrov90019765m*
dnaja4.2	*loc100497699*	*sidt2*	*Xetrov90023622m*
dnal4	*loc101731566***	*slc26a6.2*	*Xetrov90026807m*
fgfr4	*loc101731627*	*slc36a4*	*Xetrov90027278m*
fnip2	*loc101734368*	*snx7*	*Xetrov90027900m*
*foxi2***	*lrrc71*	*sox9***	*Xetrov90029512m*
gigyf1	*lrrfip2*	*spata5*	*xkrx*
gpc4	*map7d3*	*sptlc1*	*znf609*
*grhl1***	*mettl9*	*stk40*	*zrsr2*
heatr5a			

Table 2.1. Animally-localized maternal genes identified through transcriptomic analysis of the 8-cell stage embryo. ** indicates 4 genes encoding transcription factors.

Vegetally-localized maternal genes					
acsbg2	*fam43a*	*loc100494186*	*nlgn1*	*smarcc2***	*Xetrov90011196m*
acsl1	*fgfr2*	*loc100494328*	*notum*	*snai1***	*Xetrov90012204m*
aktip	*fzd1*	*loc100494519*	*ormdl2*	*soga1*	*Xetrov90012314m*
anln.1	*gab2*	*loc100495031*	*osbpl11*	*sox7***	*Xetrov90012939m*
arl8b	*galnt10*	*loc100495295*	*otx1***	*spire1*	*Xetrov90013148m*
*atf6***	*gas2l3*	*loc100496161*	*pacsin3*	*sptssa*	*Xetrov90014808m*
bicc1	*gdf1*	*loc100496367*	*parva*	*stard3*	*Xetrov90014863m*
bloc1s5	*gpbp1***	*loc100497147*	*pbx1***	*stx12*	*Xetrov90015270m*
bves	*greb1l*	*loc100498236*	*pcnt.1*	*styk1*	*Xetrov90015386m*
camk2d	*grip2*	*loc100498538*	*pcsk4*	*suclg2*	*Xetrov90018595m*
car15	*hba-l2*	*loc101730709*	*pcsk6*	*sufu*	*Xetrov90021534m*
cbx7	*hddc3*	*loc101731661*	*pi4kb*	*sulf1*	*Xetrov90022456m*
cdan1	*hexb*	*loc101732537*	*pld2*	*sumf2*	*Xetrov90022545m*
cdk5r2	*hook2*	*loc101732683*	*plk3*	*sybu*	*Xetrov90022546m*
cep57	*hpca*	*loc101732716*	*ppp1r13b*	*syngr1*	*Xetrov90022562m*
chek1	*hps6*	*loc101732837***	*ppp1r3b*	*tbc1d1*	*Xetrov90022602m*
cited2	*hsbp1*	*loc101732899*	*ppp1r9a*	*tefm*	*Xetrov90025751m*
cldn4	*igf3*	*loc101733008*	*prpsap2*	*thnsl2*	*Xetrov90025763m*
cnppd1	*kiaa1522*	*loc101733155*	*ptp4a3*	*timp3*	*Xetrov90026290m*
cntd2	*kiaa1715*	*loc101733578*	*ptplad1*	*tmprss4*	*Xetrov90026291m*
cpped1	*kif13b*	*loc101734093*	*qrich1.2*	*trib1*	*Xetrov90027567m*
cpsf4l	*kif14*	*loc101734920*	*rab11fip4l*	*trim36*	*Xetrov90028148m*
cyp26a1	*klhl11*	*loc101735136*	*ralgps2*	*usp40*	*Xetrov90028257m*
daam1	*klhl36*	*loc101735154*	*rbpms2*	*vegt***	*Xetrov90028259m*
dazl	*krt8*	*loc733735*	*rdh10*	*velo1*	*Xetrov90028649m*
dcc-like	*llgl1*	*loc779566*	*rhcg*	*wnk2*	*Xetrov90029275m*
ddx25	*loc100036665*	*lrp1*	*rnf180*	*wnt11b*	*Xetrov90029846m*
dennd5a	*loc100145165*	*lsm4*	*rnf38*	*Xetrov90000569m*	*Xetrov90029967m*
dgat2	*loc100487364*	*lurap1*	*rnf41*	*Xetrov90000785m*	*Xetrov90030080m*
dnd1	*loc100487727*	*mark4*	*rragc*	*Xetrov90005058m*	*Xetrov90030362m*
dynlt1	*loc100488945*	*mical3*	*rtn3*	*Xetrov90005215m*	*zmym3*
emc7	*loc100490959*	*mief2*	*sbf1*	*Xetrov90005273m*	*zp3*
ero1l	*loc100492173*	*mogat1.2*	*sipa1l1*	*Xetrov90008209m*	*zswim4*
exd1	*loc100493477*	*mov10*	*slc15a4*	*Xetrov90009737m*	
fam160a1	*loc100493856*	*nanos1*	*smagp*	*Xetrov90009917m*	

Table 2.2. Vegetally-localized maternal genes identified through transcriptomic analysis of the 8-cell stage embryo. ** indicates 9 genes encoding transcription factors.

nodal-type	mixer-type	fgf20-type
nodal	*mixer*	*fgf20*
lncrna_single_sw_00181079	*lefty*	*cdx2*
insm2	*cmtm8-like*	*lncrna_single_sw_00123353*
menf.1-like.2	*mespa*	*sfrp1*
Xetrov90020481m	*lhx1*	*loc733709*
loc101733237	*cass4*	*lncrna_single_sw_00221977*
uts2r-like	*foxa4*	*chrnb2*
Xetrov90006308m	*loc100488909*	*MGC107849*
Xetrov90022751m	*loc101733080*	*wnt11b*
loc100493406-like	*pnhd*	*Xetrov90023892m*
Xetrov90012033m	*sept5*	*trabd2a*
tbx3	*flrt3*	*slc12a3.2*
lncrna_single_sw_00173957	*cer1*	*nppa*
tnmd	*loc100495259*	*plod2-like*
loc100487643-like	*nodal1*	*lncrna_single_n_00078502*
dkk3	*Xetrov90013120m*	*fhl3*
Xetrov90028804m	*loc100496651*	*il2rg*
lncrna_single_sw_00179312	*ventx3.1*	*tmprss9*
kif21b-like	*frzb*	*loc100485269*
plekho1-like	*ptchd4-like*	*ca14*
lncrna-candidate-sw-20901	*Xetrov90009422m*	*lncrna_single_sw_00267705*
Xetrov90000638m	*snai1*	*esr10*
Xetrov90027386m	*lncrna-candidate-sw-50738*	*tgm4*
prr5l-like	*mespb*	*wnt11-like.1*
lncrna-candidate-sw-48848	*Xetrov90028098m*	*tmem150b*
Xetrov90030008m	*epha4*	*lncrna_single_n_00043132*
nefl	*loc101733442*	*Xetrov90029542m*
rap2b	*pcdh8.2*	*hoxd1*
hes1	*lncrna_single_sw_00026912*	*cpe*
Xetrov90009588m	*Xetrov90012981m*	*lncrna-candidate-sw-15653*
lncrna_single_sw_00237010	*mett*	*msx1*
has2	*bix1.1*	*apcs*
loc101734923	*ventx3.2*	*loc101733373*
Xetrov90017989m	*npb*	*foxj1*
mmp28	*myf5*	*lmx1b.2*
cldn5	*zswim4*	*c16orf52*
loc101732407	*Xetrov90025899m*	*cdx4*
Xetrov90014399m	*rspo2*	*fgf4*
Xetrov90005842m	*pcdh7*	*Xetrov90020611m*
Xetrov90027388m	*bix1.2*	*fras1*
Xetrov90028555m	*loc100488893*	*kcnk6*
lncrna_single_sw_00242363	*loc100489931*	*loc100488066*
Xetrov90029136m	*loc101732407*	*lncrna_single_sw_00180013*
lncrna_single_sw_00139523	*rgs16*	*lncrna_single_n_00173332*
Xetrov90027792m	*lncrna-candidate-n-6633*	*aldh1a2*
lncrna_single_sw_00087289	*Xetrov90029139m*	*loc101732886*
tesc	*lncrna-candidate-sw-20589*	*lncrna_single_n_00106984*
loc100485791	*Xetrov90017301m*	*Xetrov90000009m*
lncrna-candidate-sw-41202	*mix1*	*Xetrov90003151m*
lncrna_single_sw_00255127	*psmb2*	*lncrna_single_sw_00156986*

Table 2.3 List of most highly correlated genes to the nodal-type, mixer-type and fgf20-type co-regulation

Vegt activated	Vegt inhibited
admp2 *anpep* *ap5s1* *bloc1s1* *cd164* *cenpe* *cer1* *cyr61* *dhrs3* *dph1* *echdc2* *gtf3a* *hcar3* *loc100145507* *loc100487915* *loc100488176-* *loc100489738* *loc100490316* *loc100496651* *loc101731067* *loc101734157* *loc101735069* *loc548390* *mfsd12* *mmp1* *nefl* *nodal5.2* *nodal6* *pcyt1a* *rasd1* *rbbp7* *rps6ka1* *rspo2* *rspo3* *sfrp2* *sucla2* *tnks-like* *ventx3.2* *wnt11.1* *Xetrov90012451m* *Xetrov90013145m* *Xetrov90019859m* *Xetrov90028772m* *Xetrov90029865m* *Xetrov90030342m*	*angpt4* *arhgef3* *arl4c* *bmp7.2* *cdx4* *cyp26a1* *dll1* *fgf20* *fgf8* *foxb1* *foxd3* *foxd4l1.1* *fzd10* *grhl3* *hand2* *hes6.1* *lefty* *loc100124771* *loc100487186* *loc100493838* *loc100494294* *loc100497175* *loc100497952.2* *mespa* *mespb* *p2ry2* *sema3f* *sox2* *tfap2a* *upf1* *upk2* *wnt5b* *Xetrov90000829m* *Xetrov90013953m* *Xetrov90024231m* *zeb2* *zic1*

Table 2.4. Vegt direct targets identified using Vegt knock-down data and ChIP-seq.

Otx1 activated	Otx1 inhibited
cass4 *cyb561* *eif3i* *esr10* *foxa1* *gata4* *gatm* *kirrel2* *loc101731753* *loc101732727* *loc101733122* *ly6g6c* *nrp2* *rasd1* *sgk1*	*acrc* *chac1* *dlx6* *fgf20* *hes2* *loc100496651* *Xetrov90026650m* *Xetrov90030340m*

Table 2.5. Otx1 direct targets identified using Otx1 knock-down data and ChIP-seq.

Vegt and Otx1 activated	Vegt and Otx1 inhibited
abcb5 *ca14* *coro7* *fst* *g0s2* *gfpt1* *hal.1* *kcnk6* *loc100127584* *loc100494680* *loc100496394* *msx1* *pnhd* *ptpn13* *rasd1* *ror2* *rpl37a* *slc38a3* *tll1-like.1* *ventx3.1* *wnt8a* *znf608*	*alkbh4* *arhgap29* *fgf20* *loc100488934* *loc101734031* *nodal3.2* *rttn*

Table 2.6. Vegt and Otx1 combinatorial direct targets identified using knock-down data and ChIP-seq.

References

1.Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Developmental Cell* **42,** 316–332 (2017).

2.Minakhina, S. & Steward, R. Axes formation and RNA localization. *Current Opinion in Genetics & Development* **15,** 416–421 (2005).

3.Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What How, and Why?. *Molecular Cell* **49,** 825–837 (2013).

4.Kim, H. S. *et al.*. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells.. *Nature* **556,** 510–514 (2018).

5.Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. & Klein, P. S. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2.. *Dev Cell* **19,** 220–31 (2010).

6.Akkers, R. C. *et al.*. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos.. *Dev Cell* **17,** 425–34 (2009).

7.Gupta, R., Wills, A., Ucar, D. & Baker, J. Developmental enhancers are marked independently of zygotic Nodal signals in Xenopus.. *Dev Biol* **395,** 38–49 (2014).

8.Hontelez, S. *et al.*. Embryonic transcription is controlled by maternally defined chromatin state.. *Nat Commun* **6,** 10148 (2015).

9.Charney, R. M. *et al.*. Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program.. *Dev Cell* **40,** 595–607.e4 (2017).

10.Vastenhouw, N. L. *et al.*. Chromatin signature of embryonic pluripotency is established during genome activation.. *Nature* **464,** 922–6 (2010).

11.Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition.. *Elife* **3,** (2014).

12.Stennard, F., Carnac, G. & Gurdon, J. B. The Xenopus T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation.. *Development* **122,** 4179–88 (1996).

13.Lustig, K. D., Kroll, K. L., Sun, E. E. & Kirschner, M. W. Expression cloning of a Xenopus T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation.. *Development* **122,** 4001–12 (1996).

14.Zhang, J. & King, M. L. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning.. *Development* **122,** 4119–29 (1996).

15.Horb, M. E. & Thomsen, G. H. A vegetally localized T-box transcription factor in Xenopus eggs specifies mesoderm and endoderm and is essential for embryonic mesoderm formation.. *Development* **124,** 1689–98 (1997).

16.Zhang, J. *et al.*. The role of maternal VegT in establishing the primary germ layers in Xenopus embryos.. *Cell* **94,** 515–24 (1998).

17.Clements, D., Friday, R. V. & Woodland, H. R. Mode of action of VegT in mesoderm and endoderm formation.. *Development* **126,** 4903–11 (1999).

18.Xanthos, J. B., Kofron, M., Wylie, C. & Heasman, J. Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis.. *Development* **128,** 167–80 (2001).

19.Blitz, I. L. *et al.*. A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo.. *Dev Biol* **426,** 409–417 (2017).

20.De Domenico, E., Owens, N. D., Grant, I. M., Gomes-Faria, R. & Gilchrist, M. J. Molecular asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere transcript sequencing.. *Dev Biol* **408,** 252–68 (2015).

21.Owens, D. A. *et al.*. High-throughput analysis reveals novel maternal germline RNAs crucial for primordial germ cell preservation and proper migration.. *Development* **144,** 292–304 (2017).

22.Zhang, C., Basta, T., Fawcett, S. R. & Klymkowsky, M. W. SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in Xenopus.. *Dev Biol* **278,** 526–41 (2005).

23.Pannese, M., Cagliani, R., Pardini, C. L. & Boncinelli, E. Xotx1 maternal transcripts are vegetally localized in Xenopus laevis oocytes.. *Mech Dev* **90,** 111–4 (2000).

24.Bauer, D. V., Huang, S. & Moody, S. A. The cleavage stage origin of Spemann's Organizer: analysis of the movements of blastomere clones before and during gastrulation in Xenopus.. *Development* **120,** 1179–89 (1994).

25.Blitz, I. L. & Cho, K. W. Anterior neurectoderm is progressively induced during gastrulation: the role of the Xenopus homeobox gene orthodenticle.. *Development* **121,** 993–1004 (1995).

26.Pannese, M. *et al.*. The Xenopus homologue of Otx2 is a maternal homeobox gene that demarcates and specifies anterior body regions.. *Development* **121,** 707–20 (1995).

27.Vignali, R. *et al.*. Xotx5b, a new member of the Otx gene family, may be involved in anterior and eye development in Xenopus laevis.. *Mech Dev* **96,** 3–13 (2000).

28.Hinman, V. F., Nguyen, A. T. & Davidson, E. H. Expression and function of a starfish Otx ortholog, AmOtx: a conserved role for Otx proteins in endoderm development that predates divergence of the eleutherozoa.. *Mech Dev* **120,** 1165–76 (2003).

29.Peter, I. S. & Davidson, E. H. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Developmental Biology* **340,** 188–199 (2010).

30.Mitsunaga-Nakatsubo, K. *et al.*. Differential expression of sea urchin Otx isoform (hpOtxE and HpOtxL) mRNAs during early development.. *Int J Dev Biol* **42,** 645–51 (1998).

31.Harada, Y. *et al.*. Developmental expression of the hemichordate otx ortholog.. *Mech Dev* **91,** 337–9 (2000).

32.Williams, N. A. & Holland, P. W. Gene and domain duplication in the chordate Otx gene family: insights from amphioxus Otx.. *Mol Biol Evol* **15,** 600–7 (1998).

33.Satou, Y., Imai, K. S. & Satoh, N. Early embryonic expression of a LIM-homeobox gene Cs-lhx3 is downstream of beta-catenin and responsible for the endoderm differentiation in Ciona savignyi embryos.. *Development* **128,** 3559–70 (2001).

34.Matsuoka, T., Ikeda, T., Fujimaki, K. & Satou, Y. Transcriptome dynamics in early embryos of the ascidian, Ciona intestinalis.. *Dev Biol* **384,** 375–85 (2013).

35.Yang, K. Y. *et al.*. Transcriptome analysis of different developmental stages of amphioxus reveals dynamic changes of distinct classes of genes during development.. *Sci Rep* **6,** 23195 (2016).

36.Suda, Y. *et al.*. Evolution of Otx paralogue usages in early patterning of the vertebrate head.. *Dev Biol* **325,** 282–95 (2009).

37.Mori, H., Miyazaki, Y., Morita, T., Nitta, H. & Mishina, M. Different spatio-temporal expressions of three otx homeoprotein transcripts during zebrafish embryogenesis.. *Brain Res Mol Brain Res* **27,** 221–31 (1994).

38.Aanes, H. *et al.*. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition.. *Genome Res* **21,** 1328–38 (2011).

39.Kamimoto, M. *et al.*. Unequal distribution of otx1 mRnas among cleavage stage blastomeres in the teleost, Leucopsarion petersii (shiro-uo).. *Int J Dev Biol* **47,** 77–80 (2003).

40.Jiang, P. *et al.*. Analysis of embryonic development in the unsequenced axolotl: Waves of transcriptomic upheaval and stability.. *Dev Biol* **426,** 143–154 (2017).

41.Hwang, Y. S. *et al.*. The transcriptome of early chicken embryos reveals signaling pathways governing rapid asymmetric cellularization and lineage segregation.. *Development* **145,** (2018).

42.Wu, J. *et al.*. Chromatin analysis in human early development reveals epigenetic transition during ZGA.. *Nature* **557,** 256–260 (2018).

43.Schröder, R. The genes orthodenticle and hunchback substitute for bicoid in the beetle Tribolium.. *Nature* **422,** 621–5 (2003).

44.Lynch, J. A., Brent, A. E., Leaf, D. S., Pultz, M. A. & Desplan, C. Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp Nasonia.. *Nature* **439,** 728–32 (2006).

45.Levin, M. *et al.*. The mid-developmental transition and the evolution of animal body plans.. *Nature* **531,** 637–641 (2016).

46.Helm, R. R., Siebert, S., Tulin, S., Smith, J. & Dunn, C. W. Characterization of differential transcript abundance through time during Nematostella vectensis development.. *BMC Genomics* **14,** 266 (2013).

47.Andreazzoli, M., Pannese, M. & Boncinelli, E. Activating and repressing signals in head development: the role of Xotx1 and Xotx2.. *Development* **124,** 1733–43 (1997).

48.Puelles, E. *et al.*. Otx2 regulates the extent, identity and fate of neuronal progenitor domains in the ventral midbrain.. *Development* **131,** 2037–48 (2004).

49.Lea, R., Papalopulu, N., Amaya, E. & Dorey, K. Temporal and spatial expression of FGF ligands and receptors during Xenopus development.. *Dev Dyn* **238,** 1467–79 (2009).

50.Rana, A. A., Collart, C., Gilchrist, M. J. & Smith, J. C. Defining synphenotype groups in Xenopus tropicalis by use of antisense morpholino oligonucleotides.. *PLoS Genet* **2,** e193 (2006).

51.Owens, N. D. L. *et al.*. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development.. *Cell Rep* **14,** 632–647 (2016).

52.Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq.. *Nat Methods* **5,** 621–8 (2008).

53.Sudou, N., Yamamoto, S., Ogino, H. & Taira, M. Dynamic in vivo binding of transcription factors to cis-regulatory modules of cer and gsc in the stepwise formation of the Spemann-Mangold organizer.. *Development* **139,** 1651–61 (2012).

54.Driever, W., Thoma, G. & Nüsslein-Volhard, C. Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen.. *Nature* **340,** 363–7 (1989).

55.Conlon, F. L., Fairclough, L., Price, B. M., Casey, E. S. & Smith, J. C. Determinants of T box protein specificity.. *Development* **128,** 3749–58 (2001).

56.Hilton, E., Rex, M. & Old, R. VegT activation of the early zygotic gene Xnr5 requires lifting of Tcf-mediated repression in the Xenopus blastula.. *Mech Dev* **120,** 1127–38 (2003).

57.Houston, D. W. & Wylie, C. Maternal Xenopus Zic2 negatively regulates Nodal-related gene expression during anteroposterior patterning.. *Development* **132,** 4845–55 (2005).

58.Chiu, W. T. *et al.*. Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program.. *Development* **141,** 4537–47 (2014).

59.Yasuoka, Y. *et al.*. Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification.. *Nat Commun* **5,** 4322 (2014).

60.Nakamura, Y., de, P. A. E., Veenstra, G. J. & Hoppler, S. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules.. *Development* **143,** 1914–25 (2016).

61.Liu, G., Wang, W., Hu, S., Wang, X. & Zhang, Y. Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. *Genome Research* **28,** 998–1007 (2018).

62.Almouzni, G. & Wolffe, A. P. Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis.. *EMBO J* **14,** 1752–65 (1995).

63.Amodeo, A. A., Jukam, D., Straight, A. F. & Skotheim, J. M. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition.. *Proc Natl Acad Sci U S A* **112,** E1086–95 (2015).

64.Joseph, S. R. *et al.*. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos.. *Elife* **6,** (2017).

65.Sakabe, N. J. *et al.*. Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function.. *Hum Mol Genet* **21,** 2194–204 (2012).

66.He, M., Wen, L., Campbell, C. E., Wu, J. Y. & Rao, Y. Transcription repression by Xenopus ET and its human ortholog TBX3, a gene involved in ulnar-mammary syndrome.. *Proc Natl Acad Sci U S A* **96,** 10212–7 (1999).

67.Kawamura, A., Koshida, S. & Takada, S. Activator-to-repressor conversion of T-box transcription factors by the Ripply family of Groucho/TLE-associated mediators.. *Mol Cell Biol* **28,** 3236–44 (2008).

68.Milton, A. C. & Okkema, P. G. Caenorhabditis elegans TBX-2 Directly Regulates Its Own Expression in a Negative Autoregulatory Loop.. *G3 (Bethesda)* **5,** 1177–86 (2015).

69.Formaz-Preston, A., Ryu, J. R., Svendsen, P. C. & Brook, W. J. The Tbx20 homolog Midline represses wingless in conjunction with Groucho during the maintenance of segment polarity.. *Dev Biol* **369,** 319–29 (2012).

70.Kaltenbrun, E. *et al.*. A Gro/TLE-NuRD corepressor complex facilitates Tbx20-dependent transcriptional repression.. *J Proteome Res* **12,** 5395–409 (2013).

71.Cha, S. W. *et al.*. Inhibition of FGF signaling causes expansion of the endoderm in Xenopus.. *Biochem Biophys Res Commun* **315,** 100–6 (2004).

72.Cha, S. W. *et al.*. Spatiotemporal regulation of fibroblast growth factor signal blocking for endoderm formation in Xenopus laevis.. *Exp Mol Med* **40,** 550–7 (2008).

73.Poulain, M., Fürthauer, M., Thisse, B., Thisse, C. & Lepage, T. Zebrafish endoderm formation is regulated by combinatorial Nodal, FGF and BMP signalling.. *Development* **133,** 2189–200 (2006).

74.Schohl, A. & Fagotto, F. Beta-catenin, MAPK and Smad signaling during early Xenopus development.. *Development* **129,** 37–52 (2002).

75.Satou, Y. et al.. Phosphorylation states change Otx2 activity for cell proliferation and patterning in the Xenopus embryo.. *Development* **145**, dev159640 (2018)

76.Nieuwkoop, P. D. & Faber, J. Normal Table of Xenopus Laevis (Daudin). *Copeia* **1958,** 65 (1958).

77.Khokha, M. K. *et al.*. Techniques and probes for the study of Xenopus tropicalis development.. *Dev Dyn* **225,** 499–510 (2002).

78.Gibson, D. G. *et al.*. Enzymatic assembly of DNA molecules up to several hundred kilobases.. *Nat Methods* **6,** 343–5 (2009).

79.Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.. *Anal Biochem* **162,** 156–9 (1987).

80.Picelli, S. *et al.*. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9,** 171–81 (2014).

81.Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).

82.Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).

83.Hellsten, U. *et al.*. The genome of the Western clawed frog Xenopus tropicalis.. *Science* **328,** 633–6 (2010).

84.Karimi, K. *et al.*. Xenbase: a genomic, epigenomic and transcriptomic model organism database.. *Nucleic Acids Res* **46**, D861-D868 (2018).

85.Leng, N. *et al.*. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments.. *Bioinformatics* **29,** 1035–43 (2013).

86.R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.. (2014).

87.Harland, R. M. In situ hybridization: an improved whole-mount method for Xenopus embryos.. *Methods Cell Biol* **36,** 685–95 (1991).

88.Zhang, Y. *et al.*. Model-based analysis of ChIP-Seq (MACS).. *Genome Biol* **9,** R137 (2008).

89.Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments.. *The Annals of Applied Statistics* **5,** 1752–1779 (2011).

90.Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27,** 1653–1659 (2011).

91.Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. Quantifying similarity between motifs. *Genome Biology* **8,** R24 (2007).

92.Li, H. *et al.*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

93.Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

94.Robinson, J. T. *et al.*. Integrative genomics viewer. *Nature Biotechnology* **29,** 24–26 (2011).

95.Geisberg, J. V. & Struhl, K. Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo.. *Nucleic Acids Res* **32,** e151 (2004).

96.Howe, K. *et al.*. The zebrafish reference genome sequence and its relationship to the human genome.. *Nature* **496,** 498–503 (2013).

