UNIVERSITY OF CALIFORNIA,
IRVINE

Activating the Zygotic Genome: The Interplay between Transcription Factors, Epigenetic Modifiers, and *cis*-Regulatory Modules

DISSERTATION

Submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Jeff Jiajing Zhou

Dissertation Committee:
Professor Ken W. Y. Cho, Chair
Professor Kyoko Yokomori
Associate Professor Wenqi Wang
Professor Yongsheng Shi
Assistant Professor Evgeny Kvon

2022

DEDICATION

To

my mother

in recognition of her love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my mentor and committee chair, Dr. Ken W.Y. Cho, for his guidance, patience, encouragement, and support during my graduate study. His mentorship has molded me into a scientist with diligence, independence, and critical thinking.

I also would like to express my gratitude and appreciation to my committee member, Dr. Wenqi Wang, and his postdoctoral fellow, Dr. Han Han, for their generous support, critical comments, and collaborative work.

I am thankful to my committee members, Dr. Kyoko Yokomori, Dr. Yongsheng Shi, Dr. Evgeny Kvon and former committee member Dr. Zeba Wunderlich for their guidance, constructive criticism, and support.

I am grateful to Cho lab members for their generous help, thoughtful discussion, and immense support. I thank Dr. Ira Blitz, the general of our frog comrades, for the guidance and expertise in *Xenopus*. I thank Dr. Jin Cho for her patience, experiment assistance, and immense help on bioinformatics. I thank Paula Pham for her thoughtful discussions and emotional support. I thank Dr. Kitt Paraiso for his help and collaboration. I also thank Jessica Cheung, Tae Yeon Yoo, Clark Hendrickson, and Amina Hussein for their help and support.

I want to express my most sincere appreciation and love to my family and friends for their support and patience during my graduate school. They have been my most enormous emotional support and cheerleaders. I would not have accomplished my graduate degree without them.

VITA

Jeff Jiajing Zhou

EDUCATION

Ph.D. in Biological Sciences, University of California, Irvine, Department of Developmental and Cell Biology, June 2022

B.S. in Biochemistry and Cell Biology, University of California, San Diego, Warren College, June 2016

PUBLICATIONS

Zhou JJ, Pham PD, Han H, Blitz IL, Wang W, Cho KWY. Foxh1 engages in complex chromatin regulation revealed by protein interactome analyses. *Development, Growth & Differentiation*. (Manuscript in prep)

Zhou JJ, Cho KWY. Epigenomic dynamics of early *Xenopus* embryos. *Development, Growth & Differentiation*. (Manuscript in prep)

Zhou JJ, Cho JS, Han H, Blitz IL, Wang W, Cho KWY. Histone deacetylase 1 maintains lineage integrity through histone acetylome refinement during early embryogenesis. *Elife*. (Manuscript submitted)

Jansen C, Paraiso KD, Zhou JJ, Blitz IL, Fish MB, Charney RM, Cho JS, Yasuoka Y, Sudou N, Bright AR, Wlizla M, Veenstra GJC, Taira M, Zorn AM, Mortazavi A, Cho KWY. Uncovering the mesendoderm gene regulatory network through multi-omic data integration. *Cell Rep*. 2022 Feb 15;38(7):110364.

Han H, Nakaoka HJ, Hofmann L, Zhou JJ, Yu C, Zeng L, Nan J, Seo G, Vargas RE, Yang B, Qi R, Bardwell L, Fishman DA, Cho KWY, Huang L, Luo R, Warrior R, Wang W. The Hippo pathway kinases LATS1 and LATS2 attenuate cellular responses to heavy metals through phosphorylating MTF1. *Nat Cell Biol*. 2022 Jan;24(1):74-87.

Paraiso KD, Blitz IL, Zhou JJ, Cho KWY. Morpholinos Do Not Elicit an Innate Immune Response during Early *Xenopus* Embryogenesis. *Dev Cell*. 2019 May 20;49(4):643-650.e3.

Han H, Qi R, Zhou JJ, Ta AP, Yang B, Nakaoka HJ, Seo G, Guan KL, Luo R, Wang W. Regulation of the Hippo Pathway by Phosphatidic Acid-Mediated Lipid-Protein Interaction. *Mol Cell*. 2018 Oct 18;72(2):328-340.e8.

PRESENTATIONS AND TRAINING

Zhou JJ. Identification of FoxH1 Transcription Coregulators during Early Embryogenesis. Department of Developmental and Cell Biology Seminar Research in Progress Talk, Irvine, California, January 30th, 2019

UCI-Genomics High Throughput Facility: Topics in Bioinformatics Workshops- RNA-seq Analysis. University of California, Irvine, March 22nd, 2018

Cell and Developmental Biology of *Xenopus*: Gene Discovery& Disease, Cold Spring Harbor Laboratory, Class of 2018

PROFESSIONAL AND TEACHING EXPERIENCE

Graduate Teaching Assistant for *Eukaryotic and Human Genetics D137*: University of California, Irvine, Fall 2019-2021

Graduate Teaching Assistant for *Developmental Biology D104*: University of California, Irvine, Summer 2018, Winter 2019-2020

Graduate Admin Teaching Assistant for *DNA to Organisms Bio93*: University of California, Irvine, Fall 2018

Graduate Teaching Assistant for *Cell Biology D103*: University of California, Irvine, Fall 2017

Undergraduate Instructional Assistant for *Introductory Biology Lab BILD4*: University of California, San Diego, Spring 2016

Undergraduate Instructional Assistant for *Molecular Biology BIMM100*: University of California, San Diego, Winter 2016

Undergraduate Research Assistant at Shannon Lauberth Lab: University of California, San Diego, June 2015- March 2016

Laboratory Technician Assistant in Aquatic Animal Facility: University of California, San Diego, June 2014- May 2016

ABSTRACT OF THE DISSERTATION

Activating the Zygotic Genome: The Interplay between Transcription Factors, Epigenetic Modifiers, and *cis*-Regulatory Modules

by

Jeff Jiajing Zhou

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2022

Professor Ken W. Y. Cho, Chair

During early animal development, maternal mRNAs and proteins direct the biosynthetic activities of embryos. Through the maternal-to-zygotic transition, when both maternal clearance and zygotic genome activation (ZGA) occur, the zygote takes over the control of its developmental programs. ZGA, accompanied by epigenetic remodeling, is coordinated by the interplay between maternal transcription factors, epigenetic modifiers, and *cis*-regulatory modules. How such interactions are orchestrated within a chromatin environment to ensure proper embryogenesis is a fundamental issue in biology. Understanding mechanisms that underlie genome activation will advance applications in cellular reprogramming in culture, diagnosis of diseases, and therapeutic development.

Here, I uncovered the protein-protein interaction network of Foxh1, which is a maternal transcription factor functioning in *Xenopus* mesendoderm development. Proteomic analyses revealed that Foxh1 interacts with both positive and negative regulators, suggesting a dual role for Foxh1 in regulating zygotic transcription. My findings infer a mechanism by which Foxh1

recruits repressive epigenetic complexes to suppress mesendodermal genes in developing ectoderm.

Next, I conducted genome-wide analyses of Hdac1, an enzyme that removes acetyl groups from acetylated histones, in early *Xenopus* embryos. My results revealed a dual function model of Hdac1. On the one hand, Hdac1 keeps inactive chromatin free of histone acetylation resulting in the spatial and temporal repression of developmental genes. On the other hand, Hdac1 participates in dynamic histone acetylation-deacetylation cycles on active chromatin, sustaining gene expression in different germ layers. Thus, Hdac1 controls embryonic cell identity through epigenetic regulations.

Lastly, I developed a high-throughput screening protocol to identify functional *cis*-regulatory modules in *Xenopus* embryos by modifying STARR-seq method. My results demonstrated that the activities of candidate *cis*-regulatory modules can be effectively and quantitatively assessed; however, the complexity of *cis*-regulatory modules needs to be further determined. This modified protocol provides a basis to examine functional enhancers during ZGA *in vivo*.

My work illustrates a complex regulation of ZGA by transcription factors, epigenetic modifiers, and *cis*-regulatory modules. The findings presented here will have broad impacts on developmental biology, stem cell biology, and human diseases.

CHAPTER 1

Introduction

Cellular differentiation, beginning at fertilization, is programmed by a hard-wired genetic control system. The coordinated interaction between transcriptional regulators, epigenetic modifiers, and *cis*-regulatory modules (CRMs) is manipulated within a chromatin environment to ensure proper differentiation trajectories. Dissecting the mechanism governing cellular differentiation has broad implications in the fields of human development, stem cell biology, and pathology. With the immense advancement in technology, genome-wide analyses have been instrumental in drawing associations between behaviors of transcriptional regulators, profiles of epigenetic signatures, and function of DNA sequences to uncover the mechanisms governing gene expression, cell lineage specification, and disease states (ENCODE Project Consortium, 2012; Berdasco and Esteller, 2010; Furey, 2012; Berger and Mardis, 2018). However, the close link between the molecular actions of regulators and subsequent delineation of cell lineages during early development remains elusive.

Animal development starts with a union of two specialized cells, an egg and a sperm. The fertilized zygote is loaded with maternally supplied factors but is mostly transcriptionally silent (Tadros and Lipshitz, 2009; Vastenhouw et al., 2019). A process known as the maternal-to-zygotic transition (MZT) reprograms the zygote to a transient totipotent state through both nuclear and cytoplasmic activities. MZT is tightly dictated by two coordinated processes: maternal clearance and zygotic genome activation (ZGA). Maternal mRNAs and proteins inherited in the egg cytoplasm direct the initial biosynthetic processes in the transcriptionally silent zygote. Subsequently, zygotic genes are transcribed from the nucleus leading to an active

clearance of maternal mRNAs. This mother-to-embryo hand-off in genetic control is also accompanied by remodeling the zygotic epigenome to ensure a robust and stabilized lineage commitment of embryonic cells (Jukam et al., 2017; Schulz and Harrison, 2018; Wu et al., 2020; Blitz and Cho, 2021).

Shortly after ZGA, there begins the first cell lineage segregation, the delineation of three primary germ layers ectoderm, mesoderm, and endoderm. Cells from ectodermal lineage give rise to skin epidermis and nervous system; cells from mesodermal lineage form notochord, somite, heart, kidney, muscles, and blood system; cells in endodermal lineage become respiratory tracts and gastrointestinal system (Keller, 2005). This early stage of animal development is ideal for studying the mechanisms of gene regulation. In contrast to later developmental transitions, early embryos must undergo a drastic reprogramming of gene expression where numerous maternal transcription factors (TFs) specify embryonic cell fates. This provides a system to unveil the hierarchical order of key transcription regulators involved in gene regulatory networks controlling germ layer specification. In addition, the epigenome inherited from the parental gametes must be remodeled such that zygotic developmental programs robustly maintain the integrity of embryonic cell fates. This allows us to probe zygotically defined epigenome dynamics that entail chromatin structure, nucleosome organization, and post-translational modifications to histones and DNA. Moreover, transcriptional regulators and epigenetic modifiers must act on specific CRMs to produce coherent outputs of gene activities defining embryonic cell identities. This enables us to assess the architecture of DNA elements that contribute to the functionality of CRMs. In this chapter, I discuss the current understanding of early *Xenopus* development, emphasizing the role of

maternal transcription factor Foxh1, epigenetic regulatory mechanisms governing ZGA, and CRMs controlling zygotic genes.

The role of maternal transcription factor Foxh1 during ZGA

The fertilized embryo undergoes rapid cell division to give rise to numerous embryonic cells without any notable change to its size (Yartseva and Giraldez, 2015). Synchronized cell divisions partition maternal mRNAs and proteins into different blastomeres. Transcripts encoding maternal TFs are distributed such that they can be localized or uniform spatially. This process is essential to specify germ layer identity along the animal-vegetal axis of the embryo. The animal cap cells are fated to become the future ectoderm dictated by animally localized TFs such as Foxi2 (Cha et al., 2012). The vegetal mass cells are specified to form the prospective endoderm directed by vegetally localized TFs such as Vegt and Otx1 (Zhang et al., 1998; Kofron et al., 1999; Paraiso et al., 2019). Molecular signaling pathways initiated in the endoderm are critical to induce the overlying ectoderm layer into the mesoderm formed at the marginal zone, which is followed by the dorsal-ventral patterning of the mesoderm. Besides the spatially localized TFs, spatially ubiquitous TFs such as Sox3, Pou5f3, and Foxh1 also play pivotal roles in germ layer specification (Penzel et al., 1997; Chiu et al., 2014; Charney et al., 2017; Gentsch et al., 2019).

Foxh1, formerly known as Fast1, is a winged-helix transcription factor initially identified as a transcriptional co-activator of TGF-β signaling pathway (Chen et al., 1996). TGF-β superfamily of cytokines plays a central role in mesoderm induction and axial patterning by activating a full range of dorsal and ventral mesoderm marker genes in early embryos (Wu and Hill, 2009). The immediate transcriptional response, such as the activation of Mix1 and Mix2, is independent of protein translation, where a 290-base-pair promoter element (activin-responsive

element, ARE) at Mix2 gene directs the immediate transcriptional effect of TGF-β signals (Chen et al., 1996). Yeast one-hybrid screening with a *Xenopus* oocyte cDNA library identified Foxh1 as the factor acting on ARE, unveiling a direct interaction between cytosolic signal transducer Smads and a site-specific DNA-binding protein Foxh1 in mediating TGF-β signaling pathway.

Foxh1 is a major transcriptional co-activator of Nodal signaling pathway- part of the TGF-β superfamily conserved in amniotes (Osada et al., 2000; Whitman, 2001; Schier, 2003; Shen, 2007). Mature Nodal ligands binding to type II receptor (ActRII or ActRIIB), together with the type I serine-threonine kinase receptor (ALK4), stimulates the activation of receptors where the type II receptor kinase phosphorylates both type I receptor and cytosolic transducers Smad2 or Smad3. Phosphorylated Smad2/3 associates with Smad4 and translocates to the nucleus. The Smad2/3-Smad4 heterodimer then interacts with co-activator Foxh1 to induce the transcription of target genes. This core pathway of Nodal signaling is also modulated at various levels. For instance, the initiation by the binding between Nodal ligands and receptors can be disrupted by extracellular inhibitors Bambi, Lefty and Cerberus (Onichtchouk et al., 1999; Chen and Shen, 2004; Tavares et al., 2007). Phosphorylation on Smads can be reversed by Ppm1A phosphatase, which then stimulates the nuclear export of Smads (Lin et al., 2006). Transcriptional co-repressor Drap1 can inhibit the DNA binding of Foxh1 (Iratni et al., 2002). These complex regulations to Nodal signaling allow the developmental cues to exploit Nodal activity at different developmental stages.

In *Xenopus*, Foxh1 is maternally expressed and its temporal expression peaks before gastrulation (Chen et al., 1996; Owens et al., 2016; Charney et al., 2017). Early ARE reporter studies showed that Foxh1 is required to assemble the prototypical Foxh1/Smad2/Smad4 complex and activate transcription fully at ARE (Zhou et al., 1998; Yeo et al., 1999). The

temporally restricted availability of Foxh1 may act as a mechanism by which the TGF- β mediated induction of mesodermal genes occurs prior to gastrulation but not later in development. Hence, Foxh1 is a temporal determinant in the early embryonic response to Activin/ Nodal induction. The nodal-dependent role of Foxh1 in mesoderm specification is elucidated in many studies. Ectopic expression of constitutive active Foxh1 induces mesodermal (*tbxt, gsc, chordin, and wnt8*) and endodermal genes (*edd, sox17b*) in the animal cap; in contrast, the ectopic expression of repressor fused Foxh1 or the injection of anti-Foxh1 antibody blocks the expression of endogenous mesodermal and endodermal genes (Watanabe and Whiteman, 1999). Knockdown of endogenous Foxh1 also confers the activator role of Foxh1 in the formation of mesoderm and endoderm. Depleting endogenous Foxh1 with morpholino antisense oligonucleotides impairs activin-induced convergent extension movements accompanied by defective blastopore closure (Howell et al., 2002). Foxh1-depleted embryos display severe anterior malformation and axial abnormality with reduced expression of endodermal and mesodermal genes (*cer, dkk, lhx1, tbxt, mix, nodal1, mixer*) (Kofron et al., 2004). The study also revealed a transcriptional repressor role of Foxh1 where the expression of *nodal5* and *nodal6* is upregulated in Foxh1-depleted embryos. Genome-wide studies reinforce the Nodal-dependent role of Foxh1 during gastrulation. Foxh1 shares an overlapping binding profile with Smad2/3 on mesodermal and endodermal gene CRMs and regulates the expression of genes sensitive to the Nodal signaling inhibitor (Chiu et al., 2014; Charney et al., 2017). Together, these studies demonstrate an essential role of Foxh1 acting as a Nodal signaling co-activator during early *Xenopus* germ layer specification.