97.Warren, W. C. *et al.*. A New Chicken Genome Assembly Provides Insight into Avian Genome Structure. *G3 (Bethesda)* **7,** 109–117 (2017).

Chapter 3

Divergent designs and activities of endodermal cis-regulatory regions during *Xenopus* zygotic gene activation

Abstract

Cis-regulatory modules (CRMs) are gene regulatory regions in the DNA which contain clusters of DNA binding motifs of transcription factors (TFs). In early *Xenopus* development, recent genomic advances have facilitated identification of TF binding sites, and consequently, CRMs. Identification of these CRMs enables critical assessment of qualitative differences in CRM functionality. In this chapter, I describe two endodermal CRMs showing divergent design and function: Type I and Type II. Type I and Type II modules contain clusters of motifs of ubiquitously expressed TFs and motifs of endodermally-active TFs, respectively. Type I modules are bound by Foxh1 along with either Vegt and/or Otx1, while Type II modules are bound by Vegt and Otx1 only prior to ZGA. Both types of modules are active during ZGA. However, post-ZGA, Type I modules retain activity, likely through the function of zygotic TFs, while Type II modules are decommissioned. My findings suggest that TFs could bind to the genome through multiple mechanisms within the same cellular context, and these mechanisms are dictated by cis-regulatory motif composition seen in Type I and Type II modules. While further functional experiments are necessary, my findings uncovered key qualities of these endodermal CRMs during ZGA.

Introduction

Maternal transcription factors (TFs) coordinate their activity during zygotic genome activation (ZGA). In Chapter 2, I showed that in *Xenopus*, ubiquitously-expressed Foxh1[1] coordinates with the vegetally-localized Vegt[2-5] and Otx1[6] to specify the endodermal germ layer (Fig 2.9). Foxh1, Vegt and Otx1 co-bind specific chromatin regions prior to ZGA during early cleavage stages. Subsequently, these cis-regulatory modules (CRMs) are marked by the general enhancer mark H3K4me1 and active enhancer marks H3K27ac and Ep300 during early blastula stages (Fig 2.10a,c-e), corresponding to the time of ZGA. In addition, these regions are bound by zygotic mesendodermal factors in during gastrula stage (Fig 2.10a,b). These markings suggest that these triple bound CRMs are functional and are active during ZGA and that the emergence of cis-regulatory regions depends on the combinatorial binding of maternal TFs.

If TFs initiate the formation of active CRMs, then the discriminatory binding of TFs to specific regions of the genome is critical in initializing this process. This begs the question of how TFs select their binding sites in our *in vivo* system. *In vitro* experiments show that the DNA binding domains of TFs have high affinity to specific short ~5-8-mer DNA patterns called consensus binding motifs. The *in vivo* cellular context contains more variables that can affect TF binding including chromatin density and epigenetic marks, in addition to consensus binding motifs. How TFs identify their target CRMs prior to ZGA is unclear. In this Chapter, I show that Vegt and Otx1 preferentially select regions of the genome based

on clusters of consensus binding motifs. Vegt and Otx1 bind to modules made up of clusters of binding motifs, which I call, Type I and Type II. Type II modules contain the Vegt (T-box) and Otx1 (Otx) along with the motif of other vegetally-active TFs. Type I modules do not contain the Vegt and Otx1 motifs. However, it contains the motifs of ubiquitously expressed factors including Foxh1, Sox3 and Pou5f3. Both modules are decorated with H3K27ac and Ep300 during blastula stages, suggesting that both types are active CRMs during ZGA. However, only Type I modules are marked with H3K4me1 after ZGA. In addition, during gastrula stages, Type I modules are more likely to be bound by zygotic TFs than Type II modules. Overall, these findings identify two key cis-regulatory module designs used during endoderm specification which show differences in motif composition, types of TF binding and epigenetic activity.

Results

Vegt and Otx1 scan the genome for their consensus motifs flanked by other maternal TF motifs

In order to elucidate the mode of chromatin association of the maternal and endodermal TFs Vegt and Otx1 *in vivo*, I used ChIP-seq datasets from Chapter 2 (Fig 2.8). I used the same peaks that were obtained using the irreproducible discovery rate (IDR), which yielded high confidence and reproducible peaks and identified the genome-wide binding sites for both TFs[7]. As *in vitro* studies highlight the significance of consensus binding motifs

in DNA binding of TFs, I first performed my search in regions containing consensus binding motifs. Through a combination of PCR-based binding site selection and DNA mobility shift assays, the *Xenopus* Vegt has been shown to associate to the half-site T-box consensus binding sequence[8]. Although the same dataset does not exist for *Xenopus* Otx1, the vertebrate Otx1 orthologs in mouse[9] and zebrafish[10] has been shown to associate with the Bicoid-type homeobox[11] motif. My previous *de novo* motif analysis on the Vegt and Otx1 binding peaks yielded motifs similar to those previously published with slight variations: TAATCCCY for Otx1 and TCACACCT for Vegt (Fig 2.7). I performed a genome-wide search for these motifs to identify possible docking sites for Vegt and Otx1 in the genome. While there are 100,899 Otx and 58,740 T-box motifs genome-wide, only a fraction of these motifs are occupied by their corresponding TFs, where Otx1 occupies 1,640 (1.6%), while Vegt occupies 14,064 (23.9%) of their respective motifs (Fig 3.1a,b). As not all motifs are associated with TF binding, this suggests that there are other variables that regulated chromatin association of TFs.

I looked into four possibilities: (1) epigenetic marking, (2) open chromatin, (3) CpG density or (4) coordinated binding with other TFs. First, I was able to rule out epigenetic marking because the enhancer marks H3K4me1, Ep300 and H3K27ac are largely deposited after ZGA[12,13]. Similarly, the promoter mark H3K4me3 and the polycomb repressor mark H3K27me3 are established after ZGA as well[14]. Second, I was able to rule out the open chromatin as well using a combination of indirect evidence. Evidence from early embryos of *Drosophila*[15] and zebrafish[16] using ChIP-seq for nucleosomal histone H3 and ATAC-seq, respectively, showed that opening of the chromatin is established after ZGA. In *Xenopus*,

while the same genomic datasets are unpublished, histones have been shown to compete with transcription factors such as TBP resulting in genome quiescence[17,18], indirectly showing that the chromatin is closed and inaccessible during stages prior to ZGA. Third, with regards to CpG density, as most Vegt and Otx1 binding occurs in enhancer regions (rather than promoters), it is unlikely that CpG density affects maternal TF binding (Fig 2.8c). Consistently, search for CG sequences in regions near Vegt and Otx1 binding show that these 2-mers are not enriched in Vegt and Otx1 binding (data not shown). I therefore pursued the fourth possibility.

I hypothesized that Vegt and Otx1 coordinate with other maternal TFs that regulated endodermal gene expression, and that the motifs of these endodermal regulators would be enriched near T-box and Otx motifs that are bound by Vegt and Otx1. Among the candidates for this analysis includes Foxh1[1,19,20], Sox3[21] and Pou5f3/Oct60[1,22], which are ubiquitously expressed but have been shown to regulate endodermal genes. More importantly, I looked into endodermally-active TFs including Sox7[21,23]; Smad2/3 as the signal transducer of the endodermally-active Nodal signaling[24,25]; and β-catenin with Tcf/Lef as the signal transducer of Wnt signaling [26-29]. I used previously identified motifs for these TFs using published *Xenopus* ChIP-seq datasets: ACAAWRV for Sox7 and Sox3[30], CAGAC for Smad2/3[1], β-catenin/Tcf/Lef[31,32], and ATWYRCA for Pou5f3/Oct60[30]. I find that the region around T-box and Otx motifs bound by Vegt and Otx1, respectively, tend to be enriched for motifs for other endodermally-active factors (Fig 3.1c,d). In contrast, I find that neither TF-bound Otx or T-box motifs regions are enriched for motifs for the ubiquitously expressed Fox and Pou motifs. This finding supports the idea that Vegt and Otx1 binding acts by

coordinated binding with other TFs, and this mechanism is regulated by clusters of their consensus binding motifs.

Vegt and Otx1 associate with two types of cis-regulatory modules

In chapter 2, I showed that Vegt and Otx1 collaborate with Foxh1 and that the Fox motif is enriched by *de novo* motif analysis in the Vegt and Otx1 peaks. This is inconsistent with the motif analysis shown above as the Fox motif is not enriched near the T-box and the Otx motifs. This inconsistency suggests that there are multiple sets of peaks that Vegt and Otx1 associate with: the first set of peaks contain their respective motifs and the motifs of other endodermally-active TFs (Fig 3.1c,d), while a second set of peaks contain the Fox motif. To deconvolute the different types of peaks, I looked into subsets of Vegt and Otx1 binding. I grouped genomic regions bound by Vegt, Otx1 and Foxh1 in four categories: bound by Vegt + Otx1 only, Foxh1 + Vegt only, Foxh1 + Otx1 only and Foxh1 + Vegt + Otx1 (Fig 2.9a). I then looked into the motifs present in each category using motifs described above allowing for no sequence mismatches. In the category Vegt + Otx1 only, the motifs that are enriched are those of the endodermally-active TFs (Fig 3.2a). However, the Pou and Fox motifs are not enriched in this category, consistent with the previous finding. Alternatively, the categories of Foxh1 + Vegt only, Foxh1 + Otx1 only and Foxh1 + Vegt + Otx1 all showed similar motif composition. These regions are enriched for Fox, Pou and Sox motifs, but not T-box, Otx, Smad or Tcf/Lef (Fig 3.2a). These regions appear to be the alternative set of peaks that Vegt and Otx1 associate with that contains the Fox motif.

As consensus binding motifs are important components of TF binding *in vitro*, I found it surprising that these Vegt and Otx1 binding regions do not contain their respective motifs. Since I used a stringent criterion whereby I only use perfect matches for the T-box and Otx motifs, I considered the possibility of wobbly motifs, and allowed for 1-base mismatch during my motif search. With this analysis, the category Vegt + Otx1 are still enriched for T-box and Otx motifs while the categories Foxh1 + Vegt only, Foxh1 + Otx1 only and Foxh1 + Vegt+Otx1 are not (Fig 3.2b). Overall, my motif analyses suggest that Vegt and Otx1 interact with two distinct designs of cis-regulatory modules based on motif composition (Fig 3.2c). Type I modules are enriched for Fox and Pou motifs (motifs of ubiquitous TFs); while Type II modules are enriched for T-box, Otx, Smad2/3 and Tcf/Lef motifs (motifs of endodermally-active TFs); and both module designs are enriched for the Sox motif (either Sox3 or Sox7).

Type I and Type II modules are active during ZGA

In Chapter 2, I showed that the Foxh1 + Vegt + Otx1 triple bound regions are likely active enhancers as they are decorated with enhancer marks during ZGA and are bound by zygotic TFs at later stages (Fig 2.10). This finding indicates that at least a subset of Type I modules are active. I therefore asked whether other Type I modules (Foxh1 + Vegt only, and Foxh1 + Otx1 only) and Type II modules are active as well, and whether active chromatin marks H3K27ac and Ep300 decorate both module types. Previous studies

showed that at the whole genome level, the H3K27ac mark consistently increased from pre-ZGA (stage 8) through post-ZGA (stage 9 and 10)[12]. In Type I and Type II modules, this mark is present at stage 8 prior to ZGA, and increases in signal at stage 9 (Fig 3.3a). In contrast to the whole genome analysis, the mark decreases from stage 9 to 10, and is even absent in the Type II modules at this later stage. Similarly, both module types are decorated with Ep300 marks in the late blastula and early gastrula stages (stage 9 and 10), and the signal decreases in strength in late gastrula (Stage 12) (Fig 3.3b). This finding suggests that both Type I and Type II modules are active during the early stages of development surrounding ZGA.

Type I but not Type II modules are active post-ZGA

To further support my findings, I looked into the general enhancer mark H3K4me1 as well. In both module types H3K4me1 is absent prior to ZGA (stage 8) (Fig 3.4a). This finding is unsurprising as previous publications and my own findings (Fig 2.10e) have shown that H3K4me1 is established in enhancers after ZGA[12,13,30]. However, what is surprising is that this mark is present only in Type I modules, but not Type II modules. The signal in Type II modules is not only absent, but also shows a characteristic dip suggesting that H3K4me1 is specifically prevented from Type II modules. In contrast, the H3K4me1 signal in Type I modules persists through neurula stages (stage 16).

Similarly, a survey of zygotic TF binding suggests that Type I modules are more likely to be

used as zygotic cis-regulatory regions than Type II modules. The zygotically-active mesendodermal genes Smad2/3[1], Foxa[30], Gsc[33], Otx2[33] and Ctnnb1/β-catenin[31] at early gastrula stage 10 tend to bind to Type I over Type II modules (Fig 3.4b). The preference is particularly strong in Type I modules that were triple bound by all three maternal TFs. Interestingly, the binding of Smad2/3, Gsc and Otx2 shows a characteristic dip in Type II modules similar to H3K4me1 signal in these regions, suggesting that these TFs are prevented from binding to these modules. At late gastrula (stage 12), the T-box transcription factors Eomes, Tbxt/Brachyury and zygotic Vegt[34] preferentially bind Type I over Type II modules (Fig 3.4c). The H3K4me1 marking and zygotic TF binding suggest that Type I module activity persists after ZGA, while Type II modules are decommissioned.

Discussion

In this chapter, I analyzed our previous ChIP-seq datasets of maternal TFs Foxh1, Vegt and Otx1 using a cis-regulatory module-centric approach. In the early embryo, endoderm specifiers Vegt and Otx1 associate with two different types of modules. Type I modules contain motifs for ubiquitously expressed factors including the Fox motif and is bound by Vegt and Otx1 likely through association with Foxh1, while Type II modules can be bound by Vegt and Otx1 directly through their respective motifs, which are flanked by motifs of other endodermally-active TFs (Fig 3.2). Functionally, both types of modules appear to be active during ZGA, as both are decorated by H3K27ac and Ep300 particularly in developmental stages surrounding ZGA (Fig 3.3). However, after ZGA, Type I but not Type

II modules are marked by the general enhancer mark H3K4me1 suggesting that Type I modules maintain activity through later embryonic development, while Type II modules are decommissioned (Fig 3.4). Overall, this chapter elucidates a previously unknown connection between regulatory module design, TF binding, and temporally-resolved functionality of enhancers based on epigenetic marking during *in vivo* endoderm formation. While correlative, these findings provide a useful view of enhancer activity during early vertebrate embryogenesis.

Modes of transcription factor association to the genome

I set out to identify the mechanistic nature of TF binding *in vivo*. In the case of the endoderm specifying maternal factors Vegt and Otx1, I identified two distinct mechanisms that are co-occurring prior to zygotic genome activation. First, in Type II modules, Vegt and Otx1 associate with the chromatin containing their respective motifs, which are flanked by motifs of other endodermally-active TFs. This configuration suggests that Vegt and Otx1 associate with Type II modules through coordinated binding with other endodermally-active TFs. Alternatively, while the T-box and Otx motifs are not present in Type I modules, the motifs of ubiquitously expressed maternal TFs are clustered in these modules, which includes the motifs of Foxh1 and Pou5f3/Oct60. Rather than direct binding through their respective motifs, Vegt and Otx1 could associate with this chromatin through recruitment by other factors bound in this Type I module. In fact, when I knocked-down Foxh1 and tested for Vegt and Otx1 binding by ChIP-qPCR, I find less association of Vegt and Otx1 to

Type I modules (Fig 2.9d). This suggests a Foxh1 requirement for association to these modules.

The second mechanism has similarities with the pioneering role of the TF Zelda during the *Drosophila* ZGA. In Zelda mutants, the chromatin binding of lineage-specific factors such as Dorsal[35] (germ layer formation) and Bicoid[36] (anterior-posterior patterning) show loss of chromatin association in regions where Zelda co-binds with these factors. However, not all binding is affected; and ~40% of Bicoid binding and ~60% of Dorsal binding are maintained in Zelda mutants. This suggests that there exist alternative mechanism of Bicoid and Dorsal chromatin association and that the pioneering function of Zelda is insufficient to explain Bicoid and Dorsal binding. Possibly, these mechanisms could involve coordinated binding with other lineage-specific factors as I have seen with Vegt and Otx1.

Birth and death of enhancers

In Chapter 2 and Chapter 3, I explored the sequence of events leading to an active enhancer. My findings, along with previous publications, all converge on the idea that the emergence of active enhancers leading up to ZGA is initiated by maternal TF binding to the chromatin[30,37]. Particularly, the combinatorial docking of maternal TFs is an important component of establishing these enhancers. These enhancers are then are decorated with active enhancer marks such as H3K27ac and Ep300. This suggests that epigenetic marks are a consequence of TF binding and that TFs recruit writers of epigenetic marks. While

data regarding this mechanism is sparse, Ctnnb1/β-catenin can directly interact with the methyltransferase Prmt2[37]. Prmt2 then deposits H3R8 dimethyl marks resulting in the activation of genes such as *sia1* and *nodal3*. My findings suggest that since Foxh1, Vegt and Otx1 bind to regions that are later enriched with H3K4me1 and H3K27ac, these maternal TFs could interact with methyltransferases and acetyltransferases to deposit these marks. As our genomics approach is limited in that it only provides correlative data, protein-protein interactions using co-immunoprecipitation and mass spectrometry could be used to provide data for mechanistic models of TF and epigenetic writer recruitment.

As genes are temporally regulated, it should be unsurprising that cis-regulatory regions are only active at specific windows of time in development. In the case of endoderm development, Type I and Type II modules appear to have different life spans. While both are established by maternal TFs during cleavage stages, generally, Type I modules persist in activity at least throughout neurula stages, while Type II modules are decommissioned after ZGA. The process of enhancer decommissioning was initially described in mouse embryonic stem cells, whereby enhancers of active genes that are required for pluripotency are demethylated at H3K4/H3K9 by LSD1 during trophectoderm differentiation[38]. As Type II modules show a characteristic dip in the H3K4me1 marks after ZGA, possibly, a homologous gene to the mouse LSD1 is involved with Type II modules in regulating endodermal genes. Indeed, many lysine-specific demethylases are maternally-expressed in *X. tropicalis* embryos including the orthologous *kdm1a* and *kdm1b*[39]. Other lysine-specific demethylases, which are specific to lysine 4 of histone H3 that are maternally expressed include *kdm5a*, *kdm5b* and *kdm5c*. Interaction of maternal Vegt or

Otx1 with one of these proteins presents an attractive hypothesis in decommissioning Type II modules.

Materials and Methods

Genome-wide identification of the location of consensus binding motifs: Position weight matrices were generated for the T-box [TCACACCT], Otx [TAATCCCY], Sox [ACAAWRV], Smad2/3 [CAGAC], Tcf/Lef [CTTTGAW], Fox [AATHMACA] and Pou [ATWYRCA] motifs using Homer v4.7[40] using the function *seq2profile.pl*. The Homer v4.7[40] function *scanMotifGenomeWide.pl* was used to identify genomic location inputting the previously generated position weight matrices. This function creates a BED file containing motif locations. **Heatmap of motif signals:** A BED file which contains peak regions identified as Type I and Type II modules were expanded such that the peak summits are located at the center with 2500 bases flanking both the 5' and 3' regions of the summit using UNIX. Using the Bedtools v2.19.1[41] function *coverageBed*, the BED file containing the motif locations were overlaid to the BED file containing the Type I and Type II module locations. The output of *coverageBed* was then pre-processed into matrix form, and plotted in R v3.1.10[42] using the *heatmap* function as seen in Fig 3.1.

Lineplot of motif signals and epigenetic signals: The ChIP signal of histone modifications, Ep300 and zygotic TFs near Vegt, Otx1 and Foxh1 peaks were generated as follows. Datasets were downloaded from NCBI GEO using the accession numbers:

GSE85273 for RNA polymerase II and stage 10 Foxa[30]; GSE67974 for H3K4me1 and Ep300[13]; GSE56000 for H3K4me1 and H3K27ac[12]; GSE72657 for stage 10 Ctnnb1/β-catenin[31]; GSE48663 for stage 12 Vegt, Eomes and Tbxt/Brachyury[34]; and GSE53654 for stage 10 Smad2/3[1]. The stage 10 Gsc and Otx2 ChIP-seq were obtained from the DDBJ Sequence Read Archive using the accession numbers DRA000508, DRA000510, DRA000576 and DRA000577[33]. Reads for these ChIP-seq datasets were aligned to the *X. tropicalis* genome v9.0[43,44] using Bowtie v2.2.7[45] with default options. Using Samtools v0.1.19[46], aligned reads in SAM format were converted to BAM format (*samtools view –bS*) and duplicates were removed (*samtools rmdup*). Bedtools v2.19.1[41] was used to convert BAM files to BED format and finally the IGV v2.3.2[47] igvtools function *sort* (default options) was used to generate sorted BED files. Using the Bedtools v2.19.1[41] function *coverageBed*, the BED file containing the motif location (see above) and ChIP-seq signal were overlaid to the BED file containing the peak locations with 2500 bases flanking of Type I and Type II modules. The output of *coverageBed* was then pre-processed into matrix form, the *plot* function in R v3.1.10[42] was used to generate lineplots of the average signal across each column, as seen in Fig 3.2 and Fig 3.4.

Chapter 3 Figures and Tables

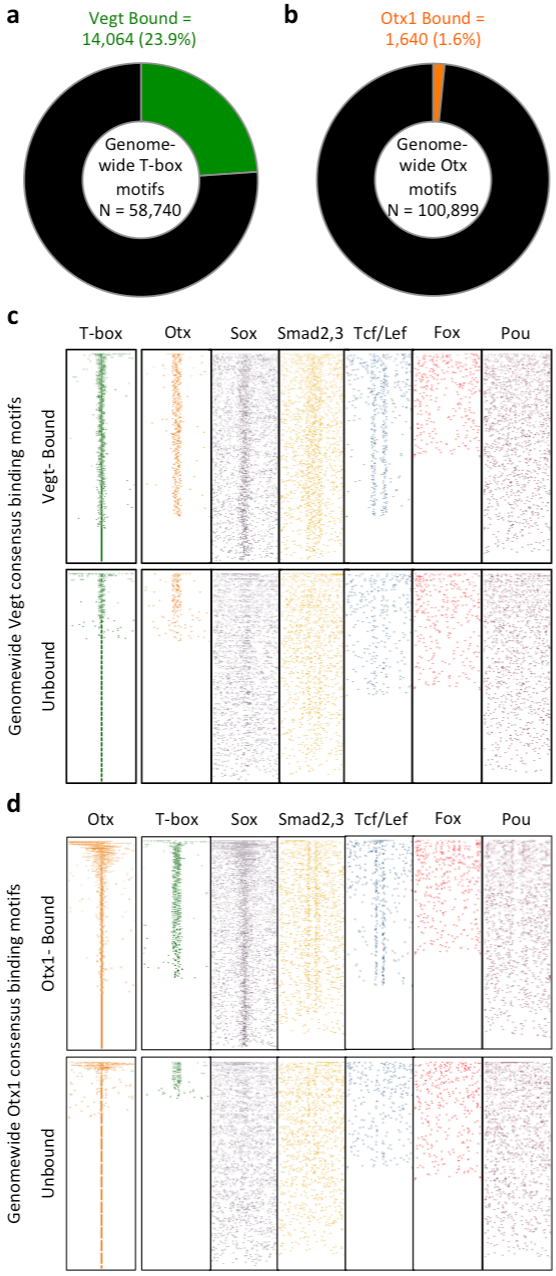

Figure 3.1. *Vegt and Otx1 bind to clusters of DNA binding motifs of endodermally-active transcription factors.* Percent of genome-wide T-box and Otx motifs bound by Vegt (a) and Otx1 (b)*, respectively.* Genome-wide analysis of T-box (c) and Otx (d) motifs that are either bound or unbound by their respective TFs to identify enrichment of endodermal or permissive motifs. Motifs used for analysis: T-box [TCACACCT], Otx [TAATCCCY], Sox [ACAAWRV], Smad2/3 [CAGAC], Tcf/Lef [CTTTGAW], Fox [AATHMACA] and Pou [ATWYRCA].

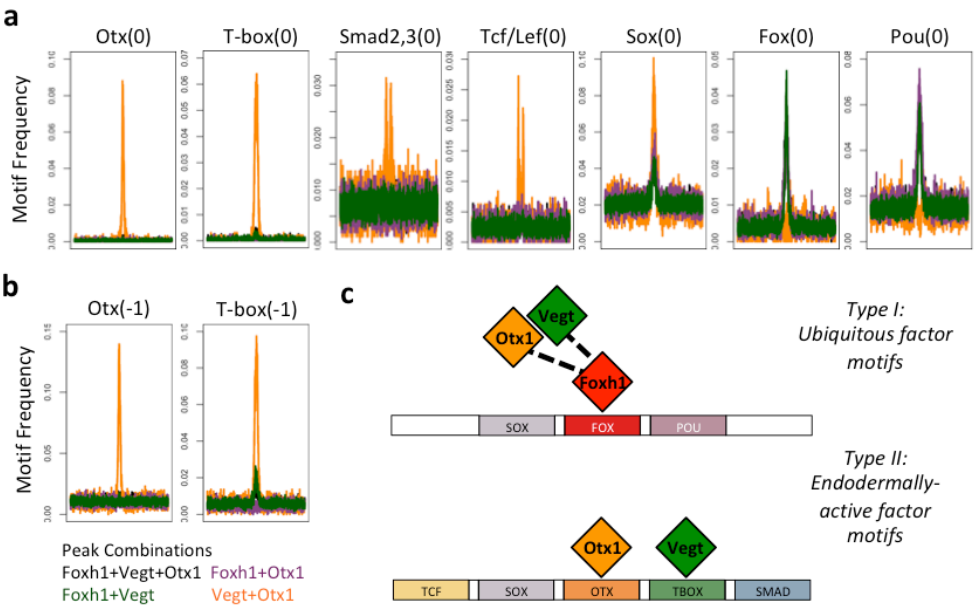

Figure 3.2. *Vegt and Otx1 bind to type I and type II regulatory modules.* (a) Motif composition allowing for 0 mismatches from canonical sequences of regions bound by different combinations of Foxh1, Vegt and Otx1 binding. (b) Wobbly Otx and T-box motifs in the same regions allowing for 1 mismatch from the canonical motif sequences. (C) Model of Vegt and Otx1 association with the type I and type II modules which contain different set of motifs.