Interestingly, Foxh1 can also function independently of Nodal signaling. Comparing genes regulated by Foxh1 and Nodal signaling, more than 60% of genes regulated by Foxh1 are

neither sensitive to Nodal inhibition nor bound by Smad2/3 (Chiu et al., 2014). These genes are expressed ubiquitously in the early gastrula, like the expression profile of Foxh1 before its rapid decline. A further investigation of the Nodal-independent function of Foxh1 illustrated that Foxh1 protein levels are relatively consistent and abundant before gastrulation (Charney et al., 2017). Strikingly, the binding of Foxh1 to genomic loci is detected as early as the 32~64 cell stage, which is hours before the activation of Nodal signaling. The persistent binding of Foxh1 to mesendodermal gene loci is not affected by the Nodal inhibitor suggests the genomic binding of Foxh1 does not rely on Nodal signaling. This study proposed a passive pioneer transcription factor role for Foxh1 where the binding of Foxh1 early in time is essential to induce ZGA and epigenetic remodeling related to mesendoderm formation. Recently, Foxh1 is also shown to function cooperatively with other TFs before the activation of Nodal signaling. For example, Foxh1, together with Vegt and Otx1, establishes endodermal super-enhancers (high H3K4me1) to activate endodermal genes during ZGA (Paraiso et al., 2019). Foxh1, together with Sox3 and Vegt, occupies genomic regions to facilitate the chromatin engagement of zygotic signals (Gentsch et al., 2019). Despite the findings that Foxh1 functions in both Nodal dependent and independent manner, there remain questions as to what protein complexes are recruited to Foxh1 bound loci to modulate transcriptional outputs and chromatin states.

Dynamics of epigenetic remodeling during ZGA

Proper embryonic development requires an extensive reorganization of the zygotic epigenome, which underlies a transition from transcriptionally silent joined gametes to a transcriptionally active zygote (Skvortsova et al., 2018; Eckersley-Maslin et al., 2018). This epigenome remodeling begins shortly after fertilization when maternal TFs establish the initial totipotent

state of the zygotic epigenome in cooperation with epigenetic modifiers. The zygotic epigenome is continuously modified to accompany ZGA to ensure a robust control by zygotic developmental programs (Schulz and Harrison, 2018), thereby maintaining the integrity of cell fates in response to diverse developmental cues. Significant efforts have been placed to uncover the relationship between epigenetic regulation and gene expression during the earliest cell fate decision process (Figure 1.1).

Higher-order chromatin structure

The spatial organization of chromatin imposes a critical influence on genome functions (Vallot and Tachibana, 2020). Around ZGA, higher-order chromatin structures undergo extensively reorganized to accompany the gradual activation of zygotic genes. Chromosome conformation capture-based methods revealed two segregated compartments, compartment A and compartment B, in the interphase chromatin (Lieberman-Aiden et al., 2009). In general, compartment A correlates with transcriptionally active genes, and compartment B associates with transcriptionally silent genes (Wang et al., 2016). Within a compartment, numerous chromatin contact domains, or topologically associating domains (TADs), are spatially segregated to form distinct folding configurations (Pombo and Dillon, 2015). Extensive chromatin loops (chromatin-chromatin interactions) mediated by the cohesion complex are present within TADs where *cis*-elements of DNA sequences are in close physical proximity to each other. Boundary elements such as CCCTC-binding factor (CTCF) are essential to maintain the transient stability of TADs and prevent the chromatin contacts outside of TADs (Ong and Corces, 2014).

In *Xenopus*, the structural organization of chromatin is not detected at the mid blastula (st8) stage (Niu et al., 2021). Weak, TAD-like structures emerge at the late blastula (st9) stage,

suggesting that TADs formation coincides with ZGA. The chromatin loops between TAD borders and chromatin compartments are continuously consolidated from the gastrula and onward. Inhibition of transcription by knocking down RNA polymerase II subunit RPB1 or α-Amanitin injection does not affect the formation of TADs at gastrula. Depletion of CTCF and Rad21, known factors involved in TAD regulation and loop structure formation (Heidari et al., 2014), results in compromised TAD formation with weakened TAD borders and attenuated chromatin loop interactions within TADs, respectively. The genomic binding of CTCF to establish TADs relies on the activity of chromatin remodeling complexes such as ISWI complex.

The mechanism controlling TADs formation is context-dependent and regulated differently in different species. In fruit flies, frogs, mice, and humans, TADs initially form at ZGA and are consolidated onward (Sexton et al., 2012; Ogiyama et al., 2018; Niu et al., 2021; Ke et al., 2017; Du et al., 2017). On the contrary, zebrafish chromatin is highly structured before ZGA, undergoes a drastic loss in structure at ZGA, and re-establishes the organization after ZGA (Kaaij et al., 2018). The mechanism for such a discrepancy is unclear. The requirement of active transcription for TADs formation is also controversial. In fruit flies, frogs, and mice, transcription is dispensable to TADs formation (Hug et al., 2017; Niu et al., 2021; Ke et al., 2017; Du et al., 2017); however, ZGA is required for TADs establishment in human embryos (Chen et al., 2019). Transcription-dependent TADs formation is also shown in prokaryotes *Caulobacter* (Le et al., 2016) and post-mitotic human monocyte-derived macrophages infected by influenza A virus (Heinz et al., 2018). Adult tissue cells carry distinct configurations of TADs. For example, mammalian primary liver cells exhibit robustly conserved syntenic regions and binding profiles of CTCF (Vietri Rudan et al., 2015). Conversely, adult frog liver cells have low levels of CTCF and Rad21 correlated with weak TAD structures (Niu et al., 2021). These

different observations on TADs formation across tissues and organisms demonstrate a highly complex regulation of the higher-order chromatin structure dictating cellular identities and physiological functions.

Nucleosomes and histone variants

Nucleosome positioning is a major chromatin regulatory mechanism by which it determines the chromatin accessibility for TFs and general transcription machinery (Tsompana and Buck, 2014). Incorporation of histone variants into nucleosome also play a role in the accessibility of underlying DNA to transcriptional regulators (Venkatesh and Workman, 2015). Thus, nucleosomes and histone variants are critical players in regulating DNA-dependent processes such as transcription, DNA replication, and repair. Transcription of zygotic genes during ZGA is tightly linked to the dynamic establishment of the nucleosome landscape and incorporation of histone variants.

The position of nucleosomes significantly dictates chromatin accessibility (Radman-Livaja and Rando, 2010). Open or accessible DNA regions (e.g., promoters and enhancers) to transcriptional regulators are often nuclease hypersensitive due to nucleosome destabilization or eviction that exposes the underneath regulatory DNA sequences (Gross and Garrard, 1988; Henikoff, 2008). DNAse-seq can detect regulatory elements such as promoters and enhancers in early *Xenopus* gastrula (Cho et al., 2019). Pluripotent factors PouV and Sox3 exhibit pioneering activity to open chromatin around TF-bound regions during ZGA (Gentsch et al., 2019). Interestingly, regions previously occupied by pluripotency factors lose chromatin accessibility at the end of gastrula (Esmaeili et al., 2020). Promoters of dorsal Wnt targets (*sia1* and *nodal3.1*) are not accessible in early gastrula, partially due to the loss of histone acetylation in these

chromatin regions. Post zygotic chromatin gradually gains accessibility after late blastula (st9), largely correlating with the appearance of co-activator Ep300 and promoter histone mark H3K4me3 (Bright et al., 2021). The gradual increase in accessibility of zygotic chromatin observed in frogs is generally consistent with findings in zebrafish and mice (Liu et al., 2018; Lu et al., 2016).