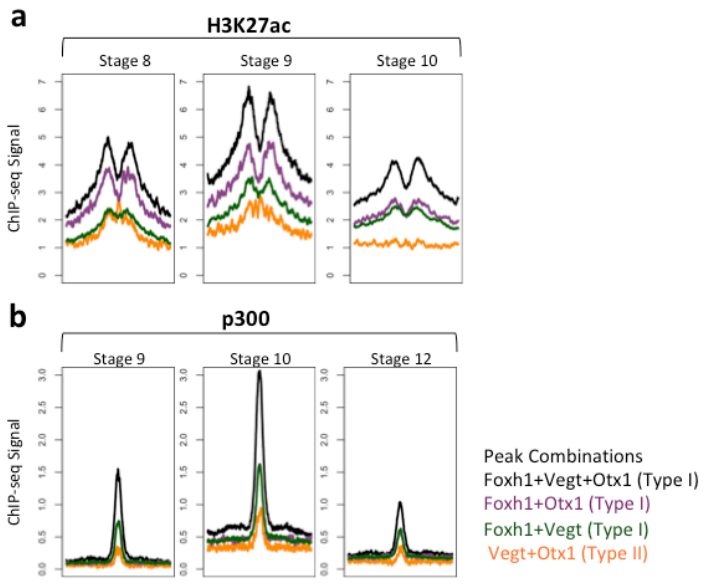

Figure 3.3. *Modules I and II are marked as active enhancers during ZGA.* Active enhancer signals H3K27ac (a) and p300 (b) are present in both modules I and II around ZGA.

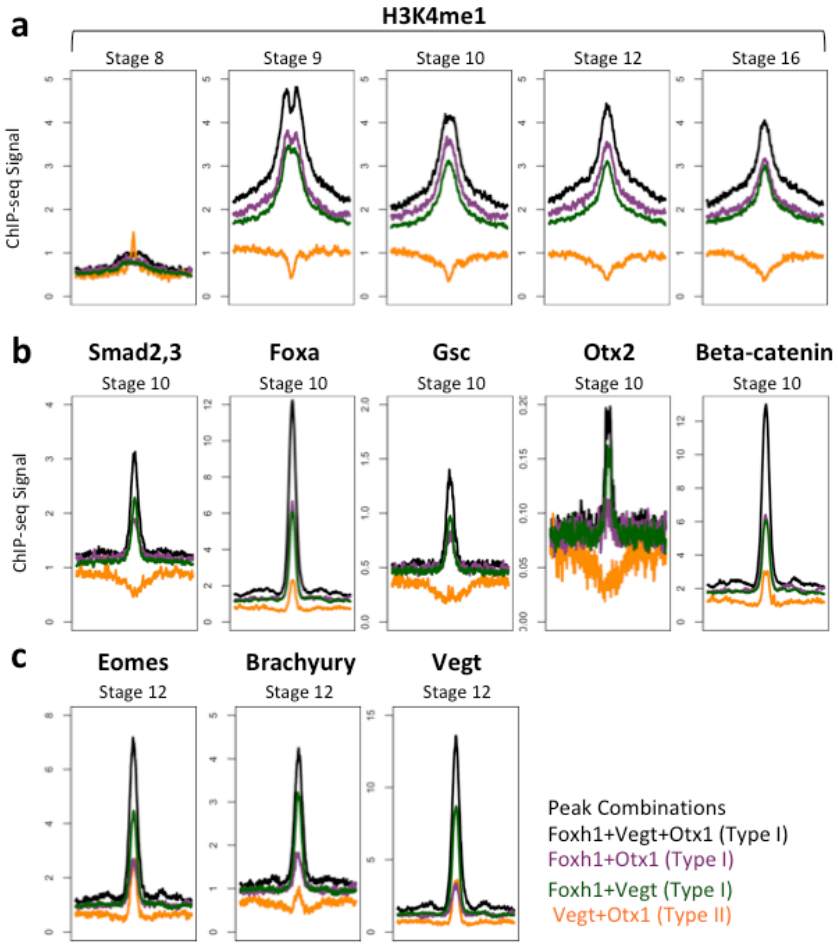

Figure 3.4. *Type I modules are associated with active enhancer activity post-ZGA.* (a) ChIP-seq signal of the enhancer mark H3K4me1. Zygotic transcription factor binding of mesendodermal genes Smad2/3, Foxa, Gsc, Otx2 and Beta-catenin at early gastrula stage 10 (b) and mesodermal genes Eomes, Tbxt/Brachyury and Vegt at late gastrula at stage 12 (c).

References

1.Chiu, W. T. *et al.*. Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program.. *Development* **141,** 4537–47 (2014).

2.Stennard, F., Carnac, G. & Gurdon, J. B. The Xenopus T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation.. *Development* **122,** 4179–88 (1996).

3.Lustig, K. D., Kroll, K. L., Sun, E. E. & Kirschner, M. W. Expression cloning of a Xenopus T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation.. *Development* **122,** 4001–12 (1996).

4.Zhang, J. & King, M. L. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning.. *Development* **122,** 4119–29 (1996).

5.Horb, M. E. & Thomsen, G. H. A vegetally localized T-box transcription factor in Xenopus eggs specifies mesoderm and endoderm and is essential for embryonic mesoderm formation.. *Development* **124,** 1689–98 (1997).

6.Pannese, M., Cagliani, R., Pardini, C. L. & Boncinelli, E. Xotx1 maternal transcripts are vegetally localized in Xenopus laevis oocytes.. *Mech Dev* **90,** 111–4 (2000).

7.Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* **5,** 1752–1779 (2011).

8.Conlon, F. L., Fairclough, L., Price, B. M., Casey, E. S. & Smith, J. C. Determinants of T box protein specificity.. *Development* **128,** 3749–58 (2001).

9.Simeone, A. *et al.*. A vertebrate gene related to orthodenticle contains a homeodomain of the bicoid class and demarcates anterior neuroectoderm in the gastrulating mouse embryo.. *EMBO J* **12,** 2735–47 (1993).

10.Bellipanni, G., Murakami, T. & Weinberg, E. S. Molecular dissection of Otx1 functional domains in the zebrafish embryo.. *J Cell Physiol* **222,** 286–93 (2010).

11.Driever, W., Thoma, G. & Nüsslein-Volhard, C. Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen.. *Nature* **340,** 363–7 (1989).

12.Gupta, R., Wills, A., Ucar, D. & Baker, J. Developmental enhancers are marked independently of zygotic Nodal signals in Xenopus.. *Dev Biol* **395,** 38–49 (2014).

13.Hontelez, S. *et al.*. Embryonic transcription is controlled by maternally defined chromatin state.. *Nat Commun* **6,** 10148 (2015).

14.Akkers, R. C. *et al.*. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos.. *Dev Cell* **17,** 425–34 (2009).

15.Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition.. *Elife* **3,** (2014).

16.Liu, G., Wang, W., Hu, S., Wang, X. & Zhang, Y. Inherited DNA methylation primes the establishment of accessible chromatin during genome activation.. *Genome Res* **28,** 998–1007 (2018).

17.Almouzni, G. & Wolffe, A. P. Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis.. *EMBO J* **14,** 1752–65 (1995).

18.Amodeo, A. A., Jukam, D., Straight, A. F. & Skotheim, J. M. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition.. *Proc Natl Acad Sci U S A* **112,** E1086–95 (2015).

19.Chen, X., Rubock, M. J. & Whitman, M. A transcriptional partner for MAD proteins in TGF-beta signalling.. *Nature* **383,** 691–6 (1996).

20.Chen, X. *et al.*. Smad4 and FAST-1 in the assembly of activin-responsive factor.. *Nature* **389,** 85–9 (1997).

21.Zhang, C. & Klymkowsky, M. W. The Sox axis, Nodal signaling, and germ layer specification.. *Differentiation* **75,** 536–45 (2007).

22.Cao, Y. *et al.*. POU-V factors antagonize maternal VegT activity and beta-Catenin signaling in Xenopus embryos.. *EMBO J* **26,** 2942–54 (2007).

23.Zhang, C., Basta, T., Fawcett, S. R. & Klymkowsky, M. W. SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in Xenopus.. *Dev Biol* **278,** 526–41 (2005).

24.Agius, E., Oelgeschläger, M., Wessely, O., Kemp, C. & De, R. E. M. Endodermal Nodal-related signals and mesoderm induction in Xenopus.. *Development* **127,** 1173–83 (2000).

25.Birsoy, B., Kofron, M., Schaible, K., Wylie, C. & Heasman, J. Vg 1 is an essential signaling molecule in Xenopus development.. *Development* **133,** 15–20 (2006).

26.Heasman, J. *et al.*. Overexpression of cadherins and underexpression of beta-catenin inhibit dorsal mesoderm induction in early Xenopus embryos.. *Cell* **79,** 791–803 (1994).

27.Schneider, S., Steinbeisser, H., Warga, R. M. & Hausen, P. Beta-catenin translocation into nuclei demarcates the dorsalizing centers in frog and fish embryos.. *Mech Dev* **57,** 191–8 (1996).

28.Larabell, C. A. *et al.*. Establishment of the dorso-ventral axis in Xenopus embryos is presaged by early asymmetries in beta-catenin that are modulated by the Wnt signaling pathway.. *J Cell Biol* **136,** 1123–36 (1997).

29.Tao, Q. *et al.*. Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos.. *Cell* **120,** 857–71 (2005).

30.Charney, R. M. *et al.*. Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program.. *Dev Cell* **40,** 595–607.e4 (2017).

31.Nakamura, Y., de, P. A. E., Veenstra, G. J. & Hoppler, S. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules.. *Development* **143,** 1914–25 (2016).

32.Kjolby, R. A. S. & Harland, R. M. Genome-wide identification of Wnt/β-catenin transcriptional targets during Xenopus gastrulation.. *Dev Biol* **426,** 165–175 (2017).

33.Yasuoka, Y. *et al.*. Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification.. *Nat Commun* **5,** 4322 (2014).

34.Gentsch, G. E. *et al.*. In vivo T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency.. *Cell Rep* **4,** 1185–96 (2013).

35.Sun, Y. *et al.*. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.. *Genome Res* **25,** 1703–14 (2015).

36.Schulz, K. N. *et al.*. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo.. *Genome Res* **25,** 1715–26 (2015).

37.Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. & Klein, P. S. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2.. *Dev Cell* **19,** 220–31 (2010).

38.Whyte, W. A. *et al.*. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation.. *Nature* **482,** 221–5 (2012).

39.Owens, N. D. L. *et al.*. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development.. *Cell Rep* **14,** 632–647 (2016).

40.Heinz, S. *et al.*. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.. *Mol Cell* **38,** 576–89 (2010).

41.Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.. *Bioinformatics* **26,** 841–2 (2010).

42.R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.. (2014).

43.Hellsten, U. *et al.*. The genome of the Western clawed frog Xenopus tropicalis.. *Science* **328,** 633–6 (2010).

44.Karimi, K. *et al.*. Xenbase: a genomic, epigenomic and transcriptomic model organism database.. *Nucleic Acids Res* **46,** D861–D868 (2018).

45.Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2.. *Nat Methods* **9,** 357–9 (2012).

46.Li, H. *et al.*. The Sequence Alignment/Map format and SAMtools.. *Bioinformatics* **25,** 2078–9 (2009).

47.Robinson, J. T. *et al.*. Integrative genomics viewer.. *Nat Biotechnol* **29,** 24–6 (2011).

Chapter 4

A gene regulatory program controlling early *Xenopus* mesendoderm formation: Network conservation and motifs

* The text of this chapter of the dissertation is a reprint of the material as it appears on *Seminars in cell & developmental biology*, vol. 66, pp. 12-24. Academic Press, 2017, DOI: doi.org/10.1016/j.semcdb.2017.03.003. The co-author Ken W.Y. Cho listed in this publication directed and supervised research which forms the basis for the thesis/dissertation. I would also like to acknowledge Rebekah M. Charney and Ira L. Blitz for their work on this manuscript.

Abstract

Germ layer formation is among the earliest differentiation events in metazoan embryos. In triploblasts, three germ layers are formed, among which the endoderm gives rise to the epithelial lining of the gut tube and associated organs including the liver, pancreas and lungs. In frogs (*Xenopus*), where early germ layer formation has been studied extensively, the process of endoderm specification involves the interplay of dozens of transcription factors. Here, we review the interactions between these factors, summarized in a transcriptional gene regulatory network (GRN). We highlight regulatory connections conserved between frog, fish, mouse, and human endodermal lineages. Especially prominent is the conserved role and regulatory targets of the Nodal signaling pathway and the T-box transcription factors, Vegt and Eomes. Additionally, we highlight network topologies and motifs, and speculate on their possible roles in development.

1. Vertebrate mesendoderm formation

During early embryogenesis, cell fate specification proceeds through the combinatorial interactions of several signaling pathways and numerous transcription factors (TFs), which function within a broader chromatin landscape. The integration of these factors ("inputs") leads to a specific transcriptome profile ("outputs") that determines the identity of a particular cell. Critical genomic regions for this integration are cis-regulatory modules (CRMs) – combinations of regulatory elements, such as enhancers, where TFs bind to specific sequence motifs and recruit the necessary co-factors [1]. CRMs are critical for the proper implementation of gene regulatory programs in development, because they modulate the rate of gene transcription, and control when a gene is turned "on" or "off" in both time and space. These complex programs can be organized into gene regulatory networks (GRNs) and visualized through logic maps [2], [3]. Elucidating GRNs will enhance our mechanistic understanding of developmental processes, and will enable comparisons across different organ systems, and across different species. These insights will also enhance our understanding of the causes of developmental defects.

Germ layer specification is one of the earliest developmental events in metazoan organisms, preceding the establishment of the organ and tissue primordia that form the complex adult organism. Cells of the three primary germ layers – the ectoderm, mesoderm, and endoderm – become further specified along distinct lineages. Ectodermal cells form the epidermis and nervous system; mesodermal cells become blood, muscle, kidneys, notochord, and connective tissue; and endodermal cells become the

gastrointestinal and respiratory tracts. In amphibian development, the germ layers form along the animal to vegetal axis, which is established during oogenesis. The ectoderm develops from the animal region, while the endoderm arises from the opposite, vegetal pole. The mesoderm forms from the equatorial cells. During early stages, there is overlap between endodermal and mesodermal domains – and therefore it is common to refer to these jointly as the "mesendoderm." Two major advantages of the *Xenopus* system are the ease of obtaining thousands of synchronously developing embryos from a single clutch of eggs, and the ease of performing both gain- and loss-of-function studies to ascertain gene function. These advantages, combined with the relatively close evolutionary distance between *Xenopus* and other vertebrates, makes *Xenopus* a powerful model for elucidating the mechanisms underlying cell fate specification.

A comprehension of the complex GRN architecture that contributes to the specification of the germ layers *in vivo* is a critical unanswered question in developmental and evolutionary biology. Since germ layer formation, like most biological processes, is controlled by a hierarchy of regulatory steps, examining the earliest inputs in the regulation of germ layer development is important. This is underscored by the realization that germ layer specification is not programmed by molecules acting in a linear fashion, but instead is controlled by a set of TFs acting in a complex network. In addition, the study of the GRN controlling amphibian germ layer specification will enable powerful comparisons across different developmental systems, and across evolutionary taxa, to identify core, conserved, GRN structures, as well as subnetworks that were modified during evolution. Beyond animal development, numerous human congenital diseases result from abnormal

formation of the mesoderm and endoderm [4], [5]. Furthermore, recent advances in regenerative medicine and stem cell biology bring the promise of a new era of personalized medicine, aided by highly efficient *in vitro* differentiation techniques [5]. These goals can be significantly aided by a strong understanding of complex *in vivo* cellular differentiation programs, namely "GRN science."

A little over a decade ago, efforts were made to compile the available molecular data into GRNs describing *Xenopus* mesendoderm [6] and Spemann organizer [7] development. Since then, the widespread use of high-throughput technologies (e.g. microarrays, sequencing) provides us with the capacity to significantly broaden the number of network connections and, therefore, our understanding of the structure of the mesendoderm GRN. For this review, we have utilized recent findings to update the *Xenopus* mesendoderm GRN from fertilization through the beginning of gastrulation, linking together critical signaling pathways with transcriptional targets. We discuss the network structure and its motifs, and review areas of conservation across vertebrates.

2. Generation of the mesendoderm gene regulatory network

2.1. Mesendoderm factors

We have assembled a *Xenopus* mesendoderm GRN that reflects data obtained from both *Xenopus laevis* and the closely related diploid species *Xenopus tropicalis*. Each species greatly contributes to the GRN assembly, as *X. laevis* has been traditionally used in the

study of mesendoderm formation, while *X. tropicalis* has been adapted more recently for genomic approaches. The "inputs" in this GRN represent the TFs and signaling molecules (transduced via intracellular TFs) important for mesendoderm formation and early endoderm patterning, many of which have been elucidated [8], [9]. Recently, genome-wide approaches have identified additional localized maternal and zygotic transcripts encoding TFs [10], [11], [12], [13], [14], [15], [16]. Based on a comprehensive catalogue of *X. tropicalis* TFs [14], 130 TFs are found to be enriched vegetally (in comparison to the animal pole), and we have focused on the ~50 TFs expressed at relatively high abundance in the vegetal tissue (Transcripts Per Million values ≥ 50). This corresponds approximately to the expression level of *siamois1* – a critical and localized mesendoderm TF.

2.2. Criteria for identifying transcriptional targets

In order to build the mesendoderm GRN, we have taken the following rigorous approach to determine direct connections between the above TF "inputs" and their downstream target genes [17]. This approach is similar to that previously taken by both Koide et al. [7] and Loose and Patient [6]. First, we enforced that there should be a strong correlation between perturbation of a regulatory TF and the expression changes of the suspected target genes. Regulation can be measured following gain- and/or loss-of-function experiments (e.g., injection of RNA encoding a TF, or a translation blocking antisense morpholino oligonucleotide, respectively) by analyzing changes in target RNA expression (e.g., RNA-seq, RT-PCR, northern blotting, *in situ* hybridization).

Second, we required that the 'inputs' and 'outputs' be expressed in a spatiotemporal manner consistent with regulation and the proposed direction (activation or repression). For a proposed activating ("positive") connection, the upstream TF and proposed target gene must be expressed in the same or overlapping regions, and during overlapping developmental time. Conversely, a proposed inhibitory ("negative") connection should find the target excluded from the spatiotemporal domain of the repressor. However, we note the possibility of finding some exceptions to this rule based on the known biological properties of TFs, and these knowledge-based connections were also included in the network. One example includes a negative autoregulatory feedback loop by *gsc*[18], where the gene modulates its own expression.

Third, we required demonstration of a direct physical interaction between the TF and the regulatory region of the proposed target. As perturbation experiments alone are insufficient to distinguish between direct and indirect connections, we find this criterion essential. This criterion was satisfied experimentally through chromatin immunoprecipitation (ChIP), gel electromobility shift assay (EMSA), DNAse footprinting, or reporter gene assays (containing appropriate mutations). Only connections that satisfied all three criteria were defined as direct. However, we note that DNA binding is only suggestive of functional regulation, and that the 'gold standard' evidence is to mutate the binding site and examine the effect on gene expression. While the rise in the use of ChIP-seq – ChIP coupled with high-throughput sequencing (HTS) – has produced large datasets of physical connections, the vast majority of these sites have not been subjected to laborious mutagenesis assays. In building the network, we therefore distinguished between

functionally validated CRMs and physical interactions lacking this support. Finally, we have also looked to experiments in which protein synthesis is inhibited (e.g. cycloheximide) as a way to determine "directness." Regulatory connections proposed based on perturbation analyses that were validated in the presence of cycloheximide, although lacking in physical binding evidence, were identified as "putative direct" targets.

2.3. Building network connections

To assemble the updated network, we have analyzed hundreds of manuscripts published over the approximately 25 year history of the investigation into the molecular basis of *Xenopus* mesendoderm formation. Building upon the networks presented by Koide et al. [7] and Loose and Patient [6] over a decade ago, we have made extensive use of recently published HTS data using *X. tropicalis* in multiple aspects. First, RNA-seq transcriptome profiling studies have revealed, in great detail, the timing of gene activation [19], [20], [21], [22]. These data have allowed us to incorporate, to a greater extent, temporal information into the graphical organization of our network. Additionally, HTS data has revealed vegetally enriched transcripts [14], [15], [16], which can be difficult to visualize by whole mount *in situ* hybridization.

Second, perturbation experiments, coupled with HTS or microarrays, have provided a wealth of regulatory connections, and ChIP-seq allows for the considerable improvement in the identification of direct target genes *in vivo*. Genome-wide binding of β-catenin [23], [24], T-box TFs [25], Smad2/3 [26], [27], Foxh1 [26], and several organizer-

specific TFs [28]have all been investigated in *X. tropicalis*. Due to differences in data formatting across publications, we have mapped all available ChIP-seq datasets performed from blastula to early gastrula (Supplemental Table S1) to the version 9 *Xenopus tropicalis* genome (www.xenbase.org) using Bowtie2 [29], and identified peaks using the software MACS2 [30]. We then used Bedtools [31] to associate ChIP-seq peaks with TFs in our network, where peaks were filtered using a q-value of 0.01 and a peak distance of 10 kilobases (kb) from the gene body.

Using the criteria described above (Section 2.2), we present a model of the GRN contributing to *Xenopus* mesendoderm development from fertilization through early gastrula (Nieuwkoop-Faber stage 10.5) (Fig. 4.1). In total, this network includes 35 TFs and 12 growth factors. As we have focused on direct transcriptional responses, we have not included well-characterized secreted signaling antagonists such as *chrd*, *nog* or *dkk1* in this network. However, we have chosen to include connections into the multi-signaling antagonist *cer1*, as the regulation of this gene has been extensively characterized [32], [33]. We present a summary of these connections and selected supporting evidence in Table 4.1, and in Supplemental Table S2 which includes the full evidence list and additional connections between mesendodermal genes that did not satisfy our criteria. Thus, the directness of these additional connections is uncertain, and this represents an area of future investigation.

2.4. Organization of the network diagram

Using Biotapestry [34], we have built the GRN as a single bird's-eye view from the full genome [2] (Fig. 4.1). In this visualization, all connections are displayed at once, regardless of time and space. We have made an effort to arrange the genes vertically based on approximate activation time [22], and horizontally from right to left across the dorsal to ventral axis [14]. At the top of the network, maternally inherited TFs (e.g., Vegt, Foxh1) and signaling ligands (e.g., Gdf1, Wnt11b) are shown. All targets of signaling pathways are connected through chevrons indicating cell-surface receptors to their intracellular signal transducers. Zygotically activated growth factor ligands (i.e. Nodals, Bmp4) are connected back through the same signal transducer, so that all connections from a given pathway feed through a single TF node (e.g. Smad2/3). An exception is the Wnt signal transducer Ctnnb1 (β-catenin), which has been displayed twice for purposes of distinguishing between targets of the Wnt11b and Wnt8a ligands. Where the usage of a particular Wnt ligand is unknown, we display a merged path.

We have compared the connections in this network with two previous GRNs [6], [7] and find that a major improvement in the current network is the identification of more bone-fide direct transcriptional connections between TFs (Supplemental Table S3). The current network contains a total of 104 direct network connections – 91 positive and 13 negative. Direct connections are displayed in the GRN as a solid line connecting the upstream TF to its downstream target. We also identify 25 putative direct interactions, which are displayed as dashed lines. Comparatively, Koide et al. [7] and Loose and Patient [6] previously

reported a total of 41 and 60 direct network connections, respectively. Based on differing criteria from the current analysis, some of the direct connections identified by Loose and Patient [6] were here considered putative. Below, we review key features of the mesendoderm network.

3. Maternal factors and the initiation of the network

3.1. Vegetally-localized maternal transcription factors: Vegt and Sox7

Maternal factors play a critical role in the activation of the mesendoderm GRN. The best characterized TF important for the initiation of the *Xenopus* zygotic mesendoderm gene program is the T-box factor Vegt, which is asymmetrically localized vegetally [35], [36], [37]. The maternal knockdown of *vegt* in *Xenopus laevis* results in the loss of both mesoderm and endoderm [38], [39]. Several studies have revealed Vegt to be a master regulator of the endoderm lineage through transcriptional regulation along two parallel routes: the zygotic activation of the Nodal genes, and activation of endodermal TFs [38], [39], [40], [41]. Importantly, the Vegt loss-of-function phenotype can be rescued by the injection of RNA encoding various Nodal ligands, indicating that a critical function of Vegt is the zygotic activation of the Nodal signaling pathway [41]. Consistent with this, *nodal1*, *nodal5* and *gdf3 (derriere)* are direct targets of Vegt, and *nodal* is putatively direct based on protein synthesis inhibition. The expression of *nodal6* and *nodal2* are also regulated by Vegt, but it remains unclear whether these activations are direct.

In addition to regulating the pSmad2/3 signaling pathway, Vegt is important for the direct activation of core mesendodermal genes. TFs *sox17a*, *sox7*, and *gsc* are bona-fide Vegt direct targets; and *sox17b*, *mixer*, *mix1*, *hhex*, and *ventx1* are additional *putative* direct targets. The *cer1* gene, encoding a secreted Bmp/Wnt/Nodal antagonist [42], is also a direct target. At present, we have very little data to fully understand how these targets interact to establish the mesendoderm GRN.