H3.3, the replacement variant of H3, has been implicated in the epigetic regulation during early development (Szenker et al., 2011). It is enriched explicitly at highly active transcriptional regions marked by histone modifications associated with active transcription (Ahmad and Henikoff, 2002; Hake et al., 2006). An early nuclear transplant experiment demonstrated that incorporating H3.3 in *myod* promoter is required for its persistent expression in the enucleated egg through the epigenetic memory (Ng and Gurdon, 2008). H3.3 is required for proper gastrulation by activating mesoderm and neuroectoderm inductive genes and maintaining cell viability (Szenker et al., 2012). Phosphorylation at serine 31 on H3.3 promotes acetylation and represses methylation on lysine 27 in *cis* (Sitbon et al., 2020). These observations suggest a role for H3.3 in forming a permissive chromatin state for transcription. Unlike H3.3, the linker histone H1 is not assembled into the core histone particle; instead, it creates the chromatosome core particle by binding the linker region at both the entry and the exit sites of the nucleosome core particle (Fyodorov et al., 2017). Similar to fruit flies and zebrafish, early frog embryos lack H1 but are loaded with maternal subtypes of H1 (Khochbin, 2001). A linker histone variant B4 (also known as H1M) shares 30% identify with H1 (Smith et al., 1988), and predominantly functions as H1 during early *Xenopus* development (Dimitrov et al., 1993). The level of zygotic H1 reaches the level of maternal B4 at mid blastula transition and exceeds vastly at the neurula stage when the level of maternal B4 drastically diminishes. Since B4 is less basic than H1, B4

has a reduced affinity to nucleosomes than H1, establishing a more permissive transcription state of the early embryonic genome (Ura et al., 1996). Together, the presence of different types of histones can modulate the transcriptional output during ZGA.

Post-translational modifications on histones

Histone tails are ~25-30% of the mass of individual histones, and protrude from the surface of the chromatin polymer (Wolffe and Hayes, 1999). Inherently disordered histone tails are extensively modified, including acetylation, methylation, phosphorylation, and ubiquitination. The patterns of covalent histone modifications, referred to as the 'histone code', significantly impact chromatin structure and are interpreted by 'reader' proteins to carry out diverse cellular functions (Strahl and Allis, 2000). In non-mammalian systems, the early embryonic genome is rather naïve, and major chromatin modifications occur at or after ZGA (Bonn et al., 2012; Vastenhouw et al., 2010; Gupta et al., 2014; van Heeringen et al., 2014). During ZGA, the zygotic epigenome undergoes dynamic changes of histone marks, which in turn affects both temporal and spatial gene expression patterns.

Histone acetylation occurs on the ε-amino group of the lysine residues within *N*-terminal tails of all four core histones H2A, H2B, H3 and H4 (Inoue and Fujimoto, 1969; Seto and Yoshida, 2014). Histone acetylation is often indicative of permissive chromatin states because the acyl groups neutralize the positive charge on the lysine residues, thereby reducing the affinity of histones to DNA (Wang et al., 2000; Anderson et al., 2001). The acetylation status of histone is determined by the equilibrium between the activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Seto and Yoshida, 2014). In *Xenopus*, histone acetylation homeostasis is required to induce mesoderm cell fate and maintain stable states of differential

gene activities during gastrulation (Almouzni et al., 1994). H3K9ac and H3K14ac are detected on β-catenin target genes (*sia1* and *nodal3.1*) prior to ZGA at mid-blastula transition (Blythe et al., 2010). Genome-wide H3K27ac marks appear around the onset of ZGA and are widespread in the genome at gastrula stages (Gupta et al., 2014). The major HAT, Ep300 progressively binds to zygotic genomic loci from the blastula and onward (Hontelez et al., 2015). The binding of Ep300 is sensitive to the loss of zygotic transcription, suggesting the involvement of zygotic factors in Ep300 binding refinement. Conversely, the low histone acetylation level in pluripotent blastula is essential for its developmental plasticity and competence (Rao and LaBonne, 2018). These findings illustrate a close link between the establishment of histone acetylation and the process of ZGA.

The histone acetylation status on chromatin also affects the accessibility of underlying DNA sequences. *Xenopus* ectoderm explants lose Wnt-responsive competence due to the loss of histone acetylation mediated chromatin closure (Esmaeili et al., 2020), implying a positive correlation between histone acetylation and chromatin accessibility. An increase in histone acetylation appears to be a shared feature of ZGA during animal development. Histone acetylation in fruit flies emerges on a few zygotic genes during the mirror wave of ZGA, but increases dramatically at the major wave of ZGA (Li et al., 2014). In zebrafish, tissue-specific gene loci are marked by histone acetylation, which is essential to activate zygotic genes (Bogdanovic et al., 2012; Chan et al., 2019). In mice, H3K27ac increases genome-wide from oocytes to 2-cell-stage embryos correlating with the time of major ZGA (Dahl et al., 2016). Although histone acetylation generally increases from mid-blastula transition and positively influences gene expression during ZGA, there remain unanswered questions. How are enzymes

regulating histone acetylation employed at zygotic genes? How are the patterns of histone acetylation interpreted to influence the spatial and temporal expression of genes?

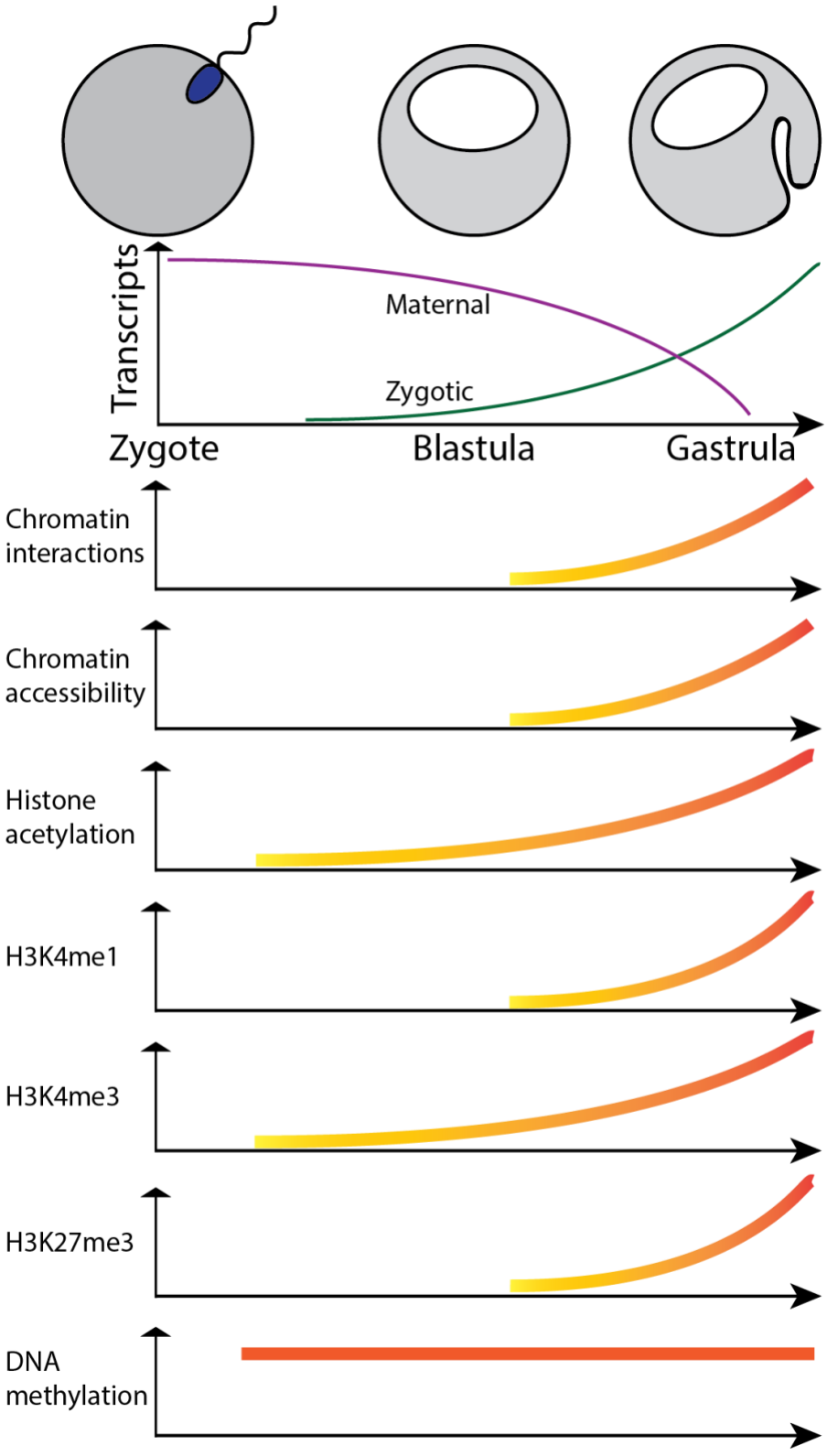

Figure 1.1: Characteristics of the epigenome during early *Xenopus* development. The patterns of different epigenetic signatures from the zygote to the gastrula stage.