In zebrafish and human, it appears that the T-box transcription factor Eomes plays a role similar to frog maternal Vegt in the activation of the endodermal gene regulatory program. Interestingly, functional studies of *Xenopus* Eomes add further support to the notion that this TF can perform an overlapping role with Vegt in specifying early mesendoderm. Eomes gain-of-function in naive animal caps results in expression of the mesendodermal genes *mix1*, *t/bra*, *wnt8*, *sox17a*, *foxa4*, and *gsc*[43], [44]. Furthermore, zygotic Vegt and Eomes cooperate in mesoderm formation in the late-gastrula embryo [25]. In zebrafish, perturbation analysis reveals that Eomesa is required for the activation of *sox17*, as well as the two nodal ligands *squint* and *cyclops*[45]. ChIP analyses confirm that Eomesa physically binds to the *squint* locus [46], [47]. Eomesa also binds regulatory regions near *sox17*, and additional binding was observed for *mixl1*, *foxa*, *foxa3*, *vent*, and *gsc*[46].

EOMES ChIP-seq and shRNA knockdown, in combination with microarray analysis, during the *in vitro* differentiation of human embryonic stem (ES) cells to definitive endoderm (DE) reveals that human Eomes regulates a similar set of target genes as *Xenopus* Vegt [48]. The genes *MIXL1*, *GDF3*, *CER1*, *SOX17*, *FOXA1*, *FOXA2*, *FOXA3*, and *VENTX* are likely direct human

Eomes targets based on the application of our criteria. Finally, the *NODAL* gene – in *Xenopus*, *direct* Vegt targets – appears to have a regulatory region associated with Eomes binding, but is unaffected by the knockdown [48]. Overall, the Vegt T-box transcription factor has a highly evolutionarily conserved relationship with Eomes orthologs in some other species, sitting at the top of the gene regulatory hierarchy to function as a master regulator of endoderm formation.

In *Xenopus*, Sox7, a maternal SoxF type TF, has been implicated in the activation of mesendoderm targets. Overexpression of Sox7 mRNA in naive animal caps reveals that *nodal, nodal5*, and *nodal6* are putative direct target genes [49]. Since Vegt also putatively activates zygotic *sox7*expression [49], this indicates that the *nodal* genes are likely co-regulated by maternal Vegt *and* both maternal and zygotic Sox7. This network structure illustrates the importance of understanding gene regulation as a network, instead of activation mechanisms at the single gene/activator level. We also note that recent transcriptome profiling of blastomeres from 8-cell stage embryos identified 65 genes reproducibly enriched in the vegetal pole [12]. Among these, the TFs o*tx1*, *pbx1*, *sox7* and *vegt* are highly enriched in the *Xenopus* vegetal tissue, most of them with poorly characterized roles in endoderm formation.

3.2. Nodal signaling and maternal Foxh1

It is clear that, in addition to directly activating endodermal TFs, a major function of Vegt is to activate expression of the *nodal* genes. The loss of Nodal signaling in *Xenopus* results in

the disruption of endoderm and mesoderm formation and a severe delay or complete disruption of gastrulation movements [26], [50], [51], [52], [53], [54]. These findings are consistent with an analysis of Nodal-deficient zebrafish embryos (*sqt*;*cyc*double mutants) where mesodermal and endodermal markers are not expressed and gastrulation is abnormal [55]. Mice have a single *Nodal* gene, and *Nodal*-null embryos fail to form a primitive streak and do not undergo proper gastrulation [56]. Taken together, Nodal signaling is necessary for mesendoderm development in vertebrates.

How this pathway functions in the early formation of the mesendoderm has been the focus of intense study. The maternal TGFβ ligand Gdf1 (also known as Vg1) is localized to the vegetal cells [57], and likely plays a role in anterior mesendoderm formation [58]; however, the endogenous role of Gdf1 has remained understudied, in part due to the inefficient conversion of the ligand precursor into its mature form [59]. Importantly, expression of *nodal5* and *nodal6* prior to the mid-blastula transition indicate that these ligands contribute to the earliest activation of the Nodal signaling pathway [60], [61], and the early onset of *gdf3* suggests that it also contributes to pathway activation. The Nodal signaling cascade activates transcription in the blastula through phospho-Smad2/3 (in complex with the maternal partner to all R-Smad signaling, Smad4). Overall, the vegetal localization of Nodals and Gdf1 is consistent with the model in which high levels of Nodal promote endoderm and low levels promote mesoderm, which has been observed in *Xenopus* explant experiments [40].

As it is difficult to distinguish between the output of different Nodal and TGFβ ligands, we have therefore fed all 7 ligands through a single Smad2/3 node. The mesendoderm GRN presented here contains 25 direct targets activated by Nodal signaling (Fig. 4.1). These targets include genes encoding core endoderm TFs such as *gata4*, *gata6*, *mix1*, *mixer*, and *foxa4*, dorsal endoderm genes such as *hhex* and *cer1*, the organizer gene *gsc,* and the pan-mesodermal gene *t* (also known as *brachyury*), among others. These targets validate the notion that Nodal signaling contributes broadly in gene activation in the mesoderm and endoderm germ layers. Additionally, positive autoregulation of *nodal1* and *nodal2*promotes further enhanced expression of the signal [62].

The activated Smad2/3-Smad4 complex regulates target genes in concert with co-TFs, and to date, identified Smad2/3 co-factors include Foxh1, Eomes, Foxh1.2, Gtf2i, Gtf2ird1, Mixer, Tcf3 (also known as E2a) and Tp53 [48], [63], [64], [65], [66], [67], [68], [69], [70], [71]. Of these, the transcriptional regulation via Smad2/3 interactions with maternal Foxh1 has been extensively investigated in *Xenopus*[26], [66], [72], zebrafish [70], [73], [74], mouse [75], [76], and differentiated human ES cells [71]. Our *Xenopus* GRN contains direct targets activated via Smad/Foxh1, including the growth factors, *nodal1*, *nodal2*, and *wnt8a*; the BMP/Wnt/Nodal antagonist *cer1*; and the TFs *gsc*, *otx2*, *mix1*, *hhex*, *lhx1*, *sebox*, and *pitx2*. Among these, to date, *gsc* and *pitx2* regulation via Foxh1/Smad2 is conserved across mouse, fish and frog [62], [77]. Finally, in human definitive endoderm differentiation *CER1*, *PITX2*, *GSC*, and *MIXL1* are also induced by FOXH1/SMAD2 [78].

It has been well-characterized in mouse, fish and frog that the loss of Foxh1 does not fully recapitulate the loss of Nodal signaling – indicating the necessity for additional Smad2/3 binding partners. To this extent, in zebrafish, Eomes has been implicated as the Smad2/3 co-factor responsible for the remaining Nodal-mediated regulation that occurred in the *Foxh1*-null [70]. Our network suggests *gata4*, *gata6*, *eomes*, and *foxa2* are also regulated by Nodal-signaling via a Foxh1-independent mechanism. It will be necessary to investigate whether Eomes, and perhaps Mixer, regulate these targets in a Foxh1-indepenent fashion.

3.3. Wnt/β-catenin signaling

In addition to germ layer specification along the animal-vegetal axis, Nodal signaling is critical in patterning the embryo along the dorsal-ventral axis. While *vegt* mRNA appears to be uniformly distributed across the vegetal tissue, Nodal signaling is higher in the dorsal mesendoderm of the *Xenopus* blastula [52], [79]. This pattern is attributed to high levels of Wnt/β-catenin signaling on the dorsal side of the embryo. While a detailed discussion of dorsal-ventral patterning of mesendoderm is beyond the scope of this review, it is useful to discuss in brief the role of maternal Wnt/β-catenin signaling. Maternal *wnt11b* is localized to the vegetal pole in the egg, relocated to the dorsal vegetal cells following cortical rotation, and activates a Wnt signaling pathway to specify dorsal fate [80], [81], [82]. Dorsal nuclear β-catenin directly regulates *sia1* and *sia2*, two homeobox genes that control dorso-anterior specification [83], [84], and many other genes, via the canonical Wnt cascade feeding through β-catenin-Lef/Tcf complexes. Maternal Wnt/β-catenin signaling also activates the expression of all Nodal genes in the dorsal mesendoderm, in particular

the early activation of *nodal5* and *nodal6*[61], [85]. Both Nodal and Wnt signaling are critical for the formation of the Nieuwkoop center and Spemann organizer [84], [86], [87], and the network reveals substantial overlap in the regulation of dorsal mesendoderm target genes such as *hhex*, *lhx1*, *otx2*, *cer1* and *gsc*. Consistent with this crosstalk model, there is a substantial co-occurrence between Foxh1 (a major Smad2/3 co-factor) and β-catenin ChIP-seq peaks [24], [26]. It should be noted that recent β-catenin ChIP-seq performed by Nakamura et al. [24]reveals β-catenin binding associated with target genes previously thought to be indirectly regulated by Wnt/β-catenin via Sia, such as *hhex* and *gsc*[84], [166]. While the biological activities of these putative enhancers remain to be validated, this finding suggests that dorsal mesendoderm targets are regulated through complex network motifs (see Section 5). Taken together, the mesendoderm GRN is highly controlled by maternal Vegt, and the signaling inputs from Nodal and Wnt signaling pathways, the activation of which coincide with the onset of zygotic gene transcription.

4. Core zygotic mesendoderm transcription factors

A number of zygotic TFs have been identified as critical for the formation of the mesendoderm. Here, we discuss the roles of the Mix family, Gata4/5/6, Foxa and Sox17.

4.1. Mix family

The critical role of the Mix family TFs in mesendoderm development has been investigated across numerous model systems [88]. In the two *Xenopus* species, the single mammalian gene encoding Mix-like 1 (Mixl1) is represented by *mix1*, *mixer*, and species-specific expansions and losses of genes referred to as *bix*[89].

In our network, we have examined connections into and from *mix1* and *mixer* based on integrated data from *X. laevis* and *X. tropicalis*. Of these, *mix1* is the earliest to be activated at the mid-blastula transition via direct regulation by Smad/Foxh1, and putative regulation by Vegt. To date, evidence also supports the direct activation of *gsc* and *cer1*, and the repression of *t* by Mix1. This network supports the notion that Mix1 is critical for the activation of dorsoanterior mesendoderm, and the exclusion of *t* from the dorsal organizer [90], [91], [92]. This places Mix1 at the top of a negative feed-forward loop, together with Gsc [93], which also represses *t* expression (see Section 5.2). This loop may be conserved in mammals, as differentiating *Mixl1*-null mouse ES cells reveal a down-regulation of *Gsc* and an up-regulation of *T*[94]. It is conceivable that Mixer − which also directly activates *Gsc* − functions in a similar capacity to repress ventrolateral mesoderm. Supporting this, Mixer deficient *Xenopus* embryos showed up-regulation of mesoderm genes including *eomes*, *fgf3*, *fgf8*, *not*, and *gata2*[95]. Similarly, *t* expression was mildly reduced in Mixer morphants. Finally, while the relationship between *X. laevis* and *X. tropicalis* Bix TFs is unclear, current evidence suggests that *bix1*, in *X. laevis*, is directly activated by T [96], and in *X. tropicalis*, is activated by Nodal/Smad2 [26]. In *X.*

laevis, *bix2* (also known as *milk*) is putatively activated by Nodal/Smad2 [97], and in turn directly activates *gsc*[65], [97]. Further elucidating the direct targets of Mix TFs in amphibians, fish, and mammals will be crucial to teasing out these subnetworks.

4.2. Gata family

The Gata transcription factors are highly conserved regulators of endoderm formation across metazoan model systems. In invertebrates, Gata transcription factors play crucial roles in the formation of the *Drosophila* midgut [98], [99], in the E lineage during *C. elegans* germ layer patterning [100], [101], [102], and in the sea urchin mesendoderm [2], [103]. In vertebrates, the Gata4/5/6 subfamily of Gata factors play a role in the formation of mouse extra embryonic endoderm [104], [105], and in the formation of *Xenopus*[106], [107] and zebrafish [108], [109] endoderm.

Despite their importance, little is known about their molecular targets. In *Xenopus*, the putative direct targets of Gata4/5/6 are endodermal genes *hnf1b* and *sox17a*. Evidence in mouse ES cells also suggests that both *Hnf1b* and *Sox17a* are Gata4/6 targets [104], [105], and zebrafish *gata5*mutants show a reduction in *sox17* expression [110]. While the directness of these interactions is unclear in the mouse and zebrafish, conservation of gene activation suggests similar network topologies are likely operating between frog, zebrafish and mouse.

A common feature of the Gata factors is the extensive mutual regulation among these three genes. In *Xenopus*, *gata5* gain-of-function upregulates *gata4* expression, while *gata4* gain-of-function upregulates *gata6* expression [107]. In zebrafish, loss of *gata5* downregulates the expression of *gata6* and vice versa [109]. Overexpression of Gata4, Gata5, or Gata6 in mouse ES cells results in the upregulation of all three factors [104], [105], [111], and *Gata4*-null mice show reduced *Gata6* expression [112].

Presently, it is unclear whether these connections are direct; however, functional analyses of a *Gata4 cis*-regulatory module in mouse supports direct co-regulation. Gata4 was shown to bind to an enhancer controlling foregut and midgut expression, suggesting an autoregulatory loop [113]. A second enhancer that controls *Gata4* expression in the septum transversum and the mesenchyme surrounding the liver are bound by all three Gata factors [114]. These data suggest a direct positive relationship between the three factors.

4.3. Sox17

Sox17 is a highly conserved endodermal transcription factor across vertebrates. Sox17 plays a role in *Xenopus*[115] and zebrafish [116]endoderm formation, and in both mouse extra embryonic and definitive endoderm formation [117], and in the definitive endoderm in human ES cell assays [118]. In *Xenopus*, little is known about the direct targets of Sox17, but putative direct targets include *foxa1*, *foxa2*, and *hnf1b*[119], [120], [121]. Sox17 also directly regulates the expression of *foxa2*orthologs in mouse and human extra embryonic and definitive endoderm [118], [122]. In addition, Sox17 targets the extra embryonic

endoderm marker *Hnf1b* in mouse ES cells [122]. However, it's not known if either of these targets are direct. Finally, functional evidence shows that Sox17 genes are subjected to a positive feedback loop [119], [121].

4.4. Foxa family

Foxa TFs are critical for endoderm development across diverse organisms [123], [124]. Of the three Foxa TFs in mouse (*Foxa1*, *Foxa2*, and *Foxa3*), Foxa2 is required for early development. *Foxa2*-null mice display defects in the node (the equivalent of the Spemann organizer) and later gut tube [125]. Both Foxa1 and 2 bind to liver-specific enhancers in mouse pluripotent gut endoderm, well before these genes become transcriptionally active [126], and genetic analyses indicated that these TFs function together in hepatic development [127]. Since they have the capacity to bind to and open compact chromatin [128], Foxa TFs have been deemed 'pioneer factors' for gut development [129].

The *Xenopus tropicalis* genome encodes three Foxa TFs (*foxa1*, *foxa2*, and *foxa4*), which are zygotically transcribed. Gain-of-function analyses indicate that Foxa2 inhibits mesoderm and anterior endoderm formation in the gastrula embryo [130]. Loss-of-function analyses in sea urchin also support the conservation of this mechanism [131]. However, as the overexpression of VP16-Foxa2 fusion protein phenocopied the overexpression of Foxa2, Foxa2 presumably functions as an activator [130], and the authors postulate that Foxa2 activates a key repressor of mesodermal cell fate. At present, the direct targets of Foxa in

the early embryo are unclear. Finally, in *Xenopus*, *foxa4* is the earliest and most abundantly expressed *foxa* gene during early mesendoderm specification. However, Foxa4's role in early mesendoderm development is not known, although by early neurula stages it promotes notochord formation and inhibits prechordal and paraxial mesoderm [132]. Our network analysis reveals that *foxa4* is activated by Smad2/3, via a Foxh1-independent mechanism [26]. Since *foxa4* expression is repressed directly by the Smad1 target *ventx1*[133], these connections support exclusion of *foxa4*expression from the ventrolateral mesoderm.

5. Network motifs in the Xenopus mesendodermal GRN

Network motifs are a subgroup of patterns found in GRN architectures. Here we analyzed the network motifs – representing autoregulatory, feedback, and feedforward loops [134] – found in the network presented here, as well as those previously reported [6], [7]. Due to the increase in the number of direct connections presented in this review over previous networks, we identified significantly more motifs (Fig. 4.2 and Supplemental Table S3), which we will discuss below.

5.1. Autoregulatory and feedback loops

Autoregulatory loops involve the self-regulation of a transcription factor or a signaling pathway (Fig. 4.2A and Supplemental Table S4). Based on our criteria, current evidence supports 4 direct autoregulatory loops. These include the positive autoregulation of Nodal

signaling in the endoderm and dorsal mesoderm [26], [62], [135]; the positive autoregulation of *ventx2* on the ventral side of the embryo [136]; the negative autoregulation of *gsc*[18], [28], [137]; and the exclusion of *wnt8a* expression in the dorsal mesendoderm through the action of Tcf/Ctnnb1 [24], [138].

We next computationally interrogated our network for feedback loops – a motif that involves mutual regulation between two genes X and Y (Fig. 4.2B and Supplemental Table S4). In a negative feedback loop, gene X positively regulates gene Y, while gene Y negatively regulates gene X; a double negative feedback loop is defined by mutual inhibition between the two genes. A positive feedback loop involves a mutual activation between genes X and Y. We identified five feedback loops in our network – one negative feedback, one double negative feedback and three positive feedback loops.

Based on gain- and loss-of-function analyses, a double negative feedback loop between the dorsal organizer gene *gsc* and the ventral gene *ventx2*has been proposed [139]. While reporter assays and EMSA experiments had confirmed *gsc* as a direct target of Ventx2 [140], it was only recently via ChIP-seq that Gsc binding to *ventx2* regulatory regions has been confirmed [28]. This type of feedback loop enables the formation of sharp expression boundaries between cell lineages, as computationally demonstrated in modeling a double negative feedback between *gsc* and *t*[141], [142].

The network reveals a negative feedback loop between Gsc and Wnt/β-catenin, whereby *gsc* – activated by maternal Wnt11b/β-catenin [82], [85], [138], [143] – represses

the expression of zygotic *wnt8a* from the dorsal organizer [28], [137]. This type of feedback loop can be useful in cases where the initial activator becomes unnecessary in the control of later gene expression.

Positive feedback loops enable continuous expression of two genes that are important for the same lineage. On the ventral side of the embryo, Smad1 mediating BMP signaling binds the regulatory region of *ventx2* and activates its expression, which in turn appears to regulate the expression of *bmp4* and increases the production of Bmp4 ligand [144]. Finally, in the mesendoderm, the Nodal and Wnt/β-catenin signaling pathways positively feed back into each other [24], [26], [85], [119], [145], [146], [147]. In this motif, maternal Wnt11b/β-catenin activates the expression of all *nodal* genes in the dorsal mesendoderm. In turn, Smad2/3 (mediated by Foxh1), activates zygotic *wnt8a*, which is excluded from the dorsal organizer via the *gsc* gene as discussed above.

5.2. Feedforward loops

One important feature found in the network presented is the prominent presence of feedforward loops in three-gene network motifs. Of the eight different types of feedforward loops [148] (Fig. 4.2C, Supplemental Table S4), the coherent feedforward type I is the most abundant. The relative abundance of coherent feedforward type I loops compared to other types of feedforward loops is a feature found in a variety of transcriptional GRNs, including *E. coli*[148], *P. aeruginosa*[149], and *S. cerevisiae*[148]*, as well as D. melanogaster* embryogenesis [150], [151]. In our network, approximately three-quarters of

identified feedforward loops are coherent type I (Supplemental Table S4). In this type of loop, a positive regulator (gene X) and its target (gene Y) both positively regulate the expression of a common downstream gene (gene Z). In the *Xenopus* mesendoderm GRN, this loop appears frequently where gene X is maternal, gene Y is a primary activated zygotic gene and gene Z is a secondary activated zygotic gene (Fig. 4.2D). In the majority of cases, the initial activator appears to be either β-catenin, Foxh1, Smad2/3, or Vegt (gene X). These maternal factors activate the expression of early and mid-blastula zygotic genes such as *wnt8a*, *sia1*, *sia2*, *mix1*, *gsc*, and the *nodal* genes (gene Y), which, in turn, activate the expression of a larger number of later expressed mesendodermal genes (gene Z). Some examples of gene Z include *cer1*, *eomes*, *ventx1*, *hhex*, and *pitx2*. The benefits of this type of loop depend on whether the co-regulation of Z by both X and Y is an 'AND-gate' (where both factors are required to activate factor Z) or an 'OR-gate' (where either factor can activate gene Z) [134]. An 'AND-gate' can be beneficial in the tight control of factor Z expression, as factor Z is only activated once factor Y is expressed. On the other hand, an 'OR-gate' enables the sustained expression of factor Z, despite the loss of the initial factor X [134]. Such would be a critical motif functioning in early *Xenopus* embryogenesis, where maternal factors that act as the initial activators are degraded during blastula stages. Then, their direct, primary, activated zygotic targets can function to maintain the expression of later, secondary, activated genes. It is not clear as to whether the coherent type I feedforward loops we identified are controlled by AND-gates or OR-gates. It will be important to address this type of question as we further refine our understanding of gene regulation in the early *Xenopus* mesendoderm.

The other types of feedforward loops involve a negative regulation between genes X, Y and Z [148] (Fig. 4.2C, Supplemental Table S4). In the *Xenopus* system, these types of loops appear enriched among genes required for dorsal-ventral patterning of the mesendoderm. For example, Smad2/3 signaling promotes both *t* and *gsc* expression [26], [43], [87], [93], [152]. In turn, *gsc* represses *t* from the dorsal organizer. In a similar loop, *otx2* also functions in the restriction of *t* expression [90], [153]. These loops, along with negative and double negative feedback loops, appear to be particularly useful in regionalizing the dorsal and ventral sides of the mesendoderm. It is likely that more of these types of loops will be identified in the future, as we better define the expression domains of more transcription factors during *Xenopus* gastrulation [14], [15], [16]. Also, as this review focuses on mesendodermal genes, the molecular mechanism for the regionalization of the embryo between mesendoderm and prospective ectoderm is not reflected on this network. However, as some animal pole factors repressing Nodal signaling have been identified (e.g., maternal Sox3) [154], similar network motifs might be used in animal-vegetal patterning.

6. Prospects

We have generated a comprehensive gene regulatory network governing *Xenopus* mesendoderm development. New findings, and the increased accessibility of HTS technologies, have contributed greatly to the number of direct regulatory interactions between critical factors, and have revealed many more possible players whose functions remain unclear. We predict that the network will become more complex as more datasets are generated. In addition, advances in the use

of CRISPR/Cas9 in *Xenopus* provides the opportunity to modify the endogenous interactions between TFs and their target CRMs. This will enable in-depth investigations into the role of CRMs in gene regulation, ultimately aiding in addressing one of the most critical questions in biology: how mutations in regulatory regions influence overall gene expression levels.

The mesendoderm network is initiated in the blastula embryo at the onset of zygotic gene activation, and the network presented here – which extends through the beginning of gastrulation – covers a timespan of approximately 3 h. During this time of rapid developmental transitions, as maternal factors – important for the initiation of the network – are degraded, and zygotic transcription ramps up, we expect the network to be highly dynamic. While single-stage analyses may be sufficient in identifying direct target genes, the investigation of TF targets over time provides valuable kinetic information to uncover the complexities of the dynamic regulatory network.

In addition to transcription factors and signaling molecules, screens for non-coding RNAs, including microRNAs [155], [156], [157], [158], [159] and long non-coding RNAs [160], [161], have identified many more potential regulators of gene expression. One example is through the negative regulation by mir-427 of the Nodal ligands *nodal5* and *nodal6*, and the nodal antagonist *lefty*[162]. Loss of function of mir-427 leads to mesodermal patterning defects. Interestingly, the interaction between Nodal signaling, Lefty and mir-427 generates an incoherent type II feedforward loop. As we learn

more about the roles of these non-coding RNAs, we will have to integrate their regulatory roles into the GRN diagram.

As the GRN increases in connectivity and becomes more complete, this will enable future researchers to investigate the GRN from a systems level perspective. Identification of various network motifs can provide some new and interesting hypotheses based on the theoretical properties of these motifs [134], which can be experimentally tested *in vivo*. Additionally, quantitative modeling of these network can delve deeper into the nature of the regulatory relationships between transcription factors, as has been done in smaller networks [141], [142], [163]. Some of the challenging questions in the future are to parse which mesendodermal factors play major roles in regulating network function and maintaining network output robustness; as well as the role of redundancy in network regulation, and feedback/feedforward loop regulation. The *Xenopus* mesendoderm GRN, with its rich history and amenability to modern genomic tools, presents itself as one of the best systems to study these types of network questions *in vivo*. We hope that this network will provide a useful framework in moving towards a greater understanding of the complex GRN controlling early mesendoderm development, and as well as the formation of later endodermal derivatives.

Supplemental Tables

Supplemental tables can be accessed at https://doi.org/10.1016/j.semcdb.2017.03.003

Acknowledgements

We thank Xenbase (http://www.xenbase.org/; RRID:SCR_003280) for Xenopus resources, the members of the Cho laboratory for useful discussion, and the UC Irvine Office of Information Technology for ongoing support of the High Performance Computing Cluster. We apologize to researchers whose work we were unable to include due to space limitations. This work was supported by NIH HD073179 to K.W.Y.C.

Chapter 4 Figures and Tables

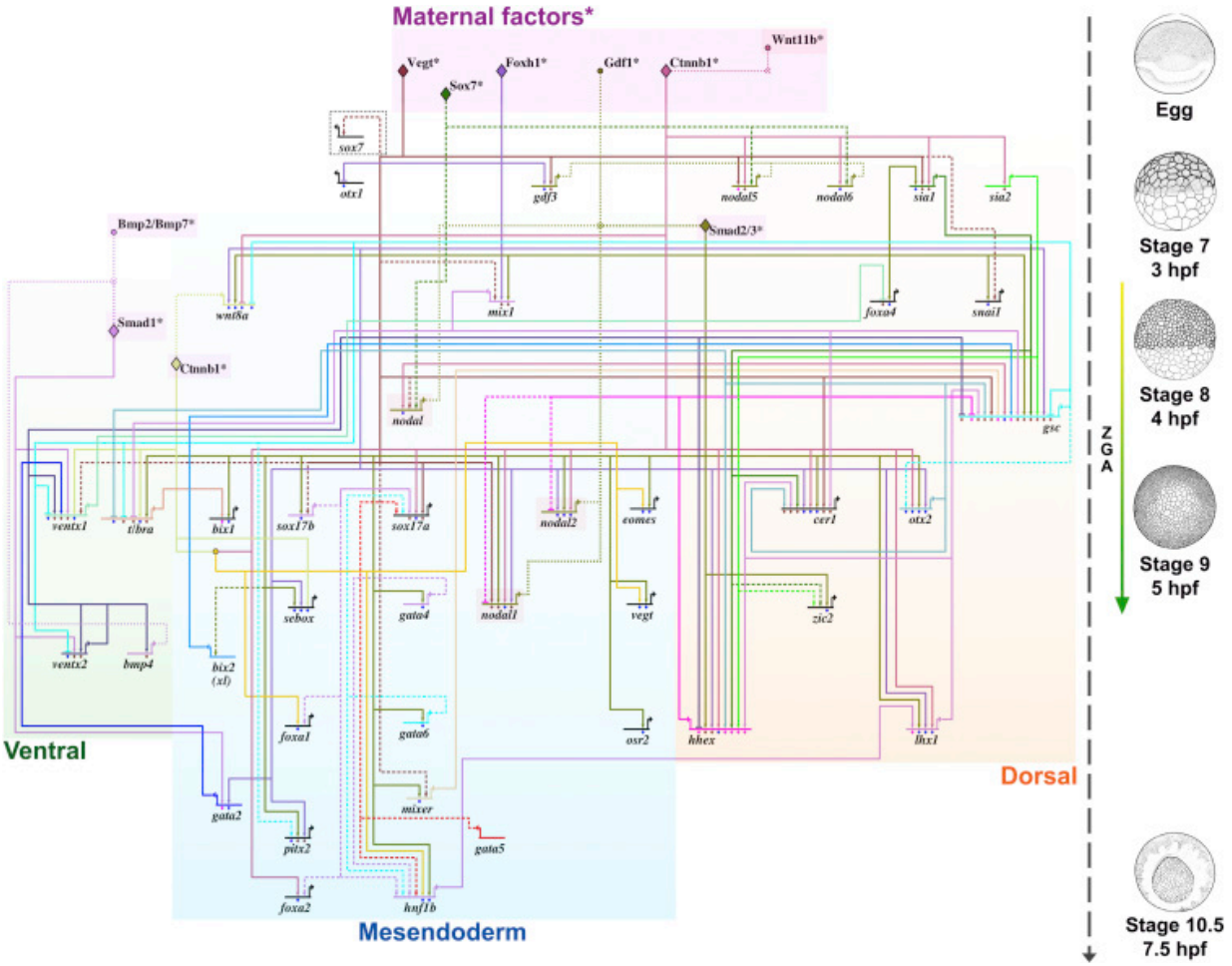

Fig. 4.1. *Xenopus mesendoderm gene regulatory network from fertilization through early-gastrula*. The network is comprised of 35 TFs and 12 growth factors. Maternal proteins are represented as diamonds, and signaling ligands as circles. Connections are drawn from the transcriptional regulator to the *cis*-regulatory region of the target gene, and direct connections are indicated as solid lines and putative connections as dashed lines. Connections from secreted ligands pass through a chevron and are mediated by their respective intracellular TFs (e.g. Smad2/3, Ctnnb1/β-catenin). Approximate spatiotemporal information is provided from top to bottom (egg through early-gastrula) and from right to left (dorsal to ventral), with some exceptions (e.g. *t/bra*). The activation time of zygotic *sox7*(boxed) is unknown. All direct connections are annotated for TF binding (blue diamond), reporter assay (pink diamond), and TF binding plus functional validation (maroon diamond). For additional connection details, including experimental evidence and references, see Table 4.1 and Table S2. ZGA, zygotic genome activation.

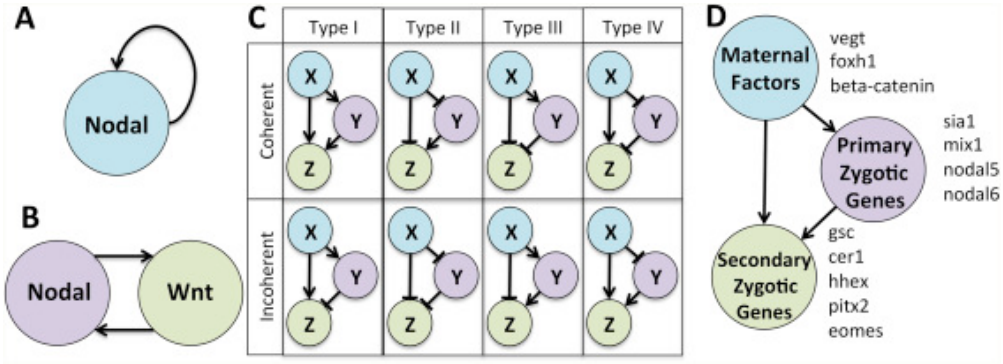

Fig. 4.2. *Network motifs found in the Xenopus mesendoderm GRN.* (A) Autoregulatory loop, for example by Nodal signaling. (B) Positive feedback loop, for example between Nodal and Wnt signaling. (C) Coherent and incoherent feedforward loops and their regulatory structure. (D) The type I feedforward loop, which appears to be the most common feedforward loop in the GRN, frequently appears in the structure such that X is a maternal factor, Y is an early zygotic gene, and Z is either an early or late zygotic gene.

Regulator	Direct Targets	Selected References
ctnnb1/β-catenin (Wnt)	cer1, eomes, foxa1, foxa2, gsc, hhex, hnf1b, lhx1, nodal1, nodal2, nodal, nodal5, nodal6, otx2, sebox, sia1, sia2, sox17a, t, vegt, ventx1, wnt8a	[23], [24], [82], [83], [84], [85], [119], [138], [145], [147], [164], [165], [166]
vegt	cer1, gdf3, gsc, mix1*, mixer*, nodal1, nodal*, nodal5, sia1, snai1*, sox17a, sox17b1*, sox7*, ventx1*	[33], [35], [36], [38], [41], [85], [121], [165], [167], [168], [169], [170], [171], [172]
foxh1	cer1, gata2, gdf3, gsc, hhex, lhx1, mix1, nodal1, nodal2, otx1, otx2, pitx2, sebox, wnt8a	[26], [62], [63], [66], [72], [166], [173], [174], [175]
sox7	nodal*, nodal5*, nodal6*	[49], [171]
smad2/3 (Nodal)	bix1, bix2*, cer1, eomes, foxa4, gata4, gata6, gsc, hhex, hnf1b, lhx1, mix1, mixer, nodal1, nodal2, osr2, otx2, pitx2, sebox, sia1, snai1, sox17b, t, vegt, wnt8a, zic2	[26], [27], [62], [63], [72], [87], [90], [166], [176], [177]
smad1 (BMP)	gata2, ventx1, ventx2	[61], [178], [179], [180], [181], [182], [183], [184], [185], [186], [187], [188]
sia1	cer1, gsc, hhex, zic2*	[33], [84], [91], [92], [152], [189], [190]
sia2	gsc, hhex, zic2*	[84], [166], [191], [192]
mix1	cer1, gsc, t	[33], [90], [91], [92], [152], [190], [193]
mixer	gsc	[65], [194], [195]
bix2	gsc	[65], [97]
gsc	gsc, otx2*, pitx2*, t, ventx1, ventx2, wnt8a	[18], [28], [90], [137], [196]
otx2	cer1, gsc, hhex, t	[32], [90], [153], [166], [197]
lhx1	cer1, gsc, hhex	[166], [190], [197]
hhex	gsc, nodal1*, nodal2*	[198], [199], [200]
gata4	hnf1b*	[106], [107], [201]
gata5	hnf1b*, sox17a*	[106], [107]
gata6	hnf1b*, sox17a*	[107], [201]
gata2	ventx1	[202]
sox17b	foxa1*, foxa2*, hnf1b*, sox17a	[119], [120], [121], [176]
hnf1b	lhx1	[203]
t	bix1	[96]
ventx1	foxa4	[133]

Table 4.1. *Summary of direct and putative connections between network transcription factors*. See also Table S2 for additional connection details. Putative direct targets are denoted with an asterisk (*).

References

[1] M. Levine. Transcriptional enhancers in animal development and evolution.

Curr. Biol., 20 (2010), pp. R754-R763, 10.1016/j.cub.2010.06.070

[2] E.H. Davidson, J.P. Rast, P. Oliveri, A. Ransick, C. Calestani, C.-H. Yuh, *et al.* A genomic regulatory network for development. *Science*, 295 (2002), pp. 1669-1678, 10.1126/science.1069883

[3] M. Levine, E.H. Davidson. Gene regulatory networks for development. *Proc. Natl. Acad. Sci*. U. S. A., 102 (2005), pp. 4936-4942, 10.1073/pnas.0408031102

[4] M. Sakabe, H. Matsui, H. Sakata, K. Ando, T. Yamagishi, Y. Nakajima. Understanding heart development and congenital heart defects through developmental biology: a segmental approach. *Congenit. Anom.* (Kyoto), 45 (2005), pp. 107-118, 10.1111/j.1741-4520.2005.00079.x

[5] J.R. Spence, J.M. Wells. Translational embryology: using embryonic principles to generate pancreatic endocrine cells from embryonic stem cells. *Dev. Dyn.*, 236 (2007), pp. 3218-3227, 10.1002/dvdy.21366

[6] M. Loose, R. Patient. A genetic regulatory network for Xenopus mesendoderm formation. *Dev. Biol.*, 271 (2004), pp. 467-478, 10.1016/j.ydbio.2004.04.014

[7] T. Koide, T. Hayata, K.W.Y. Cho. Xenopus as a model system to study transcriptional regulatory networks. *Proc. Natl. Acad. Sci. U. S. A.*, 102 (2005), pp. 4943-4948, 10.1073/pnas.0408125102

[8] J. Heasman. Patterning the early Xenopus embryo. *Development*, 133 (2006), pp. 1205-1217, 10.1242/dev.02304

[9] A.M. Zorn, J.M. Wells. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.*, 25 (2009), pp. 221-251, 10.1146/annurev.cellbio.042308.113344

[10] M. Flachsova, R. Sindelka, M. Kubista. Single blastomere expression profiling of Xenopus laevis embryos of 8 to 32-cells reveals developmental asymmetry. *Sci. Rep.*, 3 (2013), 10.1038/srep02278

[11] P.A. Grant, B. Yan, M.A. Johnson, D.L.E. Johnson, S.A. Moody. Novel animal pole-enriched maternal mRNAs are preferentially expressed in neural ectoderm. *Dev. Dyn.*, 243 (2014), pp. 478-496, 10.1002/dvdy.24082

[12] E. De Domenico, N.D.L. Owens, I.M. Grant, R. Gomes-Faria, M.J. Gilchrist. Molecular asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere transcript sequencing. *Dev. Biol.*, 408 (2015), pp. 252-268, 10.1016/j.ydbio.2015.06.010

[13] S. Gaur, M. Mandelbaum, M. Herold, H.D. Majumdar, K.M. Neilson, T.M. Maynard, *et al.* Neural transcription factors bias cleavage stage blastomeres to give rise to neural ectoderm. *Genesis*, 54 (2016), pp. 334-349, 10.1002/dvg.22943

[14] I.L. Blitz, K.D. Paraiso, I. Patrushev, W.T.Y. Chiu, K.W.Y. Cho, M.J. Gilchrist. A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. *Developmental Biology*. (in press)

http://dx.doi.org/10.1016/j.ydbio.2016.07.002.

[15] Y. Ding, G. Colozza, K. Zhang, Y. Moriyama, D. Ploper, E.A. Sosa, et al. Genome-wide analysis of dorsal and ventral transcriptomes of the Xenopus laevis gastrula. *Developmental Biology*. (in press) http://dx.doi.org/10.1016/j.ydbio.2016.02.032.

[16] I.K. Popov, T. Kwon, D.K., Crossman, M.R. Crowley, J.B. Wallingford, C. Chang. Identification of new regulators of embryonic patterning and morphogenesis in Xenopus gastrulae by RNA sequencing. *Developmental Biology* (in press)

http://dx.doi.org/10.1016/j.ydbio.2016.05.014.

[17] E. Li, E.H. Davidson. Building developmental gene regulatory networks. *Birth Defects Res. C E Embryo Today*, 87 (2009), pp. 123-130, 10.1002/bdrc.20152

[18] V. Danilov, M. Blum, A. Schweickert, M. Campione, H. Steinbeisser. Negative autoregulation of the organizer-specific homeobox gene goosecoid. *J. Biol. Chem.*, 273 (1998), pp. 627-635

[19] S.S. Paranjpe, U.G. Jacobi, S.J. van Heeringen, G.J.C. Veenstra. A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development. *BMC Genomics*, 14 (2013), p. 762, 10.1186/1471-2164-14-762

[20] M.H. Tan, K.F. Au, A.L. Yablonovitch, A.E. Wills, J. Chuang, J.C. Baker, *et al.* RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. *Genome Res.*, 23 (2013), pp. 201-216, 10.1101/gr.141424.112

[21] C. Collart, N.D.L. Owens, L. Bhaw-Rosun, B. Cooper, E. De Domenico, I.Patrushev, *et al.* High-resolution analysis of gene activity during the Xenopus mid-blastula transition. *Development*, 141 (2014), pp. 1927-1939, 10.1242/dev.102012

[22] N.D.L. Owens, I.L. Blitz, M.A. Lane, I. Patrushev, J.D. Overton, M.J. Gilchrist, *et al.* Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Reports*, 14 (2016), pp. 632-647, 10.1016/j.celrep.2015.12.050

[23] R.A.S. Kjolby, R.M. Harland. Genome-wide identification of Wnt/β-catenin transcriptional targets during Xenopus gastrulation. *Developmental Biology* (in press). http://dx.doi.org/10.1016/j.ydbio.2016.03.021.

[24] Y. Nakamura, E. de Paiva Alves, G.J.C. Veenstra, S. Hoppler. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules. *Development*, 143 (2016), pp. 1914-1925, 10.1242/dev.131664

[25] G.E. Gentsch, N.D.L. Owens, S.R. Martin, P. Piccinelli, T. Faial, M.W.B. Trotter, *et al.* In vivo T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency. *Cell Rep.*, 4 (2013), pp. 1185-1196, 10.1016/j.celrep.2013.08.012

[26] W.T. Chiu, R. Charney Le, I.L. Blitz, M.B. Fish, Y. Li, J. Biesinger, *et al.* Genome-wide view of TGF/Foxh1 regulation of the early mesendoderm program. *Development*, 141 (2014), pp. 4537-4547, 10.1242/dev.107227

[27] R. Gupta, A. Wills, D. Ucar, J. Baker. Developmental enhancers are marked independently of zygotic Nodal signals in Xenopus. *Dev. Biol.*, 395 (2014), pp. 38-49, 10.1016/j.ydbio.2014.08.034

[28] Y. Yasuoka, Y. Suzuki, S. Takahashi, H. Someya, N. Sudou, Y. Haramoto, *et al.* Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification. *Nat. Commun.*, 5 (2014), p. 4322, 10.1038/ncomms5322

[29] B. Langmead, S.L. Salzberg. Fast gapped-read alignment with bowtie 2.

Nat Meth., 9 (2012), pp. 357-359, 10.1038/nmeth.1923

[30] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, *et al.* Model-based analysis of chIP-Seq (MACS). *Genome Biol.*, 9 (2008), p. R137, 10.1186/gb-2008-9-9-r137

[31] A.R. Quinlan, I.M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26 (2010), pp. 841-842, 10.1093/bioinformatics/btq033

[32] S. Yamamoto, H. Hikasa, H. Ono, M. Taira. Molecular link in the sequential induction of the Spemann organizer: direct activation of the cerberus gene by Xlim-1, Xotx2, Mix.1, and Siamois, immediately downstream from Nodal and Wnt signaling. *Dev. Biol.*, 257 (2003), pp. 190-204, 10.1016/S0012-1606(03)00034-4

[33] N. Sudou, S. Yamamoto, H. Ogino, M. Taira. Dynamic in vivo binding of transcription factors to cis-regulatory modules of cer and gsc in the stepwise formation of the Spemann-Mangold organizer. *Development*, 139 (2012), pp. 1651-1661, 10.1242/dev.068395

[34] W.J.R. Longabaugh, E.H. Davidson, H. Bolouri. Computational representation of developmental genetic regulatory networks. *Dev. Biol.*, 283 (2005), pp. 1-16, 10.1016/j.ydbio.2005.04.023

[35] K.D. Lustig, K.L. Kroll, E.E. Sun, M.W. Kirschner. Expression cloning of a Xenopus T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation. *Development*, 122 (1996), pp. 4001-4012

[36] F. Stennard, G. Carnac, J.B. Gurdon. The Xenopus T-box gene Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation. *Development*, 122 (1996), pp. 4179-4188

[37] J. Zhang, M.L. King. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. *Development*, 122 (1996), pp. 4119-4129

[38] J. Zhang, D.W. Houston, M.L. King, C. Payne, C. Wylie, J. Heasman. The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. *Cell*, 94 (1998), pp. 515-524

[39] M. Kofron, T. Demel, J. Xanthos, J. Lohr, B. Sun, H. Sive, *et al.* Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors. *Development*, 126 (1999), pp. 5759-5770

[40] D. Clements, R.V. Friday, H.R. Woodland. Mode of action of VegT in mesoderm and endoderm formation. *Development*, 126 (1999), pp. 4903-4911

[41] J.B. Xanthos, M. Kofron, C. Wylie, J. Heasman. Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis. *Development*, 128 (2001), pp. 167-180

[42] S. Piccolo, E. Agius, L. Leyns, S. Bhattacharyya, H. Grunz, T. Bouwmeester, *et al.* The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature*, 397 (1999), pp. 707-710, 10.1038/17820

[43]K. Ryan, N. Garrett, A. Mitchell, J.B. Gurdon. Eomesodermin, a key early gene in xenopus mesoderm differentiation. *Cell*, 87 (1996), pp. 989-1000, 10.1016/S0092-8674(00)81794-8

[44] F.L. Conlon, L. Fairclough, B.M. Price, E.S. Casey, J.C. Smith. Determinants of T box protein specificity. *Development*, 128 (2001), pp. 3749-3758

[45] C.R.R. Bjornson, K.J.P. Griffin, G.H. Farr, A. Terashima, C. Himeda, Y. Kikuchi, *et al.* Eomesodermin is a localized maternal determinant required for endoderm induction in zebrafish. *Dev. Cell*, 9 (2005), pp. 523-533, 10.1016/j.devcel.2005.08.010

[46] A.C. Nelson, S.J. Cutty, M. Niini, D.L. Stemple, P. Flicek, C. Houart, *et al.* Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. *BMC Biol.*, 12 (2014), p. 81, 10.1186/s12915-014-0081-5

[47] P. Xu, G. Zhu, Y. Wang, J. Sun, X. Liu, Y.-G. Chen, *et al.* Maternal Eomesodermin regulates zygotic nodal gene expression for mesendoderm induction in zebrafish embryos. *J. Mol. Cell Biol.*, 6 (2014), pp. 272-285, 10.1093/jmcb/mju028

[48] A.K.K. Teo, S.J. Arnold, M.W.B. Trotter, S. Brown, L.T. Ang, Z. Chng, *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes. Dev.*, 25 (2011), pp. 238-250, 10.1101/gad.607311

[49] C. Zhang, T. Basta, M.W. Klymkowsky. SOX7 and SOX18 are essential for cardiogenesis in Xenopus. *Dev. Dyn.*, 234 (2005), pp. 878-891, 10.1002/dvdy.20565

[50] S.I. Osada, C.V. Wright. Xenopus nodal-related signaling is essential for mesendodermal patterning during early embryogenesis. *Development*, 126 (1999), pp. 3229-3240, 10.1073/pnas.88.15.6511

[51] B.I. Sun, S.M. Bush, L.A. Collins-Racie, E.R. LaVallie, E.A. DiBlasio-Smith, N.M.Wolfman, *et al.* derrière: a TGF-beta family member required for posterior development in Xenopus. *Development*, 126 (1999), pp. 1467-1482

[52] E. Agius, M. Oelgeschläger, O. Wessely, C. Kemp, E.M. De Robertis. Endodermal Nodal-related signals and mesoderm induction in Xenopus. *Development*, 127 (2000), pp. 1173-1183

[53] Y. Onuma, S. Takahashi, C. Yokota, M. Asashima. Multiple nodal-related genes act coordinately in Xenopus embryogenesis. *Dev. Biol.*, 241 (2002), pp. 94-105, 10.1006/dbio.2001.0493

[54] D.M. Ho, J. Chan, P. Bayliss, M. Whitman. Inhibitor-resistant type I receptors reveal specific requirements for TGF-beta signaling in vivo. *Dev. Biol.*, 295 (2006), pp. 730-742, 10.1016/j.ydbio.2006.03.050

[55] B. Feldman, M.A. Gates, E.S. Egan, S.T. Dougan, G. Rennebeck, H.I. Sirotkin, *et al.* Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature*, 395 (1998), pp. 181-185, 10.1038/26013

[56] F.L. Conlon, K.M. Lyons, N. Takaesu, K.S. Barth, A. Kispert, B. Herrmann, *et al.* A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development*, 120 (1994), pp. 1919-1928, 10.1016/0168-9525(94)90026-4

[57] D.L. Weeks, D.A. Melton. A maternal mRNA localized to the vegetal hemisphere in Xenopus eggs codes for a growth factor related to TGF-beta. *Cell*, 51 (1987), pp. 861-867

[58] B. Birsoy, M. Kofron, K. Schaible, C. Wylie, J. Heasman. Vg 1 is an essential signaling molecule in Xenopus development. *Development*, 133 (2006), pp. 15-20, 10.1242/dev.02144

[59] C. Kiecker, T. Bates, E. Bell. Molecular specification of germ layers in vertebrate embryos. *Cell. Mol. Life Sci.*, 73 (2015), pp. 923-947, 10.1007/s00018-015-2092-y

[60] J. Skirkanich, G. Luxardi, J. Yang, L. Kodjabachian, P.S. Klein. An essential role for transcription before the MBT in Xenopus laevis. *Dev. Biol.*, 357 (2011), pp. 478-491, 10.1016/j.ydbio.2011.06.010

[61] J. Yang, C. Tan, R.S. Darken, P.A. Wilson, P.S. Klein. β-Catenin/Tcf-regulated transcription prior to the midblastula transition. *Development* (2002), 10.1242/dev.00150

[62] S.I. Osada, Y. Saijoh, A. Frisch, C.Y. Yeo, H. Adachi, M. Watanabe, *et al.* Activin/nodal responsiveness and asymmetric expression of a Xenopus nodal-related gene converge on a FAST-regulated module in intron 1. *Development*, 127 (2000), pp. 2503-2514

[63] X. Chen, M.J. Rubock, M. Whitman. A transcriptional partner for MAD proteins in TGF-β signalling. *Nature*, 383 (1996), pp. 691-696, 10.1038/383691a0

[64] X. Chen, E. Weisberg, V. Fridmacher, M. Watanabe, G. Naco, M. Whitman. Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature*, 389 (1997), pp. 85-89, 10.1038/38008

[65] S. Germain, M. Howell, G.M. Esslemont, C.S. Hill. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. *Genes. Dev.*, 14 (2000), pp. 435-451, 10.1101/gad.14.4.435

[66] M. Howell, G.J. Inman, C.S. Hill. A novel Xenopus Smad-interacting forkhead transcription factor (XFast-3) cooperates with XFast-1 in regulating gastrulation movements. *Development*, 129 (2002), pp. 2823-2834

[67] C. Ring, S. Ogata, L. Meek, J. Song, T. Ohta, K. Miyazono, *et al.* The role of a Williams-Beuren syndrome-associated helix-loop-helix domain-containing transcription factor in activin/nodal signaling. *Genes. Dev.*, 16 (2002), pp. 820-835, 10.1101/gad.963802

[68] M. Cordenonsi, S. Dupont, S. Maretto, A. Insinga, C. Imbriano, S. Piccolo. Links between tumor suppressors p53 is required for TGF-β gene responses by cooperating with smads. *Cell*, 113 (2003), pp. 301-314, 10.1016/s0092-8674(03)00308-8

[69] M. Ku, S.Y. Sokol, J. Wu, M.I. Tussie-Luna, A.L. Roy, A. Hata. Positive and negative regulation of the transforming growth Factor/Activin target gene goosecoid by the TFII-I family of transcription factors. *Mol. Cell. Biol.*, 25 (2005), pp. 7144-7157, 10.1128/MCB.25.16.7144-7157.2005

[70] C.E. Slagle, T. Aoki, R.D. Burdine. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and Eomesodermin. *PLoS Genet.*, 7 (2011), p. e1002072, 10.1371/journal.pgen.1002072

[71] S.-J. Yoon, A.E. Wills, E. Chuong, R. Gupta, J.C. Baker. HEB and E2A function as SMAD/FOXH1 cofactors. *Genes. Dev.*, 25 (2011), pp. 1654-1661, 10.1101/gad.16800511

[72] M. Kofron, H. Puck, H. Standley, C. Wylie, R. Old, M. Whitman, *et al.* New roles for FoxH1 in patterning the early embryo. *Development*, 131 (2004), pp. 5065-5078, 10.1242/dev.01396

[73] H.-M. Pogoda, L. Solnica-Krezel, W. Driever, D. Meyer. The zebrafish forkhead transcription factor FoxH1/Fast1 is a modulator of Nodal signaling required for organizer formation. *Curr. Biol.*, 10 (2000), pp. 1041-1049, 10.1016/s0960-9822(00)00669-2

[74] H.I. Sirotkin, M.A. Gates, P.D. Kelly, A.F. Schier, W.S. Talbot. Fast1 is required for the development of dorsal axial structures in zebrafish. *Curr. Biol.*, 10 (2000), pp. 1051-1054, 10.1016/s0960-9822(00)00679-5

[75] P.A. Hoodless, M. Pye, C. Chazaud, E. Labbé, L. Attisano, J. Rossant, *et al.* FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes. Dev.*, 15 (2001), pp. 1257-1271, 10.1101/gad.881501

[76] M. Yamamoto, C. Meno, Y. Sakai, H. Shiratori. The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes. Dev.*, 15 (2001), pp. 1242-1256, 10.1101/gad.883901

[77] D.P. Norris, J. Brennan, E.K. Bikoff, E.J. Robertson. The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo. *Development*, 129 (2002), pp. 3455-3468

[78] S.W. Kim, S.-J. Yoon, E. Chuong, C. Oyolu, A.E. Wills, R. Gupta, *et al.* Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.*, 357 (2011), pp. 492-504, 10.1016/j.ydbio.2011.06.009

[79] S. Faure, M.A. Lee, T. Keller, P. ten Dijke, M. Whitman. Endogenous patterns of TGFbeta superfamily signaling during early Xenopus development. *Development*, 127 (2000), pp. 2917-2931

[80] M. Ku, D.A. Melton. Xwnt-11: a maternally expressed Xenopus wnt gene.