Unlike histone acetylation, histone methylation can indicate either active or repressed states of transcription. H3K4me3 is often enriched at active promoters (Heintzman et al., 2007); H3K4me3, together with H3K27ac or H3K27me3, is associated with active chromatin states or repressive chromatin states, respectively (Skvortsova et al., 2018). H3K4me3 is catalyzed by lysine methyltransferases KMT2G/F (also known as SETD1B/1A) and lysine demethylases KDM5A-D (also known as JARID1A-D) (Beacon et al., 2021). In *Xenopus*, β-catenin dependent H3K4me3 is observed at Wnt target gene (*sia1* and *nodal3.1*) promoters before the onset of their transcription (Blythe et al., 2010). Genome-wide deposition of H3K4me3 precedes or coincides with ZGA (Akkers et al., 2009). Maternal factors primarily direct the deposition of H3K4me3 in regions lacking DNA methylation (Hontelez et al., 2015). In other animals, genome-wide H3K4me3 appears shortly after ZGA in fruit flies (Chen et al., 2013; Li et al., 2014) and before the onset of ZGA in zebrafish (Vastenhouw et al., 2010; Zhang et al., 2014). Though broad domains of H3K4me3 are detected in mice oocytes, H3K4me3 drastically reduces in level and is restricted to active gene promoters at the 2-cell stage, keeping them at a poised state for gene activation (Dahl et al., 2016; Zhang et al., 2016). Human oocytes show sharp peaks of H3K4me3 at gene promoters, and weak *de novo* deposition of H3K4me3 to the genome is observed in pre-ZGA embryos (Xia et al., 2019). These findings suggest that H3K4me3 deposition is a functionally conserved feature of major ZGA in animal development.

H3K4me1 is enriched at active and primed enhancers to fine-tune enhancer activities and functions (Dorighi et al., 2017; Local et al., 2017; Rickels et al., 2017). The presence of

H3K4me1, coupled with the absence of H3K4me3, distinguishes enhancers (distal CRMs) from proximal promoters (Heintzman et al., 2009). In addition, the co-existence of H3K27ac and H3K4me1 demarcates active enhancers from primed enhancers (Creyghton et al., 2010; Rada-Iglesias et al., 2011). The primary enzymes regulating H3K4me1 catalysis are KMT2C/D (also known as MLL3/4) and KDM1A/B (Herz et al., 2012; Beacon et al., 2021). Genome-wide detection of H3K4me1 is around the onset of ZGA in *Xenopus* (Hontelez et al., 2015). H3K4me1 marks both maternally instructed and zygotically directed CRMs; however, CRMs directed by zygotic factors exhibit weak H3K4me3 but strong H3K4me1 signals, indicating that these CRMs are distally located. Recent studies suggested that dense clusters of CRMs with high H3K4me1 signals act as super-enhancers (Loven et al., 2013; Whyte et al., 2013), which are thought to underlie robust tissue-specific gene expression. In *Xenopus* gastrula endoderm, key endodermal TFs Vegt, Otx1, and Foxh1 modulate the establishment of super-enhancers (Paraiso et al., 2019). The profile of H3K4me1 becomes germ-layer specific and forms super-enhancers by a sequential function of maternal and zygotic transcription factors by early gastrula (Paraiso et al., preprint). In fruit flies, H3K4me1 signals widely spread zygotic genome after ZGA flanking the binding of Zelda in fruit flies (Li et al., 2014; Moshe and Kaplan, 2017). CRMs marked by H3K4me1 are poised for further activation or repression during tissue patterning (Koenecke et al., 2017). In zebrafish, H3K4me1 is detected in a few regions at blastula but expands greatly afterward (Bogdanovic et al., 2012). Collectively, H3K4me1 primes CRMs around ZGA for robust zygotic transcription. However, the factors involved in establishing and interpreting H3K4me1 remain elusive.

The formation of heterochromatin, a repressive chromatin state, involves several key histone methylation modifications associated with gene suppression. Histone H3K9 methylation

mediated by the interplay between KMT1 family proteins (also known as Su(var)39 family) and KDM3/4/7 (also known as JHDM2/JHDM3/PHF8) is essential for the formation of constitutive heterochromatin through the recruitment of Heterochromatin Protein 1 (HP1) (Hyun et al., 2017; Allshire and Madhani, 2017). Constitutive heterochromatin consisting of tandem repeats is formed throughout the cell cycle and plays a vital role in genome stability (Richards and Elgin, 2002). Abundant H3K9me3 is present in *Xenopus* sperm but not in oocytes or eggs (Shechter et al., 2009). H3K9me2/3, together with H4K20me3, mark *Xenopus* genomic loci at blastula and onward (Hontelez et al., 2015). Retrotransposons are marked by H3K9me2/3 with H4K20me3, while a subset of DNA transposons lose these histone methylation modifications during gastrulation (van Kruijsbergen et al., 2017). The dynamic nature of H3K9 methylation is currently understudied, and the role of H3K9 methylation in gene repression during *Xenopus* ZGA requires further investigations.

The repressive histone H3K27 methylation, H3K27me3, is dynamically regulated during ZGA. Unlike H3K9 methylation, H3K27me3 marks are associated with gene repression for cell-type-specific genes, where the profiles of H3K27me3 can vary from one cell type to another (Margueron and Reinberg, 2011). Clusters of genomic regions enriched with H3K27me3 can function as silencers by forming intra-chromatin loops (Kundu et al., 2017; Cai et al., 2021). H3K27me3 is catalyzed by Polycomb repressive complex 2 (PRC2) and removed by KDM6A/B (Pan et al., 2018). In *Xenopus*, H3K27me3 deposition occurs after the onset of ZGA at many transcription factor genes correlating with the spatial restriction of gene expression (Akkers et al., 2009). The level of H3K27me3 increases at genomic loci devoid of DNA methylation from blastula to gastrula (van Heeringen et al., 2014). These genomic loci with H3K27me3 show repressor activity in reporter assays supporting a repressive role of H3K27me3 deposition. The

early landscape of H3K27me3 marks is instructed primarily by maternal factors, which persist at many zygotic genes even at later developmental stages (Hontelez et al., 2015). *Xenopus* endodermal super-enhancers are enriched with H3K27me3 and subunits of PRC2, Ezh2, and Jarid, suggesting the critical role of H3K27me3 during cell fate specification (Paraiso et al., 2019). Nevertheless, the factors direct the spatially restricted deposition of H3K27me3 are not known.

The regulation of H3K27me3 modifications seems to diverge among different animal species. In nematode worms and fruit flies, H3K27me3 is maternally transmitted from oocytes to embryos, modulating the balance between H3K27ac and H3K27me3 at enhancers (Gaydos et al., 2014; Zenk et al., 2017). A pervasive H3K27me3 is observed in fruit flies after ZGA (Li et al., 2014). In zebrafish, H3K27me3 co-occurs with H3K4me3 at genomic regions to poise gene activation (Vastenhouw et al., 2010; Lindeman et al., 2011), which is similar to bivalent domains identified in embryonic stem cells (Bernstein et al., 2006). However, bivalent domains are not observed in embryos of fruit flies or frogs at ZGA (Li et al., 2014; Akkers et al., 2009). Mouse embryos also exhibit persistent maternally inherited H3K27me3 at distal CRMs until blastocysts (Zheng et al., 2016). H3K27me3 marks are erased from paternal alleles upon fertilization and re-deposited after post-implantation. In addition, oocyte inherited H3K27me3 also plays a role in maternal imprinting independent of DNA methylation (Inoue et al., 2017). Contrary to mice, human embryos undergo global erasure of H3K27me3 at both parental alleles resulting in the lack of H3K27me3 at ZGA (Xia et al., 2019). The expression of PRC2 core subunits, EED and SUZ12, does not occur until after ZGA. Since H3K27me3 is erased globally in human embryos, it is unclear whether oocyte inherited H3K27me3 mediates gene imprinting. Taken together, H3K27me3 is dynamically regulated between different species.

DNA methylation

Methylation of the fifth carbon of cytosine is one of the most studied epigenetic modifications conserved among fungi, plants, and animals (Feng et al., 2010). The transmission of DNA methylation patterns is a critical epigenetic mechanism to ensure the integrity of cell states after cell divisions (Bird, 2002). The pattern of DNA methylation is often unique among cell types, and its changes influence gene expression (Schübeler, 2015). The deposition and maintenance of methylation on symmetrical CpG DNA are primarily controlled by DNA methyltransferases (Dnmts) (Li et al., 1992; Okano et al., 1999; Ramsahoye et al., 2000). Dnmts often directly interact with histone-modifying enzymes causing a repressive state in the regulatory regions of genes (Moore et al., 2013). CpG methylation also imposes both positive and negative influences on the binding of TFs to DNA (Yin et al., 2017). Therefore, remodeling DNA methylation patterns is critical in cellular differentiation and ZGA.

An early biochemical study demonstrated that *in vitro* methylated DNA is transcriptionally inactive when microinjected into *Xenopus* oocytes (Vardimon et al., 1982). This suggests a functional link between DNA methylation and gene repression. In contrast to mouse embryos (Zaitseva et al., 2007; Popp et al., 2010), *Xenopus* genome does not undergo global demethylation, but sustains a global hypermethylated state throughout early embryogenesis (Veenstra and Wolffe, 2001). Embryos depleted of Dnmt1 express zygotic genes approximately two cell cycles earlier than wildtype embryos (Stancheva and Meehan, 2000). However, the ectopic expression of catalytic inactive human DNMT1 in Dnmt1 morphants abolishes the premature ZGA phenotype (Dunican et al., 2008). These findings revealed that DNA

methylation itself may not play a prominent repressive role in zygotic genes. Instead, other complexes interacting with Dnmt1 are critical for repressing zygotic genes before ZGA.

Genome-wide DNA methylome survey in *Xenopus* showed that promoters and gene bodies of actively transcribed genes are hypermethylated, but H3K4me3 enriched transcriptional start sites are hypomethylated (Bogdanovic et al., 2011). Consistent with human cell culture (Lister et al., 2009; Ball et al., 2009), *Xenopus* embryos display enriched DNA methylation in gene bodies of highly active genes, which may act as a mechanism to suppress transcription from intragenic promoters (Maunakea et al., 2010). Polycomb-mediated H3K27me3 repression domains often are more susceptible to DNA methylation in human embryonic stem cells (Lynch et al., 2012), but they are not subjected to DNA methylation in early *Xenopus* embryos (Bogdanovic et al., 2011). These findings highlight the compatibility of DNA methylation and active transcription in early *Xenopus* embryos, distinguishing from the repressive role of DNA methylation during later developmental stages such as organogenesis. Further studies are needed to understand how DNA methylation influences transcriptional activities differently in early versus late developmental times.

Cis-regulatory modules controlling zygotic genes

Proper development requires precise patterns of gene expression both in time and space. Gene expression is a coherent output from both coding sequences and non-coding *cis*-regulatory modules (CRMs) orchestrated by the activity of TFs and epigenetic modifiers (Spitz and Furlong, 2012). CRMs are typically a few hundred DNA base pairs in length, which can enhance (Levine, 2010; Bulger and Groudine, 2011), repress (Ayer and Benyajati, 1990; Petrykowska et al., 2008), or insulate (Gaszner and Felsenfeld, 2006) gene expression. CRMs can also

selectively tether promoter-enhancer interaction to stimulate gene expression (Calhoun et al., 2002). Extensive studies have been conducted to understand how CRMs instruct the activities of gene regulators to carry out complex temporal and spatial patterns of gene expression (Patwardhan et al., 2009; Melnikov et al., 2012; Kvon et al., 2014; Moorthy et al., 2014; Arnold et al., 2014). Our understanding of the operational logic of CRMs has applications in the fields of cell differentiation and evolution. Moreover, the knowledge of CRM functions can be instrumental to therapeutic development because numerous genetic variants on CRMs are significantly associated with susceptibility to diseases (Hindorff et al., 2009; Mansour et al., 2014).

Studies on required DNA sequences for eukaryotic gene transcription by RNA polymerase II emerged during the early 1980s. Viral sequences were the first elements shown to highly activate transcription in *cis* within a plasmid construct (Banerji et al., 1981, Moreau et al., 1981). Metazoan genes were also shown to contain *cis*-acting sequences that strongly stimulate transcriptional activities in a tissue-specific manner (Banerji et al., 1983, Gillies et al., 1983). These early reporter studies defined a hallmark of enhancer elements that activate transcription, often independent of their location and orientation. In *Xenopus*, CRMs modulating gene *goosecoid* (*gsc*) were well-characterized (Koide et al., 2005). *Gsc* is expressed in Spemann's organizer to direct dorsal mesoderm formation (Cho et al., 1991). Two promoter elements, distal element (DE) and proximal element (PE), are required for efficient *gsc* expression *in vitro* and *in vivo* (Watabe et al., 1995). The expression of *gsc* is induced by Activin/BVg1(BMP-Vg1 chimera) at DE directly and by Wnt8 at PE. These two CRMs of *gsc* are functionally and structurally conserved in mice (Blum et al., 1992). Sia2 (formerly known as Twin) also activates *gsc* expression through direct binding at PE (Laurent et al., 1997). Transcriptional co-activators

Foxh1 and Gtf2ird1 (formerly known as XWBSCR11) function cooperatively in Activin/Nodal-mediated *gsc* induction at DE (Ring et al., 2002). Lhx1 (formerly known as Xlim1) and Otx2 maintain the expression of *gsc* by binding to the upstream element, DE, and PE of *gsc* during gastrulation (Taira et al., 1992; Blitz and Cho, 1995; Sudou et al., 2012). Mix-related homeobox proteins, Mixer and Milk, interact with Smads to activate *gsc* expression at DE (Germain et al., 2000). In addition, Gsc itself binds to both DE and PE resulting in the downregulation of its expression (Danilov et al., 1998). These findings demonstrate a systematic cellular signaling integration at *gsc* CRMs during dorsal mesoderm formation.

The conventional reporter gene assay is a gold standard method to assess the functionality of DNA sequences. Yet, it is limited by its low-throughput, and poor proxy for the activities of CRMs scattered along the genome. High-throughput sequencing methods allow us to construct an unbiased genomic view of potential CRMs in a wide range of tissues and developmental stages. CRMs are often characterized as nucleosome depleted and DNase hypersensitive due to the binding of transcriptional regulators (Elgin, 1988; Gross and Garrard, 1988). For example, the "house-keeping" gene *eef1a1o* and the developmental gene *otx2*, but not a silent gene *hbe1* are sensitive to DNase treatment in *Xenopus* gastrulae (Cho et al., 2019). CRMs occupied by pluripotency maternal TFs PouV and Sox3 are DNase sensitive during the mid-blastula transition (Gentsch et al., 2019). Compared to surrounding somatic tissues, accessible CRMs in neural progenitor cells are associated with genes functioning in neurogenesis and neuron differentiation during *Xenopus* tail regeneration (Kakebeen et al., 2020). CRMs associated with genes regulating mesodermal and neural development and cell pluripotency lose accessibility from the early to the late gastrula, indicating the temporal usage of accessible CRMs along the developmental trajectory (Esmaeili et al., 2020). Genome-wide chromatin

accessibility increases from late blastula but not before, indicating that the number of CRMs increases globally to accompany the expression of diverse zygotic genes (Bright et al. 2021). These findings demonstrate the application of genome-wide chromatin accessibility assays as a predictive measurement of potential CRMs.

Key TFs initiate the transcriptional complex assembly at CRMs to induce the gene expression networks governing specific cell lineages. Thus, high-throughput profiling of key TFs is also employed to identify potential CRMs. CRMs of key Spemann's organizer genes *cer* and *gsc* are marked by Vegt and Sia at late blastula (Sudou et al., 2012). CRMs bound by T-box TFs, Eomes, Vegt, and Tbxt are associated with genes for mesoderm induction (Gentsch et al., 2013). Head organizer gene Otx2 binds to CRMs of head inductive genes in cooperation with Lhx1 and Gsc (Yasuoka et al., 2014). Mesendodermal CRMs are persistently bound by maternal Foxh1 at early development, and such CRMs are primed by Foxh1 as early as 32~64 cell stages (Charney et al., 2017). A combinatory binding of maternal TFs Vegt, Otx1, and Foxh1 is found at CRMs associated with endodermal genes (Paraiso et al., 2019). *Xenopus* gastrula inductive signals Wnt, Nodal, and BMP depend on the binding of PouV and Sox3 to their CRMs (Gentsch et al., 2019). These studies indicate that the binding of crucial lineage specifying TFs, or often the combinatorial binding of TFs, can reveal potential CRMs.