Development, 119 (1993), pp. 1161-1173

[81] C.A. Larabell, M. Torres, B.A. Rowning, C. Yost, J.R. Miller, M. Wu, *et al.* Establishment of the dorso-ventral axis in Xenopus embryos is presaged by early asymmetries in beta-catenin that are modulated by the Wnt signaling pathway. *J. Cell Biol.*, 136 (1997), pp. 1123-1136

[82] Q. Tao, C. Yokota, H. Puck, M. Kofron, B. Birsoy, D. Yan, *et al.* Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos. *Cell*, 120 (2005), pp. 857-871, 10.1016/j.cell.2005.01.013

[83] M. Brannon, M. Gomperts, L. Sumoy, R.T. Moon, D. Kimelman. A beta-catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in Xenopus. *Genes. Dev.*, 11 (1997), pp. 2359-2370

[84] M.N. Laurent, I.L. Blitz, C. Hashimoto, U. Rothbächer, K.W. Cho. The Xenopus homeobox gene twin mediates Wnt induction of goosecoid in establishment of Spemann's organizer. *Development*, 124 (1997), pp. 4905-4916

[85] J.B. Xanthos, M. Kofron, Q. Tao, K. Schaible, C. Wylie, J. Heasman. The roles of three signaling pathways in the formation and function of the Spemann Organizer. *Development*, 129 (2002), pp. 4027-4043

[86] D. Kimelman, J.L. Christian, R.T. Moon. Synergistic principles of development: overlapping patterning systems in Xenopus mesoderm induction. *Development*, 116 (1992), pp. 1-9

[87] T. Watabe, S. Kim, A. Candia, U. Rothbächer, C. Hashimoto, K. Inoue, *et al.* Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between Xenopus and mouse. *Genes. Dev.*, 9 (1995), pp. 3038-3050

[88] L.A. Pereira, M.S. Wong, S.M. Lim, E.G. Stanley, A.G. Elefanty. The Mix family of homeobox genes—Key regulators of mesendoderm formation during vertebrate development. *Dev. Biol.*, 367 (2012), pp. 163-177, 10.1016/j.ydbio.2012.04.033

[89] M. Watanabe, Y. Yasuoka, S. Mawaribuchi, A. Kuretani, M. Ito, M. Kondo, et al. Conservatism and variability of gene expression profiles among homeologous transcription factors in Xenopus laevis. *Developmental Biology* (in press).

http://dx.doi.org/10.1016/j.ydbio.2016.09.017.

[90] B.V. Latinkic, M. Umbhauer, K.A. Neal, W. Lerchner, J.C. Smith, V. Cunliffe. The Xenopus Brachyury promoter is activated by FGF and low concentrations of activin and suppressed by high concentrations of activin and by paired-type homeodomain proteins. *Genes. Dev.*, 11 (1997), pp. 3265-3276

[91] P. Lemaire, S. Darras, D. Caillol, L. Kodjabachian. A role for the vegetally expressed Xenopus gene Mix.1 in endoderm formation and in the restriction of mesoderm to the marginal zone. *Development*, 125 (1998), pp. 2371-2380

[92] A. Colas, J. Cartry, I. Buisson, M. Umbhauer, J.C. Smith, J.-F. Riou. Mix.1/2-dependent control of FGF availability during gastrulation is essential for pronephros development in Xenopus. *Dev. Biol.*, 320 (2008), pp. 351-365, 10.1016/j.ydbio.2008.05.547

[93] M. Artinger, I. Blitz, K. Inoue, U. Tran, K.W. Cho. Interaction of goosecoid and brachyury in Xenopus mesoderm patterning. *Mech. Dev.*, 65 (1997), pp. 187-196

[94] L.A. Pereira, M.S. Wong, A.K. Mossman, K. Sourris, M.E. Janes, K. Knezevic, *et al.* Pdgfrα and Flk1 are direct target genes of Mixl1 in differentiating embryonic stem cells. *Stem Cell Res.*, 8 (2012), pp. 165-179, 10.1016/j.scr.2011.09.007

[95] M. Kofron, C. Wylie, J. Heasman. The role of Mixer in patterning the early Xenopus embryo. *Development*, 131 (2004), pp. 2431-2441, 10.1242/dev.01132

[96] M. Tada, E.S. Casey, L. Fairclough, J.C. Smith. Bix1 a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. *Development*, 125 (1998), pp. 3997-4006

[97] V. Ecochard, C. Cayrol, S. Rey, F. Foulquier, D. Caillol, P. Lemaire, *et al.* A novel Xenopus mix-like gene milk involved in the control of the endomesodermal fates. *Development*, 125 (1998), pp. 2577-2585

[98] T. Okumura, A. Matsumoto, T. Tanimura, R. Murakami. An endoderm-specific GATA factor gene, dGATAe, is required for the terminal differentiation of the Drosophila endoderm. *Dev. Biol.*, 278 (2005), pp. 576-586, 10.1016/j.ydbio.2004.11.021

[99] T. Okumura, R. Tajiri, T. Kojima, K. Saigo, R. Murakami. GATAe-dependent and −independent expressions of genes in the differentiated endodermal midgut of Drosophila. *Gene Expr. Patterns*, 7 (2007), pp. 178-186, 10.1016/j.modgep.2006.07.001

[100] E.M. Sommermann, K.R. Strohmaier, M.F. Maduro, J.H. Rothman. Endoderm development in Caenorhabditis elegans: the synergistic action of ELT-2 and −7 mediates the specification → differentiation transition. *Dev. Biol.*, 347 (2010), pp. 154-166, 10.1016/j.ydbio.2010.08.020

[101] M.E. Boeck, T. Boyle, Z. Bao, J. Murray, B. Mericle, R. WaterstonSpecific roles for the GATA transcription factors end-1 and end-3 during C. elegans E-lineage development. *Dev. Biol.*, 358 (2011), pp. 345-355, 10.1016/j.ydbio.2011.08.002

[102] M.F. Maduro. Gut development in C. elegans. *Seminars in Cell and Developmental Biology* (in press). http://dx.doi.org/10.1016/j.semcdb.2017.01.001.

[103] I.S. Peter, E.H. Davidson. The endoderm gene regulatory network in sea urchin embryos up to mid-blastula stage. *Dev. Biol.*, 340 (2010), pp. 188-199, 10.1016/j.ydbio.2009.10.037

[104] J. Fujikura, E. Yamato, S. Yonemura, K. Hosoda, S. Masui, K. Nakao, *et al.* Differentiation of embryonic stem cells is induced by GATA factors. *Genes. Dev.*, 16 (2002), pp. 784-789, 10.1101/gad.968802

[105] D. Shimosato, M. Shiki, H. Niwa. Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. *BMC Dev.* Biol., 7 (2007), p. 80, 10.1186/1471-213x-7-80

[106] H. Weber, C.E. Symes, M.E. Walmsley, A.R. Rodaway, R.K. Patient. A role for GATA5 in Xenopus endoderm specification. *Development*, 127 (2000), pp. 4345-4360

[107] B.A. Afouda, A. Ciau-Uitz, R. Patient. GATA4, 5 and 6 mediate TGFbeta maintenance of endodermal gene expression in Xenopus embryos. *Development*, 132 (2005), pp. 763-774, 10.1242/dev.01647

[108] X. Lou, A.R. Deshwar, J.G. Crump, I.C. Scott. Smarcd3b and Gata5 promote a cardiac progenitor fate in the zebrafish embryo. *Development*, 138 (2011), pp. 3113-3123, 10.1242/dev.064279

[109] W.-F. Tseng, T.-H. Jang, C.-B. Huang, C.-H. Yuh. An evolutionarily conserved kernel of gata5, gata6, otx2 and prdm1a operates in the formation of endoderm in zebrafish. *Dev. Biol.*, 357 (2011), pp. 541-557, 10.1016/j.ydbio.2011.06.040

[110] J.F. Reiter, Y. Kikuchi, D.Y. Stainier. Multiple roles for Gata5 in zebrafish endoderm formation. *Development*, 128 (2001), pp. 125-135

[111] H.K. Turbendian, M. Gordillo, S.-Y. Tsai, J. Lu, G. Kang, T.-C. Liu, *et al.* GATA factors efficiently direct cardiac fate from embryonic stem cells. *Development*, 140 (2013), pp. 1639-1644, 10.1242/dev.093260

[112] A.J. Watt, R. Zhao, J. Li, S.A. Duncan. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev. Biol.*, 7 (2007), p. 37, 10.1186/1471-213x-7-37

[113] A. Rojas, W. Schachterle, S.-M. Xu, B.L. Black. An endoderm-specific transcriptional enhancer from the mouse Gata4 gene requires GATA and homeodomain protein-binding sites for function in vivo. *Dev. Dyn.*, 238 (2009), pp. 2588-2598, 10.1002/dvdy.22091

[114] A. Rojas, S. De Val, A.B. Heidt, S.-M. Xu, J. Bristow, B.L. Black. Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development*, 132 (2005), pp. 3405-3417, 10.1242/dev.01913

[115] C. Hudson, D. Clements, R.V. Friday, D. Stott, H.R. Woodland. Xsox17α and –β mediate endoderm formation in xenopus. *Cell*, 91 (1997), pp. 397-405, 10.1016/S0092-8674(00)80423-7

[116] J. Alexander, D. Stainier. A molecular pathway leading to endoderm formation in zebrafish. *Curr. Biol.* (1999), 10.1016/S0960-9822(00)80016-0

[117] M. Kanai-Azuma, Y. Kanai, J.M. Gad, Y. Tajima, C. Taya, M. Kurohmaru, *et al.* Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development*, 129 (2002), pp. 2367-2379

[118] C.A. Séguin, J.S. Draper, A. Nagy, J. Rossant. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell*, 3 (2008), pp. 182-195, 10.1016/j.stem.2008.06.018

[119] D. Sinner, S. Rankin, M. Lee, A.M. Zorn. Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development*, 131 (2004), pp. 3069-3080, 10.1242/dev.01176

[120] K. Dickinson, J. Leonard, J.C. Baker. Genomic profiling of Mixer and Sox17β targets duringXenopus endoderm development. *Dev. Dyn.*, 235 (2006), pp. 368-381, 10.1002/dvdy.20636

[121] L. Howard, M. Rex, D. Clements, H.R. Woodland. Regulation of the Xenopus Xsox17alpha(1) promoter by co-operating VegT and Sox17 sites. *Dev. Biol.*, 310 (2007), pp. 402-415, 10.1016/j.ydbio.2007.07.028

[122] M. Shimoda, K. Kanai-Azuma, S. Miyazaki, Y. Kanai, M. Monden, *et al.* Sox17 plays a substantial role in late-stage differentiation of the extraembryonic endoderm in vitro. *J. Cell Sci.*, 120 (2007), pp. 3859-3869, 10.1242/jcs.007856

[123] J.R. Friedman, K.H. Kaestner. The Foxa family of transcription factors in development and metabolism. *Cell. Mol. Life Sci.*, 63 (2006), pp. 2317-2328, 10.1007/s00018-006-6095-6

[124] S.B.-T. de-Leon. The conserved role and divergent regulation of foxa, a pan-eumetazoan developmental regulatory gene. *Dev. Biol.*, 357 (2011), pp. 21-26, 10.1016/j.ydbio.2010.11.027

[125] D.C. Weinstein, A. Ruiz Altaba, W.S. Chen, P. Hoodless, V.R. Prezioso, T.M.Jessell, *et al.* The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell*, 78 (1994), pp. 575-588

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

[127] C.S. Lee, J.R. Friedman, J.T. Fulmer, K.H. Kaestner. The initiation of liver development is dependent on Foxa transcription factors. *Nature*, 435 (2005), pp. 944-947, 10.1038/nature03649

[128] L.A. Cirillo, F.R. Lin, I. Cuesta, D. Friedman, M. Jarnik, K.S. Zaret. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell.*, 9 (2002), pp. 279-289, 10.1016/s1097-2765(02)00459-8

[129] M. Iwafuchi-Doi, K.S. Zaret. Cell fate control by pioneer transcription factors. *Development*, 143 (2016), pp. 1833-1837, 10.1242/dev.133900

[130] C. Suri, T. Haremaki, D.C. Weinstein. Inhibition of mesodermal fate by Xenopus HNF3β/FoxA2. *Dev. Biol.*, 265 (2004), pp. 90-104, 10.1016/j.ydbio.2003.09.017

[131] P. Oliveri, K.D. Walton, E.H. Davidson, D.R. McClay. Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo. *Development*, 133 (2006), pp. 4173-4181, 10.1242/dev.02577

[132] S. Murgan, A.M. Castro Colabianchi, R.J. Monti, L.E. Boyadjián López, C.E.Aguirre, E.G. Stivala, *et al.* FoxA4 favours notochord formation by inhibiting contiguous mesodermal fates and restricts anterior neural development in Xenopus embryos. *PLoS One*, 9 (2014), p. e110559, 10.1371/journal.pone.0110559

[133] H. Friedle, S. Rastegar, H. Paul, E. Kaufmann, W. Knochel. Xvent-1 mediates BMP-4-induced suppression of the dorsal-lip-specific early response gene XFD-1' in Xenopus embryos. *EMBO J.*, 17 (1998), pp. 2298-2307, 10.1093/emboj/17.8.2298

[134] U. Alon. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 8 (2007), pp. 450-461, 10.1038/nrg2102

[135] G. Luxardi, L. Marchal, V. Thomé, L. Kodjabachian. Distinct Xenopus Nodal ligands sequentially induce mesendoderm and control gastrulation movements in parallel to the Wnt/PCP pathway. *Development*, 137 (2010), pp. 417-426, 10.1242/dev.039735

[136] K.A. Henningfeld, H. Friedle, S. Rastegar, W. Knöchel. Autoregulation of Xvent-2B; direct interaction and functional cooperation of Xvent-2 and Smad1. *J. Biol. Chem.*, 277 (2002), pp. 2097-2103, 10.1074/jbc.M108524200

[137] V. Sander, B. Reversade, E.M. De Robertis. The opposing homeobox genes Goosecoid and Vent1/2 self-regulate Xenopus patterning. *EMBO J.*, 26 (2007), pp. 2955-2965, 10.1038/sj.emboj.7601705

[138] A. Schohl, F. Fagotto. A role for maternal β-catenin in early mesoderm induction in Xenopus. *EMBO J.*, 22 (2003), pp. 3303-3313, 10.1093/emboj/cdg328

[139] D. Onichtchouk, V. Gawantka, R. Dosch, H. Delius, K. Hirschfeld, C. Blumenstock, *et al.* The Xvent-2 homeobox gene is part of the BMP-4 signalling pathway controlling [correction of controling] dorsoventral patterning of Xenopus mesoderm. *Development*, 122 (1996), pp. 3045-3053

[140] M. Trindade, M. Tada, J.C. Smith. DNA-binding specificity and embryological function of Xom (Xvent-2). *Dev. Biol.*, 216 (1999), pp. 442-456, 10.1006/dbio.1999.9507

[141] A.M. Middleton, J.R. King, M. Loose. Bistability in a model of mesoderm and anterior mesendoderm specification in Xenopus laevis. *J. Theor. Biol.*, 260 (2009), pp. 41-55, 10.1016/j.jtbi.2009.05.016

[142] Y. Saka, J.C. Smith. A mechanism for the sharp transition of morphogen gradient interpretation in Xenopus. *BMC Dev. Biol.*, 7 (2007), p. 47, 10.1186/1471-213X-7-47

[143] A.M. Zorn, K. Butler, J.B. Gurdon. Anterior endomesoderm specification inXenopusby wnt/β-catenin and TGF-β signalling pathways. *Dev. Biol.*, 209 (1999), pp. 282-297, 10.1006/dbio.1999.9257

[144] A. Schuler-Metz, S. Knochel, E. Kaufmann, W. Knochel. The homeodomain transcription factor Xvent-2 mediates autocatalytic regulation of BMP-4 expression in Xenopus embryos. *J. Biol. Chem.*, 275 (2000), pp. 34365-34374, 10.1074/jbc.M003915200

[145] M. Rex, E. Hilton, R. Old. Multiple interactions between maternally-activated signalling pathways control Xenopus nodal-related genes. *Int. J. Dev. Biol.*, 46 (2002), pp. 217-226

[146] S.-W. Cha, E. Tadjuidje, Q. Tao, C. Wylie, J. Heasman. Wnt5a and Wnt11 interact in a maternal Dkk1-regulated fashion to activate both canonical and non-canonical signaling in Xenopus axis formation. *Development*, 135 (2008), pp. 3719-3729, 10.1242/dev.029025

[147] S.A. Blythe, S.-W. Cha, E. Tadjuidje, J. Heasman, P.S. Klein. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Dev. Cell*, 19 (2010), pp. 220-231, 10.1016/j.devcel.2010.07.007

[148] S. Mangan, U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U. S. A.*, 100 (2003), pp. 11980-11985, 10.1073/pnas.2133841100

[149] E. Galán-Vásquez, B. Luna, A. Martínez-Antonio. The regulatory network of pseudomonas aeruginosa. *Microb. Inf. Exp.*, 1 (2011), p. 3, 10.1186/2042-5783-1-3

[150] M.-S. Kim, J.-R. Kim, K.-H. Cho. Dynamic network rewiring determines temporal regulatory functions in Drosophila melanogaster development processes. *Bioessays*, 32 (2010), pp. 505-513, 10.1002/bies.200900169

[151] M.-S. Kim, J.-R. Kim, D. Kim, A.D. Lander, K.-H. Cho. Spatiotemporal network motif reveals the biological traits of developmental gene regulatory networks in Drosophila melanogaster. *BMC Syst. Biol.*, 6 (2012), p. 31, 10.1186/1752-0509-6-31

[152] B.V. Latinkic, J.C. Smith. Goosecoid and mix.1 repress Brachyury expression and are required for head formation in Xenopus. *Development*, 126 (1999), pp. 1769-1779

[153] C. Carron, A. Bourdelas, H.-Y. Li, J.-C. Boucaut, D.-L. Shi. Antagonistic interaction between IGF and Wnt/JNK signaling in convergent extension in Xenopus embryo. *Mech. Dev.*, 122 (2005), pp. 1234-1247, 10.1016/j.mod.2005.06.007

[154] C. Zhang, T. Basta, E.D. Jensen, M.W. Klymkowsky. The beta-catenin/VegT-regulated early zygotic gene Xnr5 is a direct target of SOX3 regulation. *Development*, 130 (2003), pp. 5609-5624, 10.1242/dev.00798

[155] F. Faunes, L.I. Almonacid, F. Melo, J. Larraín. Characterization of small RNAs in X. tropicalis gastrulae. *Genesis*, 50 (2012), pp. 572-583, 10.1002/dvg.22027

[156] J.L. Harding, S. Horswell, C. Heliot, J. Armisen, L.B. Zimmerman, N.M. Luscombe,*et al.* Small RNA profiling of Xenopus embryos reveals novel miRNAs and a new class of small RNAs derived from intronic transposable elements. *Genome Res.*, 24 (2014), pp. 96-106, 10.1101/gr.144469.112

[157] A. Ahmed, N.J. Ward, S. Moxon, S. Lopez-Gomollon, C. Viaut, M.L. Tomlinson, *et al.* A database of microRNA expression patterns in xenopus laevis. *PLoS One*, 10 (2015), p. e0138313, 10.1371/journal.pone.0138313

[158] C. Liu, C.-H. Lou, V. Shah, R. Ritter, J. Talley, B. Soibam, *et al.* Identification of microRNAs and microRNA targets in Xenopus gastrulae: the role of miR-26 in the regulation of Smad1. *Dev. Biol.*, 409 (2016), pp. 26-38, 10.1016/j.ydbio.2015.11.005

[159] V.V. Shah, B. Soibam, R.A. Ritter, A. Benham, J. Oomen, A.K. Sater. MicroRNAs and ectodermal specification I. Identification of miRs and miR-targeted mRNAs in early anterior neural and epidermal ectoderm. *Developmental Biology* (in press)

http://dx.doi.org/10.1016/j.ydbio.2016.08.017.

[160] A. Necsulea, M. Soumillon, M. Warnefors, A. Liechti, T. Daish, U. Zeller, *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 505 (2014), pp. 635-640, 10.1038/nature12943

[161] E. Forouzmand, N.D.L. Owens, I.L. Blitz, K.D. Paraiso, M.K. Khokha, M.J. Gilchrist, et al. Developmentally regulated long non-coding RNAs in Xenopus tropicalis. *Developmental Biology* (in press). http://dx.doi.org/10.1016/j.ydbio.2016.06.016.

[162] A. Rosa, F.M. Spagnoli, A.H. Brivanlou. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. *Dev. Cell*, 16 (2009), pp. 517-527, 10.1016/j.devcel.2009.02.007

[163] M. Paulsen, S. Legewie, R. Eils. Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signaling range and canalizes development. *Proceedings of the National Academy of Sciences*, *108*(25), 10202-10207 (2011). http://dx.doi.org/10.1074/jbc.M113.509331.