Epigenetic features are highly correlated with the states of CRMs, where different activities of CRMs may be demarcated by a range of combinations of epigenetic features (Hardison and Taylor, 2012). The widely used epigenetic signatures to predict functional CRMs include histone acetylation (Roh et al., 2005), binding of co-activator p300 (Visel et al., 2009), and H3K4me1(Heintzman et al., 2009). The presence of H3K27ac on CRMs and the activities of neighboring genes are highly correlated (Creyghton et al., 2010; Rada-Iglesias et al., 2011).

During early *Xenopus* development, zygotic genomic loci are progressively marked by H3K27ac and Ep300 from blastula and onward (Gupta et al., 2014, Hontelez et al., 2015). This observation suggests that CRMs utilized by zygotic genes are occupied by histone acetyltransferases and decorated with histone acetylation. H3K4me1 is indictive of CRMs in a primed state where such CRMs may undergo rapid upregulation or active repression of their activities (Lupien et al., 2008; Cui et al., 2009; Rada-Iglesias et al., 2011; Bonn et al., 2012). Evidence in *Xenopus* endoderm explants shows clusters of endodermal (*foxa2*, *foxa4*, *gata6*, *hhex,* etc.) CRMs exhibit strong signal of H3K4me1 (Paraiso et al., 2019). These clusters of CRMs with the high signal of H3K4me1 are endodermal-specific super-enhancers, which are major hubs of transcriptional regulation (Loven et al., 2013; Whyte et al., 2013). Observed *Xenopus* endodermal super-enhancers are subjected to both active and repressive epigenetic modifications, presumably in different cell lineages (Paraiso et al., 2019; Paraiso et al., preprint), following the notion that H3K4me1 marks primed CRMs.

The integrative analysis of different high-throughput methods is the best approach to predicting functional CRMs. Recently, *Xenopus* mesendoderm gene regulatory network has been elucidated using linked self-organizing maps (Jansen et al., 2022). Newly identified CRMs through this highly dimensional integrating study exhibit reporter activities *in vivo*. However, whether all predicted CRMs are functional remains elusive. Thousands of developmental CRMs were functionally assessed in *Drosophila* embryos to uncover their temporal and spatial activities (Kvon et al., 2014). A system for high-throughput genomic integration of reporter constructs in *Xenopus* is unavailable. Alternatively, an episomal approach, self-transcribing active regulatory region sequencing (STARR-seq), effectively assesses the functionality of DNA sequences encompassing the entire *Drosophila* genome (Arnold et al., 2013). Additional variations to

STARR-seq have been developed to examine the CRMs in vertebrates with larger genome sizes (Liu et al., 2017; Barakat et al., 2018; Schöne et al., 2018; Wang et al., 2018; Neumayr et al., 2019). Applying this method may provide insights into how regulators integrate activities of functional CRMs in early *Xenopus* embryos.

For my dissertation work, I aimed to elucidate the contribution of maternal TFs, epigenetic modifiers, and functional CRMs to ZGA and understand how each of them impacts the formation of three primary germ layers in *Xenopus*. I proposed a transcriptional repressive function of maternal Foxh1 during gastrulation (Chapter 2). Next, I examined the regulation of histone acetylome by epigenetic modifier Hdac1, which is pivotal in maintaining the integrity of germ-layer transcriptomes (Chapter 3). Lastly, I modified and applied the high-throughput CRM identification method STARR-seq to assess the functionality of CRMs in gastrula embryos (Chapter 4). My work reveals the fundamental molecular principles during early development, which will be broadly applicable to developmental biology, chromatin biology, genomic biology, and human diseases.

CHAPTER 2

Identification of Protein Interaction Networks of Foxh1 in Mouse Embryonic Stem Cells

Abstract

Precise patterns of gene expression both in time and space ensure a proper progression of developmental processes. Early embryonic cell fates are specified through the integration of gene induction and suppression. Foxh1, a maternal transcription factor, primes embryonic cells to adopt a mesendodermal fate; yet, the transcriptional coregulators in conjunction with Foxh1 at genomic loci are unknown. Here, we report the protein-protein interactome network of Foxh1 in mouse embryonic stem cells. We confirm the interaction between Foxh1 and epigenetic repressors, PRC2 subunits and Hdac1. Foxh1 modulates genes in early gastrula *Xenopus* ectoderm where a subset of genes is co-repressed by Foxh1-Hdac1 interaction. Our findings reveal a complex genomic engagement of transcription factors and epigenetic modifiers for cell lineage specification.

Introduction

Precise regulation of gene expression is crucial for embryonic cell lineage determination during vertebrate embryogenesis. Understanding basic molecular mechanisms governing cell fate specification has far-reaching implications in stem cell biology, cancer biology, etc. Extensive studies have identified many essential genes that impose positive and negative regulations on developmental programs, thereby specifying cell identity (ENCODE, 2012; ENCODE, 2020). The dynamic equilibrium between gene induction and suppression ensures a proper progression of embryonic development, such as gastrulation.

In *Xenopus*, the vegetal region of the embryo poses two important developmental properties. It gives rise to the future endoderm autonomously and induces the overlying tissues to become the prospective mesoderm (Winklbauer, 2020). The vegetally localized maternal T-box TF Vegt is crucial to confer these two developmental properties. Vegt activates pro-endodermal TFs, including Sox17, Gata5, and Mixer, and controls the zygotic transcription of nodal ligands for mesoderm development (Stennard et al., 1996; Lustig et al., 1996; Zhang et al., 1996; Xanthos et al., 2001; Whiteman et al., 2001; Paraiso et al., 2019). Nodal ligands bind to type I and type II serine-threonine kinase receptors and induce the phosphorylation of cytosolic Smad2/3, leading to their interaction with Smad4 for nuclear translocation (Hill, 2001). Foxh1, a maternal forkhead-box TF, is a critical cofactor in forming a nodal-induced Smad complex that activates mesodermal and endodermal genes (Zhou et al., 1998; Howell et al., 2002; Chiu et al., 2014). Foxh1 binds to genomic loci as early as the 32~64 cell stage, significantly earlier than the onset of ZGA (128-cell stage); Foxh1 persistently binds to many putative mesendodermal CRMs, which subsequently gain well-characterized active histone modifications such as H3K4me1 and H3K27ac at ZGA and onward (Charney et al., 2017). These observations raise whether Foxh1 recruits epigenetic regulators like histone-modifying enzymes during ZGA. Furthermore, Nodal/Foxh1 can function as a repressor that inhibits *nodal5* and *nodal6* in the ventral vegetal blastula stage embryos (Kofron et al., 2004). Interestingly, a co-repressor Tle4 (also known as Grg4) is recruited to genomic loci by Foxh1 through physical interaction (Reid et al., 2016; Charney et al., 2017) independent of nodal signaling. These findings lead us to hypothesize that Foxh1 functions as a transcriptional repressor in the absence of nodal signaling, given that maternal Foxh1 is ubiquitously distributed across the developing embryo; if so, what

co-regulators are involved in this repression, and how is this repression carried out mechanistically?

Here, we report that FOXH1 participates in protein interaction networks functioning in various aspects of chromatin regulation in mouse embryonic stem cells (mESCs). The protein-protein interactions between Foxh1 and PRC2 core subunits and Hdac1 are recapitulated using *Xenopus* proteins. Foxh1 and Hdac1 co-repress a small subset of genes in the ectoderm of developing *Xenopus* embryos. Our results elucidate a transcriptional repressor function through epigenetic regulations for Foxh1.

Results

Generation of E14 cells stably expressing mFOXH1-SFB

To identify the specific protein interactome of mFOXH1, we first generated an E14 mESC line stably expressing mFOXH1 with epitope tags. We here constructed a mFOXH1 recombinant protein where a S-protein peptide, two copies of FLAG peptides, and a streptavidin-binding peptide (SBP), respectively, were added to the *C*-terminus (Figure 2.1A) (Wang et al., 2014). The S-protein peptide and SBP enable a sequential immunoprecipitation (IP) to effectively eliminate non-specific proteins, while the FLAG peptide allows for the detection of bait protein. The lentiviral transduction was used to integrate the mFOXH1-SFB cassette into the genome of E14 mESCs. On day seven of the antibiotic selection, colonies of E14 mESCs transduced by lentivirus survived and propagated while non-transduced E14 mESCs did not (Figure 2.1B). Western blot analyses showed the expression of mFOXH1-SFB recombinant protein in transduced but not non-transduced mESCs (Figure 2.1C).