[164] M. Heasman. β catenin signaling activity dissected in the early xenopus embryo: a novel antisense approach. *Dev. Biol.*, 222 (2000), pp. 124-134, 10.1006/dbio.2000.9720

[165] E. Hilton, M. Rex, R. Old. VegT activation of the early zygotic gene Xnr5 requires lifting of Tcf-mediated repression in the Xenopus blastula. *Mech. Dev.*, 120 (2003), pp. 1127-1138, 10.1016/j.mod.2003.08.004

[166] S.A. Rankin, J. Kormish, M. Kofron, A. Jegga, A.M. Zorn. A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer. *Dev. Biol.*, 351 (2011), pp. 297-310, 10.1016/j.ydbio.2010.11.037

[167] D.W. Houston, M. Kofron, E. Resnik, R. Langland, O. Destree, C. Wylie, *et al.* Repression of organizer genes in dorsal and ventral Xenopus cells mediated by maternal XTcf3. *Development*, 129 (2002), pp. 4015-4025

[168] R.J. White, B.I. Sun, H.L. Sive, J.C. Smith. Direct and indirect regulation of derrière, a Xenopus mesoderm-inducing factor, by VegT. *Development*, 129 (2002), pp. 4867-4876

[169] D. Clements, I. Cameleyre, H.R. Woodland. Redundant early and overlapping larval roles of Xsox17 subgroup genes in Xenopus endoderm development. *Mech. Dev.*, 120 (2003), pp. 337-348, 10.1016/s0925-4773(02)00450-1

[170] M. Taverner, Y. Kofron, C. Shin, M.J. Kabitschke, *et al.* Microarray-based identification of VegT targets in Xenopus. *Mech. Dev.*, 122 (2005), pp. 333-354, 10.1016/j.mod.2004.10.010

[171] C. Zhang, T. Basta, S.R. Fawcett, M.W. Klymkowsky. SOX7 is an immediate-early target of VegT and regulates Nodal-related gene expression in Xenopus. *Dev. Biol.*, 278 (2005), pp. 526-541, 10.1016/j.ydbio.2004.11.008

[172] Y. Cao, D. Siegel, C. Donow, S. Knöchel, L. Yuan, W. Knöchel. POU–V factors antagonize maternal VegT activity and β-Catenin signaling in Xenopus embryos. *EMBO J.*, 26 (2007), pp. 2942-2954, 10.1038/sj.emboj.7601736

[173] E. Labbé, C. Silvestri, P.A. Hoodless, J.L. Wrana, L. Attisano. Smad2 and Smad3 positively and negatively regulate TGF beta-dependent transcription through the forkhead DNA-binding protein FAST2. *Mol. Cell.*, 2 (1998), pp. 109-120

[174]C.Y. Yeo, X. Chen, M. Whitman. The role of FAST-1 and Smads in transcriptional regulation by activin during early Xenopus embryogenesis. *J. Biol. Chem.*, 274 (1999), pp. 26584-26590

[175] H. Shiratori, R. Sakuma, M. Watanabe, H. Hashiguchi, K. Mochida, Y. Sakai, *et al.* Two-step regulation of left-Right asymmetric expression of pitx2. *Mol. Cell.*, 7 (2001), pp. 137-149, 10.1016/s1097-2765(01)00162-9

[176] D. Sinner, P. Kirilenko, S. Rankin, E. Wei, L. Howard, M. Kofron, *et al.* Global analysis of the transcriptional network controlling Xenopus endoderm formation. *Development*, 133 (2006), pp. 1955-1966, 10.1242/dev.02358

[177] C.D. Reid, Y. Zhang, M.D. Sheets, D.S. Kessler. Transcriptional integration of Wnt and Nodal pathways in establishment of the Spemann organizer. *Dev. Biol.*, 368 (2012), pp. 231-241, 10.1016/j.ydbio.2012.05.018

[178] V. Gawantka, H. Delius, K. Hirschfeld, C. Blumenstock, C. Niehrs. Antagonizing the Spemann organizer: role of the homeobox gene Xvent-1. *EMBO J.*, 14 (1995), pp. 6268-6279

[179] G. Lagna, A. Hata, A. Hemmati-Brivanlou, J. Massagué. Partnership between DPC4 and SMAD proteins in TGF-β signalling pathways. *Nature*, 383 (1996), pp. 832-836, 10.1038/383832a0

[180] A.F. Candia, T. Watabe, S.H. Hawley, D. Onichtchouk, Y. Zhang, R. Derynck, *et al.* Cellular interpretation of multiple TGF-beta signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by Smads. *Development*, 124 (1997), pp. 4467-4480, 10.1101/gad.10.15.1880

[181] R. Ladher, T.J. Mohun, J.C. Smith, A.M. Snape. Xom: a Xenopus homeobox gene that mediates the early effects of BMP-4. *Development*, 122 (1996), pp. 2385-2394

[182] C. Zhang, T. Evans. BMP-like signals are required after the midblastula transition for blood cell development. *Dev. Genet.*, 18 (1996), pp. 267-278, 10.1002/(SICI)1520-6408(1996)18:3<267:AID-DVG7>3.0. CO;2-7

[183] K.A. Henningfeld, S. Rastegar, G. Adler, W. Knochel. Smad1 and Smad4 are components of the bone morphogenetic protein-4 (BMP-4)-induced transcription complex of the Xvent-2B promoter. *J. Biol. Chem.*, 275 (2000), pp. 21827-21835, 10.1074/jbc.M000978200

[184] H.-S. Lee, M.J. Park, S.-Y. Lee, Y.-S. Hwang, H. Lee, D.-H. Roh, *et al.* Transcriptional regulation of Xbr-1a/Xvent-2 homeobox gene: analysis of its promoter region. *Biochem. Biophys. Res. Commun.*, 298 (2002), pp. 815-823, 10.1016/s0006-291x(02)02570-6

[185] Y. Miyanaga, I. Torregroza, T. Evans. A maternal Smad protein regulates early embryonic apoptosis in Xenopus laevis. *Mol. Cell. Biol.*, 22 (2002), pp. 1317-1328, 10.1128/mcb.22.5.1317-1328.2002

[186] D.A. Peiffer, A. von Bubnoff, Y. Shin, A. Kitayama, M. Mochii, N. Ueno, *et al.* A Xenopus DNA microarray approach to identify novel direct BMP target genes involved in early embryonic development. *Dev. Dyn.*, 232 (2005), pp. 445-456, 10.1002/dvdy.20230

[187] I. Torregroza, T. Evans. Tid1 is a Smad-binding protein that can modulate Smad7 activity in developing embryos. *Biochem. J.*, 393 (2006), pp. 311-320, 10.1042/bj20050785

[188] A. Wills, K. Dickinson, M. Khokha, J.C. Baker. Bmp signaling is necessary and sufficient for ventrolateral endoderm specification in Xenopus. *Dev. Dyn.*, 237 (2008), pp. 2177-2186, 10.1002/dvdy.21631

[189] D.S. Kessler. Siamois is required for formation of Spemann's organizer. *Proc. Natl. Acad. Sci. U. S. A.*, 94 (1997), pp. 13017-13022, 10.1073/pnas.94.24.13017

[190] S. Yamamoto, H. Hikasa, H. Ono, M. Taira. Molecular link in the sequential induction of the Spemann organizer: direct activation of the cerberus gene by Xlim-1, Xotx2, Mix.1, and Siamois, immediately downstream from Nodal and Wnt signaling. *Dev. Biol.*, 257 (2003), pp. 190-204, 10.1016/s0012-1606(03)00034-4

[191] S. Bae, C.D. Reid, D.S. Kessler. Siamois and Twin are redundant and essential in formation of the Spemann organizer. *Dev. Biol.*, 352 (2011), pp. 367-381, 10.1016/j.ydbio.2011.01.034

[192] S.L. Klein, S.A. Moody. Early neural ectodermal genes are activated by Siamois and Twin during blastula stages. *Genesis*, 53 (2015), pp. 308-320, 10.1002/dvg.22854

[193] P.E. Mead, I.H. Brivanlou, C.M. Kelley, L.I. Zon. BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1. *Nature*, 382 (1996), pp. 357-360, 10.1038/382357a0

[194] T. Hayata, K. Tanegashima, S. Takahashi. Overexpression of the secreted factor Mig30 expressed in the Spemann organizer impairs morphogenetic movements during Xenopus gastrulation. *Mech. Dev.*, 112 (2002), pp. 37-51, 10.1016/s0925-4773(01)00638-4

[195] R.A. Randall, S. Germain, G.J. Inman. Different Smad2 partners bind a common hydrophobic pocket in Smad2 via a defined proline-rich motif.

The EMBO journal, 21(2002), pp.145-156

[196] J. Yao, D.S. Kessler. Goosecoid promotes head organizer activity by direct repression of Xwnt8 in Spemann's organizer. *Development* 128, no. 15 (2001): 2975-2987.

[197] T. Mochizuki, A.A. Karavanov, P.E. Curtiss, K.T. Ault, N. Sugimoto, T. Watabe, *et al.* Xlim-1 and LIM domain binding protein 1 cooperate with various transcription factors in the regulation of the goosecoid promoter. *Dev. Biol.*, 224 (2000), pp. 470-485, 10.1006/dbio.2000.9778

[198] J.M. Brickman, C.M. Jones, M. Clements, J.C. Smith, R.S. Beddington. Hex is a transcriptional repressor that contributes to anterior identity and suppresses Spemann organiser function. *Development*, 127 (2000), pp. 2303-2315

[199] L.E. Smithers, C.M. Jones. Xhex-expressing endodermal tissues are essential for anterior patterning in Xenopus. *Mech. Dev.*, 119 (2002), pp. 191-200, 10.1016/s0925-4773(02)00361-1

[200] A.L. Zamparini. Hex acts with −catenin to regulate anteroposterior patterning via a Groucho-related co-repressor and Nodal. *Development*, 133 (2006), pp. 3709-3722, 10.1242/dev.02516

[201] B.A. Afouda, J. Martin, F. Liu, A. Ciau-Uitz, R. Patient, S. Hoppler. GATA transcription factors integrate Wnt signalling during heart development. *Development*, 135 (2008), pp. 3185-3190, 10.1242/dev.026443

[202] H. Friedle, W. Knöchel. Cooperative interaction of Xvent-2 and GATA-2 in the activation of the ventral homeobox gene Xvent-1B. *J. Biol. Chem.*, 277 (2002), pp. 23872-23881, 10.1074/jbc.m201831200

[203] C. Drews, S. Senkel, G.U. Ryffel. The nephrogenic potential of the transcription factors osr1, osr2, hnf1b, lhx1 and pax8 assessed in Xenopus animal caps. *BMC Dev. Biol.*, 11 (2011), p. 5, 10.1186/1471-213x-11-5

[204] D. Onichtchouk, A. Glinka, C. Niehrs. Requirement for Xvent-1 and Xvent-2 gene function in dorsoventral patterning of Xenopus mesoderm. *Development*, 125 (1998), pp. 1447-1456

[205] A. Schuler-Metz, S. Knochel, E. Kaufmann, W. Knochel. The homeodomain transcription factor Xvent-2ediates autocatalytic regulation of BMP-4 expression in Xenopus embryos. *J. Biol. Chem.*, 275 (2000), pp. 34365-34374, 10.1074/jbc.M003915200

Chapter 5

Feedforward loops as a second tier control of the onset of zygotic gene activation

Abstract

Zygotic gene activation is a process in metazoan development in which the control of gene expression is transferred from maternal factors to zygotic factors. The timing of the onset of this process is highly regulated. While models exist to explain the timing, recent observations in transcriptomic datasets show multiple waves of zygotic transcription. This implies that in addition to current models, there exists a regulatory control to form the different waves. In the endoderm, the predominant model is a linear model whereby Vegt activates Nodal signaling, and Nodal signaling activates second wave genes. However, network motif analysis of the *Xenopus* mesendoderm network suggests a role for feed-forward loops to control this process. Using computational modeling, I validate that feed-forward loops can generate the multiple onsets of zygotic gene activation. Using genomics, I provide a mechanistic model of feed-forward loop formation seen using Vegt and Smad2/3 chromatin binding. Overall, the analysis in this chapter provides a useful genomic/systems biological approach to the study of the onset of ZGA.

Introduction

Zygotic gene activation (ZGA) is a period in embryonic development when the first genes from the zygotes genome are transcribed. This process is controlled temporally with the involvement of maternal transcription factors (TFs)[1-9]. While multiple models exist to explain the onset of ZGA, recent gene expression analysis showed that ZGA occurs in waves[10,11]. In *Xenopus*, this phenomenon has been called mid-blastula transition (MBT), when most genes are activated, and pre-MBT, when a small subset of genes is activated earlier[12]. Among the pre-MBT genes are important developmental genes that regulate formation of one of the three germ layers, such as *gdf3*, *nodal5* and *nodal6*, which encode ligands that activate the Activin/Nodal signaling pathway[13]. These findings suggest a second tier of ZGA control that differentiates pre-MBT and MBT gene transcription.

In the vegetal region of the early embryo, the predominant model of mesendodermal gene activation[14-17] involves the activity of the maternal vegetally-localized TF Vegt[18-21]. During cleavage and early blastula stages, the T-box TF Vegt directly activates the expression of the pre-MBT Nodal ligands *gdf3*[22-24], *nodal5*[25,26] and *nodal6*[27] (Table 2.4). Through activation of the Nodal signaling pathway using the TF Smad2/3, these ligands regulate the expression of second wave of zygotic genes, which are activated during mid-blastula or early gastrula[28]. Among these endodermal genes are genes activated during the second wave of zygotic transcription including *gata4*, *gata6*, *hnf1b* and *foxa4*[9,29]. The sequential activation by maternal TF of pre-MBT genes, and activation by pre-MBT genes of MBT

genes provides a convenient model to explain multiplicity in ZGA onset. However, Vegt ChIP-seq generated prior to ZGA reveals that Vegt could bind in regulatory regions near second wave zygotic genes (Fig 2.8). This suggests that Vegt can maintain control of gene expression during the activation of pre-MBT/Nodal and MBT/second wave genes, creating a positive feedforward loop (FFL). Theoretically, this motif had been shown to be able to regulate timing of gene expression[30] and has been attributed to timing of gene expression in early sea urchin development[31]. In addition to gene onset, FFLs provides additional advantages over the linear model as FFLs can provide control over the timing of gene transcription shut-down[30]. While the linear model is much simpler, the formation of FFLs provides an alternative control in generating the waves of gene expression seen in ZGA.

Here, I investigate the timing of ZGA using computational methods and genomic analysis. I hypothesize that FFLs act as a second tier of ZGA control. I compare the predominant model of gene activation (the linear model) and the FFL model, and find that computationally, these two models are capable of creating the gene expression waves during ZGA. From genomic analysis, I explore the mechanisms of FFL formation during these early stages of development, and hypothesize two different models. In the first model, Vegt pre-marks the regulatory region prior to Nodal signaling induction and Vegt binding then allows for Smad2/3 binding in these regions. If Vegt and Smad2/3 were combinatorially required for enhancer activity, then the corresponding gene would not be activated until Smad2/3 is present hence generating a delayed induction of the second wave. In the second model, Vegt and Smad2/3 coordinate binding in regulatory regions. In this model, while Vegt protein is present in endodermal cells, it does not associate with

specific regulatory regions in the absence of Smad2/3. When Nodal signaling is activated and Smad2/3 is present in the nucleus, Vegt and Smad2/3 can coordinate binding to enhancers and consequently activate the corresponding second wave genes. In this case, delayed binding of Vegt causes the delayed induction of second wave genes. While incomplete, the findings in this chapter provide a useful framework for studying a mechanistic model of ZGA onset regulation from a systems biological perspective.

Results

Type I FFLs are significantly enriched in the Xenopus GRN

In Chapter 4, I quantified the number of network motifs in the *Xenopus* mesendoderm GRN. However, it as unclear whether these quantities for the auto-regulatory, feedback and feedforward loops are statistically significant. In order to address this question, I generated 10,000 random GRNs that are of similar size to the *Xenopus* GRN where #(genes) = 37, #(direct positive connections) = 82, and #(direct negative connections) = 14. From these randomized networks, I counted the number of network motif types and compared the empirical measures from the *Xenopus* GRN. In general, most motifs are enriched in the *Xenopus* GRN, although motifs such as Type VII and Type VIII FFLs quantities fall below randomized GRN background (Table 5.1). Among motifs that only includes direct positive connections, only the Type I FFL is enriched 6.5-fold (p-value = 3.3e-19) in the *Xenopus* GRN compared to randomized GRNs, while both the positive auto-regulatory feedback and

the positive feedback loops quantity falls below randomized GRNs. This analysis suggests that only a subset of network motifs are enriched in the *Xenopus* GRN, and among them is the Type I FFL.

Two theoretical models to show waves of gene transcription during ZGA

Two competing theoretical models can describe the delayed induction of the secondary zygotic genes (Fig 5.1a). In the predominant linear model, Vegt activates Nodal signaling, and Nodal signaling activates the second wave of endodermal zygotic genes. Alternatively, in the FFL model, Vegt activates Nodal signaling and then both Vegt and Smad2/3 activate the expression of secondary zygotic genes. To verify that both models can generate multiple waves of gene expression, I computationally modeled the GRN interaction between Vegt, Nodal signaling (through Smad2/3) and second wave genes using partial differential equations. In both models, I represented the Nodal signaling pathway as simply the induction of Nodal ligands *gdf3*, *nodal5* and *nodal6.* The induction of these ligands is the critical step to induce the pathway in the endoderm[13]. As expected, both models can show staggered initial induction of zygotic genes, consistent with previous findings[32] (Fig 5.1b,c). The Nodal ligand expression is activated first, and this induction is followed by the secondary zygotic genes. However in the feed-forward loop model, this staggered induction is only feasible when both Vegt and Nodal signaling are combinatorially required to activate the secondary zygotic gene. Such a combinatorial requirement (so called *AND-gate*) for causing delayed induction has also been described[30]. If either Vegt alone or Nodal

signaling alone is capable of inducing the target genes, then the induction of secondary zygotic genes occurs alongside the induction of Nodal genes (not shown). This scenario results in a single wave of gene activation, and thereby fails to explain the multiplicity of onset of ZGA.

Molecular mechanisms of FFL formation

While these models can computationally explain the multiplicity in the onset of ZGA, I wondered mechanistically how these network motifs could form in the cis-regulatory regions of the genome. The linear model appears the most straightforward, as the binding of Vegt in CRMs of second wave genes is not required, and that the binding of Smad2/3 is sufficient to induce secondary gene expression (Fig 5.1d). In these cases, the limiting factor in the induction of the secondary zygotic genes is the activity of Smad2/3. As the activity of Smad2/3 requires the activation of zygotic nodal ligands *nodal5* and *nodal6*, this results in delayed induction of secondary zygotic genes.

In the feedforward loop models, there are two possible mechanisms. First, Vegt can pre-mark these regulatory regions pre-MBT, and then allow for Smad2/3 binding (Fig 5.1e). This is consistent with the findings in Chapter 2, as Vegt, along with other maternal TFs can pre-mark regulatory regions as early as the 32-cell stage (Fig 2.5). As T-box and Smad2/3 proteins have previously been shown to physically interact, possibly, Vegt recruits Smad2/3 into these regions[32,33]. Alternatively, the second mechanism is that Vegt cannot

bind to these regions in the absence of Smad2/3 (Fig 5.1f). However, in the presence of Smad2/3, Vegt can associate to these CRMs either through coordinated binding, or even through recruitment by Smad2/3. As Vegt is not present in the CRMs in the absence of Smad2/3, this results in the delayed onset of induction of secondary zygotic genes.

In Chapter 2, I generated Vegt ChIP-seq at Stage 8 (pre-MBT). Post-MBT, Smad2/3 ChIP-seq at Stage 10 has been published[29] and, while unpublished, Vegt ChIP-seq at Stage 10 has already been generated by a collaborator. I used these datasets to validate the molecular mechanisms of FFL formation.

Vegt and Smad2/3 chromatin binding support the mechanistic formation of FFLs

After peak calling these ChIP-seq datasets, I identified regions of chromatin overlaps and found candidate regulatory regions consistent with the three models described above (Fig 5.2a). First, the linear model does not require the binding of Vegt and hence any binding of Smad2/3 could pose as candidate regulatory regions for the linear model. In the case of FFLs which involved pre-marking by Vegt, I found 114 regulatory regions where Vegt persistently binds and is co-bound by Smad2/3. In the case of coordinated binding, I found 185 regulatory regions where Vegt did not previously bind prior to the appearance of active Nodal signaling. However, these regions are co-bound by Vegt and Smad2/3 at later stages.

Since I am specifically interested in regions that are regulated through a FFL mechanisms, I prioritized regions that are adjacent to genes that are regulated by both Vegt and Smad2/3. I used previously published datasets, combined with our own ChIP-seq datasets to identify a set of genes that are directly regulated by Vegt and Smad2/3, among which are mesendodermal genes *foxa4*, *gata4*, and *gata6*[21,29,34]. A subset of these genes contains a peak that corresponds to one or both types of FFL mechanisms (Fig 5.2b). For example, the *gata4* promoter region and the *mix1* downstream region contain a persistent Vegt peak and a Smad2/3 peak (Fig 5.2b,c). These peaks are marked by the transcriptional co-activator Ep300 suggesting that these regulatory regions are active. In addition, as I am aware that TFs can associate with the chromatin through various mechanisms, these CRMs contain at least one motif for both Vegt and Smad2/3 suggesting that these can be directly bound by both Vegt and Smad2/3. These ChIP-seq findings identified possible chromatin regions regulated through the linear model and both mechanistic models of FFLs.

Discussion

The abundance of FFLs in single-celled organism GRNs and the sea urchin developmental GRNs has previously been established[30,31]. I found this network motif to be highly abundant in the *Xenopus* network and I wondered if there is a mechanistic reason for the abundance of this network motif[17]. I looked into the function of this network motif during early *Xenopus* development and hypothesized that FFLs can generate the multiplicity in onset of zygotic gene activation. My computational modeling corroborates previous

findings that FFLs can generate the multiple waves of gene induction. ChIP-seq binding of Vegt and Smad2/3 support this hypothesis as Vegt and Smad2/3 co-bind in cis-regulatory regions. Vegt and Smad2/3 can form these FFLs either by Vegt pre-marking these CRMs followed by Smad2/3 binding or alternatively, Vegt and Smad2/3 can coordinate binding to these regions together. Further experiments required to establish these models will be discussed in Chapter 6.

FFLs as second tier control of ZGA

While multiple models have been proposed to explain the timing of the onset of ZGA, recent RNA expression analysis showed that there are multiple waves of gene induction. While the linear model of gene induction is sufficient to explain the formation of multiple waves, the network motif analysis suggests that an alternative model using FFLs is at play. Maternal transcription factors can retain control of zygotic gene activation even after the first waves of factors are activated. In this chapter, I focus on the role of Vegt and Smad2/3 in the endoderm although this mechanism could be more ubiquitous. For example, in the endoderm, Mixer[13] is an early expressed TF activated pre-MBT and could also possibly collaborate with Vegt in the activation of second wave zygotic genes. In addition, this mechanism could be involved in other regions of the embryo. For example, the maternal β-catenin is active in the dorsal regions of the embryo, and has been found to be active through the gastrula stages, like Vegt. Possibly, β-catenin[35,36] collaborates with dorsally-activated zygotic transcription factors such as Otx2[37,38,40] or Gsc[39,40] to form the Spemann

organizer in the dorsal mesoderm. Since the *Xenopus* ectodermal network is not as established as the mesendodermal network, possibly, knowledge of network motifs in the mesendoderm can provide useful hypothesis in the ectodermal network. For example, the maternal factors Grhl3 and Foxi2[41] (Table 2.1) might collaborate with zygotic genes such as Foxi1[42] to activate the secondary zygotic genes in the ectoderm.

FFLs in developmental regulatory networks

As I uncovered that the Xenopus mesendoderm GRN has many FFLs, like the sea urchin network, I wondered if this was the case as well in other published developmental systems. I looked into published GRNs where network connections have been identified as direct. I used the published *C. elegans* endoderm GRN[43], the neural crest GRN from zebrafish and chick[44], and the mammalian pancreatic GRN[45], and compared them to the *Xenopus* mesendoderm and sea urchin ectoderm GRNs. Using computational pattern matching for 2-connection and 3-connection network motifs, all five GRNs showed an abundance of the type I FFLs, while other network motifs vary in abundance (Table 5.1). For example, the *C. elegans* endoderm GRN is formed through feedforward loops during both the specification and the differentiation processes. This finding suggests the importance of FFLs in developmental systems and the possibility that this is a fundamental feature of control of gene expression timing in many contexts.

Chapter 5 Figures and Tables

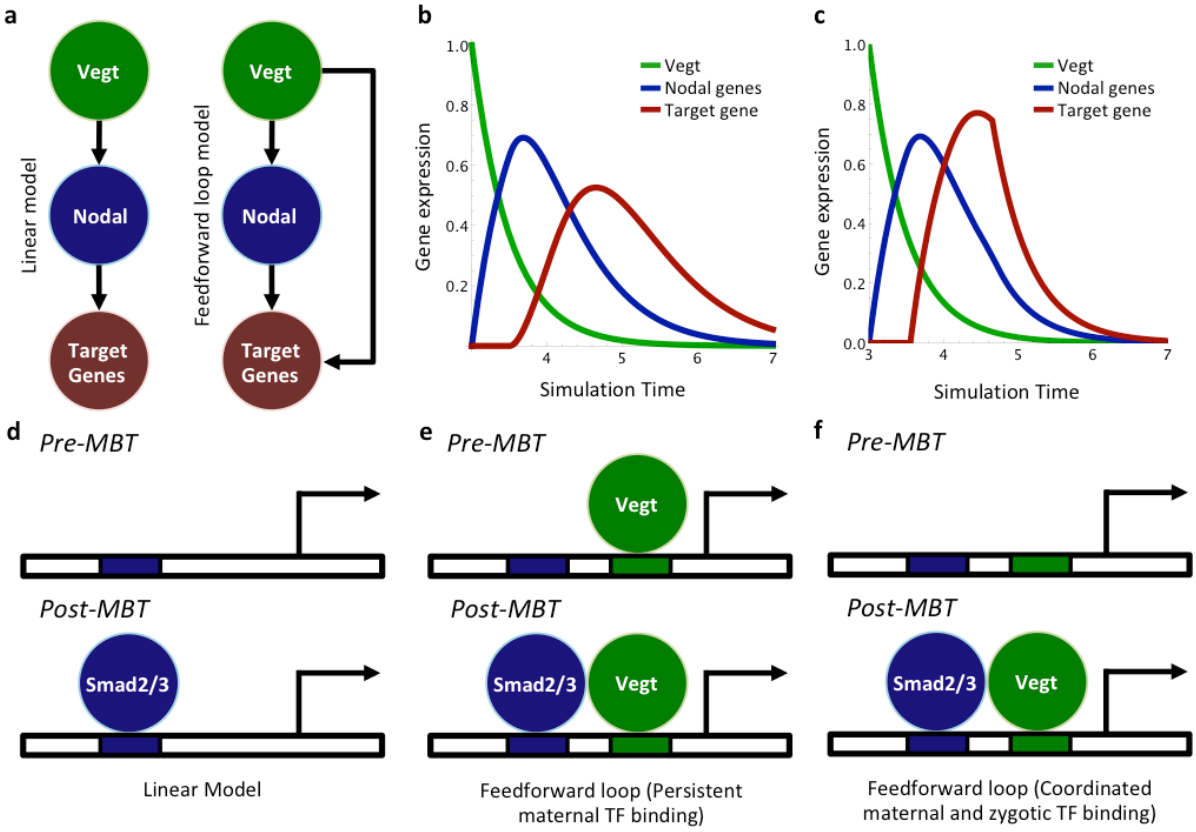

Figure 5.1 *Model of transcription factor binding and feedforward loops.* (a) The linear and the feedforward loop models are competing models to explain multiple waves of transcription during ZGA. Computational modeling of gene networks following the linear model (b) or the feedforward loop model (c) showing temporal gene expression. (d) Chromatin binding of Smad2/3 in the linear model. Chromatin binding of Smad2/3 and Vegt in feedforward loop mechanisms showing pre-marking by Vegt (e) and coordinated binding by Vegt and Smad2/3 (f).

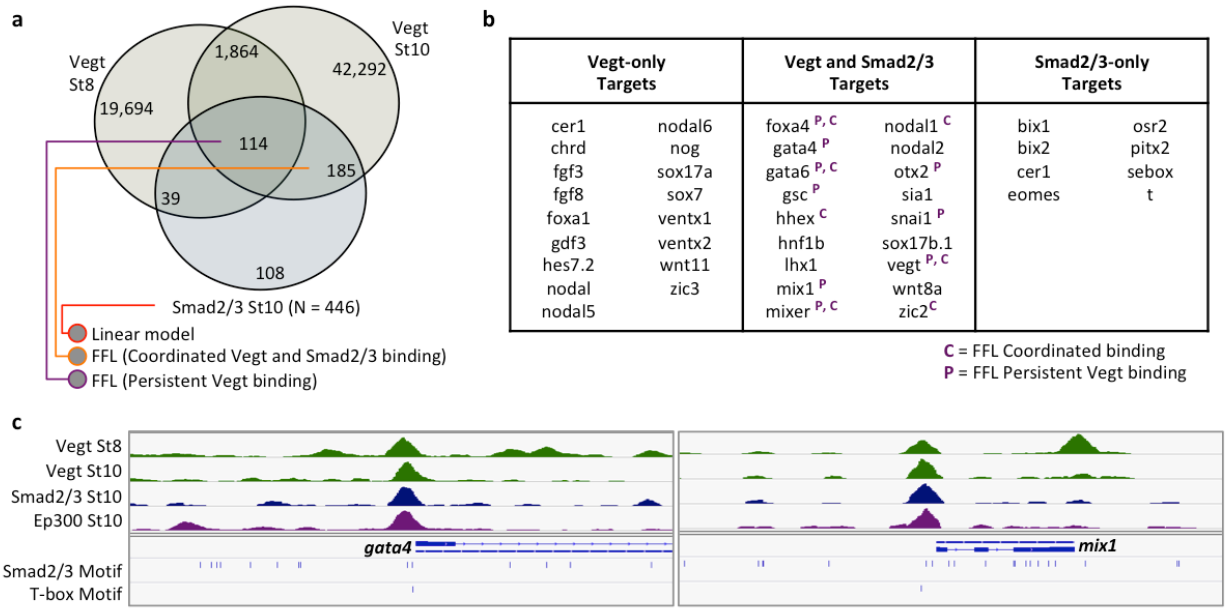

Figure 5.2 *Candidate regulatory regions of linear and feedforward loop models* (a) Overlap of Vegt St 8, Vegt St 10 and Smad2/3 St 10 ChIP-seq peaks highlighting potential regulatory regions for the linear model (red); and the feedforward loop model using coordinate binding (orange) or persistent maternal TF binding (violet). (b) Set of direct mesendodermal targets regulated by Vegt-only, Smad2/3-only or both. For the co-regulated genes, labels indicate whether their binding regions are consistent with the FFL coordinated binding or FFL persistent Vegt binding models seen in panel A. (c) Genome browser showing Vegt (St 8 and St 10), Smad2/3 (St 10) and Ep300 (St 10) along with the location of Smad2/3 and T-box consensus binding motifs. Specifically shown are the genomic regions of *gata4* and *mix1* which both contain a candidate for FFL persistent Vegt binding.

	Developmental Gene Regulatory Networks					*Xenopus* mesendoderm randomized (N = 10,000)
Motif Type	***Xenopus*** **Mesendoderm**	***C. Elegans*** **endoderm**	**Neural Crest**	**Pancreatic**	**Sea Urchin Ectoderm**	
Auto-regulation (positive)	2	2	4	1	0	2.44
Auto-regulation (negative)	2	0	0	0	5	0.38
Positive Feedback	3	1	0	1	3	4.80
Double Negative Feedback	1	0	0	0	0	0.86
Negative Feedback	1	0	0	0	0	0.48
Feedforward loops						
Type I	63	15	15	5	24	9.62
Type II	3	0	9	0	0	0.28
Type III	6	0	0	0	0	0.28
Type IV	3	0	0	0	0	0.28
Type V	13	0	0	0	8	1.71
Type VI	1	0	0	0	2	0.04
Type VII	0	0	0	0	0	1.70
Type VIII	0	0	0	0	0	1.68

Table 5.1 *Abundance of common regulatory motifs across diverse gene regulatory networks.* Network motif composition of five developmental GRNs. On the right-most column is the average motif composition of the randomized *Xenopus* mesendoderm GRN simulated 10,000 times.

References

1.Veenstra, G. J., Destrée, O. H. & Wolffe, A. P. Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition.. *Mol Cell Biol* **19,** 7972–82 (1999).

2.Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. & Klein, P. S. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2.. *Dev Cell* **19,** 220–31 (2010).

3.Lee, M. T. *et al.*. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition.. *Nature* **503,** 360–4 (2013).

4.Leichsenring, M., Maes, J., Mössner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates.. *Science* **341,** 1005–9 (2013).

5.Liang, H. L. *et al.*. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila.. *Nature* **456,** 400–3 (2008).

6.Hendrickson, P. G. *et al.*. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons.. *Nat Genet* **49,** 925–934 (2017).

7.De, I. A. *et al.*. DUX-family transcription factors regulate zygotic genome activation in placental mammals.. *Nat Genet* **49,** 941–945 (2017).

8.Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network.. *Nat Genet* **49,** 935–940 (2017).

9.Charney, R. M. *et al.*. Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program.. *Dev Cell* **40,** 595–607.e4 (2017).

10.Lee, M. T., Bonneau, A. R. & Giraldez, A. J. Zygotic genome activation during the maternal-to-zygotic transition.. *Annu Rev Cell Dev Biol* **30,** 581–613 (2014).

11.Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates.. *Dev Cell* **42,** 316–332 (2017).

12.Yang, J., Tan, C., Darken, R. S., Wilson, P. A. & Klein, P. S. Beta-catenin/Tcf-regulated transcription prior to the midblastula transition.. *Development* **129,** 5743–52 (2002).

13.Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L. & Klein, P. S. An essential role for transcription before the MBT in Xenopus laevis.. *Dev Biol* **357,** 478–91 (2011).

14.Loose, M. & Patient, R. A genetic regulatory network for Xenopus mesendoderm formation.. *Dev Biol* **271,** 467–78 (2004).

15.Koide, T., Hayata, T. & Cho, K. W. Xenopus as a model system to study transcriptional regulatory networks.. *Proc Natl Acad Sci U S A* **102,** 4943–8 (2005).

16.Heasman, J. Patterning the early Xenopus embryo.. *Development* **133,** 1205–17 (2006).

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

18.Stennard, F., Carnac, G. & Gurdon, J. B. The Xenopus T-box gene, Antipodean, encodes a vegetally localised maternal mRNA and can trigger mesoderm formation.. *Development* **122,** 4179–88 (1996).

19.Zhang, J. & King, M. L. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning.. *Development* **122,** 4119–29 (1996).

20.Lustig, K. D., Kroll, K. L., Sun, E. E. & Kirschner, M. W. Expression cloning of a Xenopus T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation.. *Development* **122,** 4001–12 (1996).

21.Horb, M. E. & Thomsen, G. H. A vegetally localized T-box transcription factor in Xenopus eggs specifies mesoderm and endoderm and is essential for embryonic mesoderm formation.. *Development* **124,** 1689–98 (1997).

22.Kofron, M. *et al.*. Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors.. *Development* **126,** 5759–70 (1999).

23.White, R. J., Sun, B. I., Sive, H. L. & Smith, J. C. Direct and indirect regulation of derrière, a Xenopus mesoderm-inducing factor, by VegT.. *Development* **129,** 4867–76 (2002).

24.Messenger, N. J. *et al.*. Functional specificity of the Xenopus T-domain protein Brachyury is conferred by its ability to interact with Smad1.. *Dev Cell* **8,** 599–610 (2005).

25.Houston, D. W. & Wylie, C. Maternal Xenopus Zic2 negatively regulates Nodal-related gene expression during anteroposterior patterning.. *Development* **132,** 4845–55 (2005).

26.Hilton, E., Rex, M. & Old, R. VegT activation of the early zygotic gene Xnr5 requires lifting of Tcf-mediated repression in the Xenopus blastula.. *Mech Dev* **120,** 1127–38 (2003).

27.Houston, D. W. *et al.*. Repression of organizer genes in dorsal and ventral Xenopus cells mediated by maternal XTcf3.. *Development* **129,** 4015–25 (2002).

28.Agius, E., Oelgeschläger, M., Wessely, O., Kemp, C. & De, R. E. M. Endodermal Nodal-related signals and mesoderm induction in Xenopus.. *Development* **127,** 1173–83 (2000).

29.Chiu, W. T. *et al.*. Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program.. *Development* **141,** 4537–47 (2014).

30.Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences* **100,** 11980–11985 (2003).

31.Peter, I. S. & Davidson, E. H. Assessing regulatory information in developmental gene regulatory networks.. *Proc Natl Acad Sci U S A* **114,** 5862–5869 (2017).

32.Slagle, C. E., Aoki, T. & Burdine, R. D. Nodal-dependent mesendoderm specification requires the combinatorial activities of FoxH1 and Eomesodermin.. *PLoS Genet* **7,** e1002072 (2011).

33.Nelson, A. C. *et al.*. Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression.. *BMC Biol* **12,** 81 (2014).

34.Xanthos, J. B., Kofron, M., Wylie, C. & Heasman, J. Maternal VegT is the initiator of a molecular network specifying endoderm in Xenopus laevis.. *Development* **128,** 167–80 (2001).

35.Tao, Q. *et al.*. Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos.. *Cell* **120,** 857–71 (2005).

36.Nakamura, Y., de, P. A. E., Veenstra, G. J. & Hoppler, S. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules.. *Development* **143,** 1914–25 (2016).

37.Blitz, I. L. & Cho, K. W. Anterior neurectoderm is progressively induced during gastrulation: the role of the Xenopus homeobox gene orthodenticle.. *Development* **121,** 993–1004 (1995).

38.Pannese, M. *et al.*. The Xenopus homologue of Otx2 is a maternal homeobox gene that demarcates and specifies anterior body regions.. *Development* **121,** 707–20 (1995).

39.Cho, K. W., Blumberg, B., Steinbeisser, H. & De, R. E. M. Molecular nature of Spemann's organizer: the role of the Xenopus homeobox gene goosecoid.. *Cell* **67,** 1111–20 (1991).

40.Yasuoka, Y. *et al.*. Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification.. *Nat Commun* **5,** 4322 (2014).

41.Cha, S. W., McAdams, M., Kormish, J., Wylie, C. & Kofron, M. Foxi2 is an animally localized maternal mRNA in Xenopus, and an activator of the zygotic ectoderm activator Foxi1e.. *PLoS One* **7,** e41782 (2012).

42.Mir, A. *et al.*. FoxI1e activates ectoderm formation and controls cell position in the Xenopus blastula.. *Development* **134,** 779–88 (2007).

43.Maduro, M. F. Gut development in C. elegans.. *Semin Cell Dev Biol* **66,** 3–11 (2017).

44.Simões-Costa, M. & Bronner, M. E. Establishing neural crest identity: a gene regulatory recipe.. *Development* **142,** 242–57 (2015).

45.Servitja, J. M. & Ferrer, J. Transcriptional networks controlling pancreatic development and beta cell function.. *Diabetologia* **47,** 597–613 (2004).

Chapter 6

Outlook

Maternal transcription factors and endoderm formation

The aim of my thesis was to uncover the combinatorial function of maternal transcription factors (TFs) during zygotic gene activation to specify the endodermal germ layer. To accomplish this goal, I used different frameworks and approaches. In Chapter 2, I've used developmental genetic approaches to understand the function of *otx1* within the context of early endoderm formation. Using standard *Xenopus* gain- and loss-of-function methods, I uncovered the role of Otx1 in combination with another maternal endodermal factor Vegt in activating endoderm regulatory network. In Chapters 2 and 3, I used genomic approaches to understand the endogenous *in vivo* chromatin binding of these maternal TFs. I found that Vegt and Otx1 co-bind with the ubiquitously expressed Foxh1 in cis-regulatory regions of the genome. Together, these three maternal factors pre-mark active endodermal enhancers, prior to establishment of enhancer-specific epigenetic marks. In Chapter 4, I used a systems biological approach to understand the role of maternal TFs within the context of the mesendodermal differentiation process. I curated existing datasets to build an updated *Xenopus* mesendodermal GRN. In Chapters 4 and 5, I zoomed into the GRN at the level of network motifs. I combined computational modeling and genomic approaches to understand the role of specific motifs, particularly, ones that involved the function of maternal TFs. As the primary conclusions for each project are already covered at the end of each chapter, I will focus on a few of the many questions that I have encountered in my project that I was unable to address.

What is the role of the larger set of maternal transcription factors?

In my screen for candidates of maternal specifiers of germ layer formation, I obtained 9 vegetally- and 4 animally-localized TFs (Fig 2.2). Among the vegetal RNAs are *vegt* and *otx1*, whose function I addressed in Chapters 2 and 3. Using developmental genetic approaches, I uncovered their combinatorial roles in the activation of the endoderm regulatory network (Fig 2.4). Using ChIP-seq and ChIP-qPCR analysis, I show that Vegt and Otx1 interact in the genome, along with the ubiquitously expressed Foxh1 (Fig 2.8-2.9).

While my analysis was limited to Vegt, Otx1 and Foxh1, both my screen for maternal specifiers of germ layer formation and my ChIP-seq analysis suggest a role for combinatorial regulation by a larger set of maternal TFs. Such is the case for the TFs *sox7*, *sox3*, *zic2* and *pou5f3*/*oct60*. All four TFs are maternally expressed and can possibly interact with Vegt and Otx1, and when I analyzed the motifs under Vegt and Otx1 binding from ChIP-seq, the motifs for all four factors are enriched. These findings suggest that other maternal TFs collaborate with *vegt* and *otx1* during endoderm differentiation. As the function of these genes during endoderm differentiation are generally unknown, similar developmental genetics and genomic approaches can be applied to understand their function.

Analysis of the role of individual genes is relatively straightforward. The hurdle will be understanding how all these factors combinatorially regulate. With *vegt* and *otx1*, there is

only one combination to experimentally test. However, increasing the number of factors increases the number of combinations such that experimental testing as performed in this thesis is not feasible. Answering these important developmental questions will rely more on systems biological and genomic approaches.

What distinguishes the activating and inhibitory roles of transcription factors?

In Chapter 2, I showed that *vegt* and *otx1* can act as an activator of endodermal genes and as a repressor of mesodermal genes, and by ChIP-seq, I showed that both activation and repression functions could be direct (Table 2.4-2.6). What is unclear is how *vegt* and *otx1* can perform either function in a gene specific manner within the same cellular context. As I have performed experiments to uncover the molecular nature of *otx1* duality, I will focus the discussion on *otx1*, although parts of the discussion could apply to *vegt* as well.

In the case of anterior-posterior patterning of the neuroectoderm, the mouse OTX1 can act as a repressor by through a region containing the WSP/EH1 domain, which mediates interactions with the TLE (transducin-like enhancer of split) family of co-repressors[1]. Interaction with TLE co-repressors is crucial in order to inhibit the anterior spread of FGF8 made at the midbrain-hindbrain boundary into the fore- and midbrain. *Xenopus* Otx1 may generally act as an activator, and this function may be switched through interaction with any of the three maternally expressed TLE *Xenopus* homologs Tle1, Tle2, or Tle4. Indeed, when I inhibited the expression of Tle1, Tle2, Tle4, or all three using antisense morpholino

oligonucleotides, mesodermal genes were upregulated in the vegetal mass, similar to the Otx1 morphant (not shown). This finding suggested that Otx1 and Tle1/2/4 inhibit the expression of mesodermal genes in the endoderm. Possibly, through the mechanism established in mouse, Otx1 and Tle1/2/4 interact to perform this function. This hypothesis is supported by the chromatin binding of Otx1 and Tle1/2/4. When I performed correlation analysis of Tle1/2/4 ChIP-seq to maternal TF ChIP-seq, I found Otx1 to be highly correlated to Tle1/2/4 binding (not shown). This is in contrast to Vegt binding which did not correlate as strongly to Tle1/2/4 binding. Both the morpholino experiment and chromatin binding correlation supported the role of Tle1/2/4 in the switch between activating and repressing roles of Otx1.

This interpretation is made complicated by two findings. First, I hypothesized that Otx1 binding can be categorized as Tle-bound and Tle-unbound, and these two categories represent the inhibitory (mesodermal genes) and the activating (endodermal genes) roles of Otx1, respectively. When I performed this analysis, I found that approximately half of Otx1 peaks are co-bound by Tle1/2/4 and the other half are not (not shown). However, when I associated these two categories to gene expression by assigning peaks to the closest genes, I found no difference between their germ layer expression of these two sets of genes. The second pertains to the experiments that involved mutations in the *otx1* WSP/EH1 domain (not shown). The hypothesis for these experiments was that removal of the WSP/EH1 domain would result in an Otx1 protein that does not interact with Tle and hence, does not act as a repressive factor. I have generated two different mutations: complete deletion of WSP (*otx1*-ΔWSP) and alanine substitution of the eight amino acids

(*otx1*-ASUB). As shown in Chapter 2, when I co-inject wild type *otx1* mRNA with *vegt* mRNA, otx1 inhibits the induction of mesodermal genes by *vegt* (Fig 2.2b,c). When I performed a similar experiment by co-injecting vegt mRNA with *otx1*-ΔWSP or *otx1*-ASUB, surprisingly, these mutant *otx1* mRNAs performed similarly as wild type *otx1*. Using a different experimental design, I microinjected wild type *otx1*, *otx1*-ΔWSP or *otx1*-ASUB into the vegetal masses. Wild type *otx1* further down-regulated the expression of mesodermal genes. However, both *otx1*-ΔWSP or *otx1*-ASUB performed similarly as wild type *otx1*. These experiments suggest that loss of the WSP domain is insufficient to explain the inhibitory function of Otx1.

If the WSP domain is not necessary for the inhibitory function of *otx1*, this suggests an alternative mechanism. Here, I list three hypotheses. First, one possibility is that otx1 does not act as inhibitor at all during *Xenopus* endoderm formation. Possibly, *otx1* activates a repressor that targets mesodermal genes. To establish this model, identification of specific transcription factor targets of *otx1* is necessary. As we have performed LOF and GOF experiments on *otx1*, in addition to ChIP-seq, these datasets can be leveraged to identify a set of candidates.

Second, another possibility is that Otx1 can interact with Tle1/2/4 via a separate domain. In the mouse OTX1 studies, the WSP/EH1 domain was pinpointed to be the region which TLE interacts with because deletion of a large portion of the mouse OTX1 which contained the WSP resulted in loss of inhibitory properties[1]. In addition, since in the deleted region,

the WSP region is the only domain conserved between OTX1 and OTX2, the study concluded that the WSP region must be the region of interaction. There are two experiments missing that are needed to make this conclusion. First, the study showed that the OTX2 mutant cannot interact with TLE, but this was not shown with regards to OTX1. Second, the OTX1 deletion contained a total of 75 amino acid deletions (out of 352). Absence of these two experiments open up the possibility that the OTX1 WSP/EH1 domain is not the sole interaction domain for TLE. Surprisingly, experiments in zebrafish using equally large deletions in *otx1* suggested that the WSP domain is actually required for the activating roles of *otx1*^{2}, rather than inhibitory function in contrast to the mouse data. While the study in zebrafish did not focus on the interaction between *tle* and *otx1*, the study did highlight that further analysis of molecular domains of *otx1* is needed to elucidate the function of *otx1*.

Third, Otx1 might not interact with Tle at all, and Otx1 can interact with a different co-repressor. As not much is known about Otx1 co-factors, an exploratory approach in search of Otx1-interacting proteins could provide novel mechanisms of Otx1 gene activation or repression.

Overall, the current model whereby Tle acts as a switch between the activator and repressor roles of Otx1 role is insufficient. Further analysis of the protein domains and protein-protein interactions of Otx1 will be required to understand this switch.

Why are Vegt and Otx1 bound regions depleted of H3K4me1 marks after zygotic genome activation?

In Chapters 2 and 3, I showed that maternal TFs combinatorially bound regions of the DNA that are likely active cis-regulatory regions. These regions are decorated by active enhancer marks H3K27ac and Ep300 around the stages of zygotic gene activation. In addition, I've used these cis-regulatory regions for reporter assays and they appear to be endodermally-active (not shown). However, one of the most surprising results I have encountered during this analysis is the characteristic anti-correlation between the enhancer mark H3K4me1 and Vegt + Otx1 only co-bound regions (without Foxh1). In contrast, regions that are bound by Foxh1 + Vegt + Otx1, Foxh1 + Vegt only and Foxh1 + Otx1 only are decorated by this enhancer mark from the blastula through the neurula stages. Why then are H3K4me1 enhancer marks specifically depleted in Vegt + Otx1 only co-bound regions?

This question represents a large gap in knowledge in establishment of H3K4me1 marks in *Xenopus* in general. Specifically, we are missing (1) enzymes involved in establishment of H3K4me1 marks and (2) how enzymes associate with specific regions of the chromatin. First, in various eukaryotic systems, H3K4 mono-methylation and demethylation is performed by enzymes found in various eukaryotes including *Xenopus*. MLL/Set are highly conserved proteins, which play a major role in establishing monomethyl marks[3]. While the mechanism for demethylation is not as well established, this role seems to be performed by the protein LSD1 at least in human embryonic stem cells[4]. Possibly, the establishment and

specific inhibition of H3K4me1 in specific regulatory regions is performed by orthologous *Xenopus* genes. Indeed, orthologous histone H3 lysine 4 methyltransferases (MLL/SET *Xenopus* orthologs *kmt2a*, *kmt2b*, *kmt2c* and *kmt2d*) and demethylases (LSD1 *Xenopus* orthologs *kdm1a*, *kdm1b*, *kdm5a*, *kdm5b* and *kdm5c*) are expressed maternally[5].

Second, if these same enzymes are involved in establishment of chromatin marks in *Xenopus* as they are in yeast or mammals, then the question becomes how do these enzymes associate with specific regions of the chromatin. As Chapter 4 indicated, TF binding motifs appear to control the specificity of establishment of DNA regulatory regions. Likely, these same *Xenopus* maternal TFs directly recruit or are involved in the recruitment of these enzymes. While not much is known regarding the interaction of *Xenopus* TFs and histone modifying enzymes, there are a few cases elucidated. During early dorsal-ventral patterning of the mesendoderm, β-catenin recruits Prmt2, which catalyzes dimethylation of H3R8 in cis-regulatory regions of β -catenin target genes[6]. In later stages, during neurogenesis, the TF Neurog2 mediates H3K9me2 demethylation through interaction with the MLL/Set ortholog Kdm3a[7]. Perhaps, in the case of epigenetic marks established during or after zygotic gene activation, the maternal factors Foxh1, Otx1 and Vegt are involved in recruiting their respective modifiers.

Addressing these questions require proper functional studies of these histone modifiers in the early embryogenesis of *Xenopus*. First, these studies will require gain- and loss-of-

function approaches in conjunction with chromatin binding assays to establish the role of these proteins in *Xenopus*. Second, these studies will require assays of protein-protein interactions using co-immunoprecipitation and mass spectrometry to identify transcription factor and histone modifier interactions.

Are FFLs second tier regulators of the onset of zygotic genome activation?

In Chapters 4 and 5, I describe the abundance and possible function of FFLs in a variety of developmental systems. In the case of early *Xenopus* germ layer formation, FFL contains maternal TFs, which activates a gene that encodes for a zygotic TF or signaling molecule (Fig 4.2d). Subsequently, the combinations of the maternal and zygotic factors regulate a set of second wave zygotic factors. Possibly, this network motif could have implications in the regulation of the multiple waves of gene activation during this period of development, as feed-forward loops have been attributed to regulation of timing of gene expression[8]. In Chapter 5, I use computational modeling to validate this function of FFLs. Additionally, I use ChIP-seq datasets to describe the mechanistic model of FFL formation using the maternal factor Vegt and the Nodal signaling TF Smad2/3.

While the network connections which generated the FFLs in the early *Xenopus* GRN have been validated, the temporal regulatory aspects of feed-forward loops have not been investigated. In the case of the *Xenopus* endoderm, it is essential to show various aspects of Vegt and Smad2/3 co-regulation. By computational modeling, I showed that the delayed

onset of second wave genes is dependent on a mechanism whereby Vegt and Smad2/3 are combinatorially required, and the absence of either one cannot activate second wave gene expression (Fig 5.1a). This specific feature of FFLs could be used to validate the model, and specifically, this feature could distinguish the FFL model of gene regulation from the linear model. Experimentally, as loss-of-function of Vegt also inhibits Smad2/3 activity, then the types of experiments must focus on the regulatory regions of second wave genes. One such experiment is through reporter gene analysis where deletions of the T-box or Smad binding sites can show that the expression of the reporter is lost in the absence of either. Positive results in these types of experiments would indicate that both Vegt and Smad2/3 are required to regulate the enhancer, thus validating the FFL. Alternatively, instead of using reporter constructs, endogenous regulatory DNA function could be assayed using the new generation of CRISPR/Cas9[9] which can create controlled mutations at the single base level. The use of this new technology has not been previously applied in *Xenopus*, however successful application this approach can transform functional analysis of regulatory regions, or at the very least provide supporting evidence along with reporter gene experiments.

Final remarks

In conclusion, my thesis revealed the critical role of combinatorial regulation of maternal TFs at both the cis-regulatory level and the GRN levels. Hopefully, the ideas discussed and the evidence generated will act as useful stepping stones in future scientific ventures in the

fields of cell fate specification, gene regulatory network science, evo-devo, chromatin biology and that which unifies the body of my thesis, developmental biology.

References

1.Puelles, E. *et al.*. Otx2 regulates the extent, identity and fate of neuronal progenitor domains in the ventral midbrain.. *Development* **131,** 2037–48 (2004).

2.Bellipanni, G., Murakami, T. & Weinberg, E. S. Molecular dissection of Otx1 functional domains in the zebrafish embryo.. *J Cell Physiol* **222,** 286–93 (2010).

3.Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Molecular Cell* **25,** 15–30 (2007).

4.Whyte, W. A. *et al.*. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482,** 221–225 (2012).

5.Owens, N. D. L. *et al.*. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development.. *Cell Rep* **14,** 632–647 (2016).

6.Blythe, S. A., Cha, S. W., Tadjuidje, E., Heasman, J. & Klein, P. S. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2.. *Dev Cell* **19,** 220–31 (2010).

7.Lin, H. *et al.*. KDM3A-mediated demethylation of histone H3 lysine 9 facilitates the chromatin binding of Neurog2 during neurogenesis.. *Development* **144,** 3674–3685 (2017).

8.Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif.. *Proc Natl Acad Sci U S A* **100,** 11980–5 (2003).

9.Gaudelli, N. M. *et al.*. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage.. *Nature* **551,** 464–471 (2017).