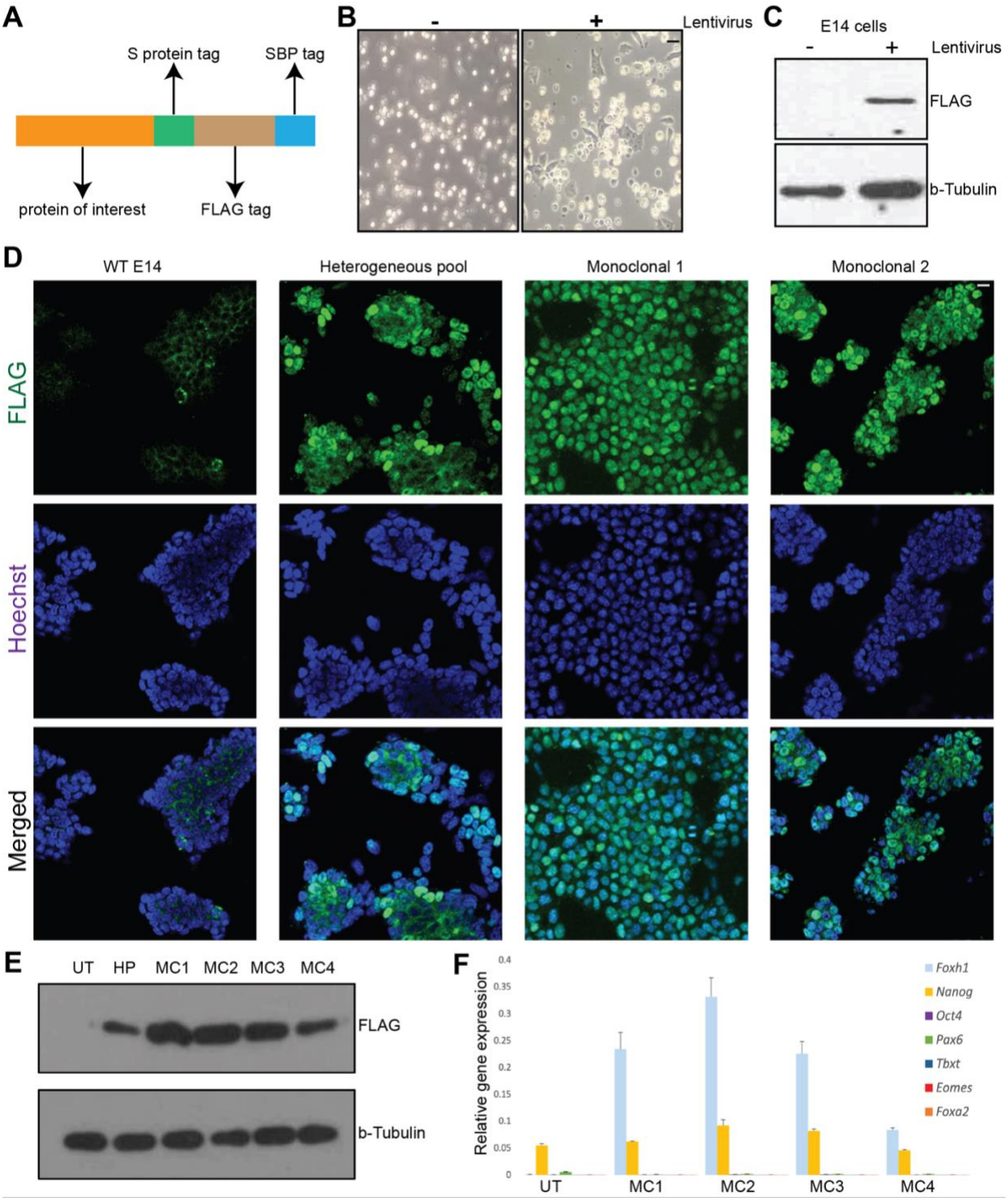

Figure 2.1: Generation of E14 mESCs stably expressing recombinant mFOXH1-SFB.

(**A**) A schematic diagram of the recombinant protein with SFB triple tags, S-protein tag, FLAG tag, and streptavidin-binding peptide (SBP) tag at *C*-terminus. (**B**) E14 cells transduced by

lentivirus show resistance to puromycin after seven days of selection. (**C**) Western blot of E14 cells expressing mFOXH1-SFB recombinant proteins, mFOXH1 is detected by FLAG antibody; b-Tubulin is used as a loading control. (**D**) Immunofluorescence images illustrating that mFoxh1-SFB is localized to the nucleus of E14 cells. The white dash on the right top corner denotes 10 microns. (**E**) Western blot showing the protein levels of mFOXH1-SFB in non-transduced E14 (UT), heterogeneous cell pool (HP), monoclonal line 1 (MC1), monoclonal line 2 (MC2), monoclonal line 3 (MC3), and monoclonal line 4 (MC4) of mESCs. mFOXH1 is detected by FLAG antibody; b-Tubulin is used as a loading control. (**F**) RT-qPCR analyses of pluripotent marker and early germ layer marker genes in E14 mFOXH1-SFB cell lines. Error bars represent standard deviation values from two technical replicates.

Given that this initially established mFOXH1-SFB E14 mESC line is a heterogeneous pool consisting of cells with different integration sites and copies, we isolated monoclonal lines of mFOXH1-SFB mESCs that descend from a single cell using the dilution method. To confirm the physiological state of mFOXH1-SFB proteins, we used immunofluorescence staining to show that mFOXH1-SFB is strictly localized to the nucleus (Figure 2.1D), which follows the role of mFOXH1 as a transcription factor. Compared to the initial heterogeneous pool, isolated monoclonal mFOXH1-SFB E14 mESCs exhibit more uniform expression levels of mFOXH1-SFB. Western blot analyses demonstrated variable protein levels of mFOXH1-SFB in clones of mFOXH1-SFB E14 mESCs (Figure 2.1E). Gene expression analyses showed high RNA levels of mFoxh1 and pluripotency genes but not early germ layer marker genes, confirming that mFoxh1 overexpression does not affect the pluripotency of embryonic stem cells (Figure 2.1F).

FOXH1 participates in diverse protein-protein interaction networks

Liquid chromatography-mass spectrometry was performed on mFOXH1-SFB IP eluate to identify mFOXH1 associated proteins. 342 proteins consistently recovered from three independent samples (heterogeneous pool cells, monoclonal 1 cells, and monoclonal 2 cells) are considered as potential putative mFOXH1 interactants (Table 2.1). Gene ontology enrichment analysis showed that mFOXH1 participates in diverse cellular processes (Figure 2.2A). We highlight that mFOXH1 interacts with proteins functioning in various aspects of chromatin regulation (Figure 2.2B), consistent with the notion that TFs often recruits coregulators to influence the local chromatin states (Spitz and Furlong, 2012). We obtained a list of mFOXH1 interacting proteins functioning in chromatin organization for further analyses (Figure 2.2C). SMAD2 and SMAD3, known FOXH1 interactants through Activin/Nodal signaling (Massagué, 2012), are captured. Proteins positively or negatively regulating chromatin are identified as potential mFOXH1 interacting proteins. For example, SMARCA5 and HELLS are known regulators related to SWI-SNF chromatin remodeling complexes (Oppikofer et al., 2017; Dennis et al., 2001). NSD1 is a histone methyltransferase depositing H3K36 methylation during transcription elongation (Lucio-Eterovic et al., 2010). Alternative, TET1 and OGT are assembled into DNA methyltransferase complexes (Hrit et al., 2018). SUZ12, JARID2, and EZH2 are subunits of Polycomb repressive complex 2 (PRC2) that deposits H3K27me3 modification on inactive genes (Chammas et al., 2019). HDAC1 is a histone deacetylase removing active histone acetylation modifications, thereby facilitating the repression of genes (Seto and Yoshida, 2014). These results suggest that mFOXH1 participates in diverse cellular processes which can either positively or negatively regulate transcription.

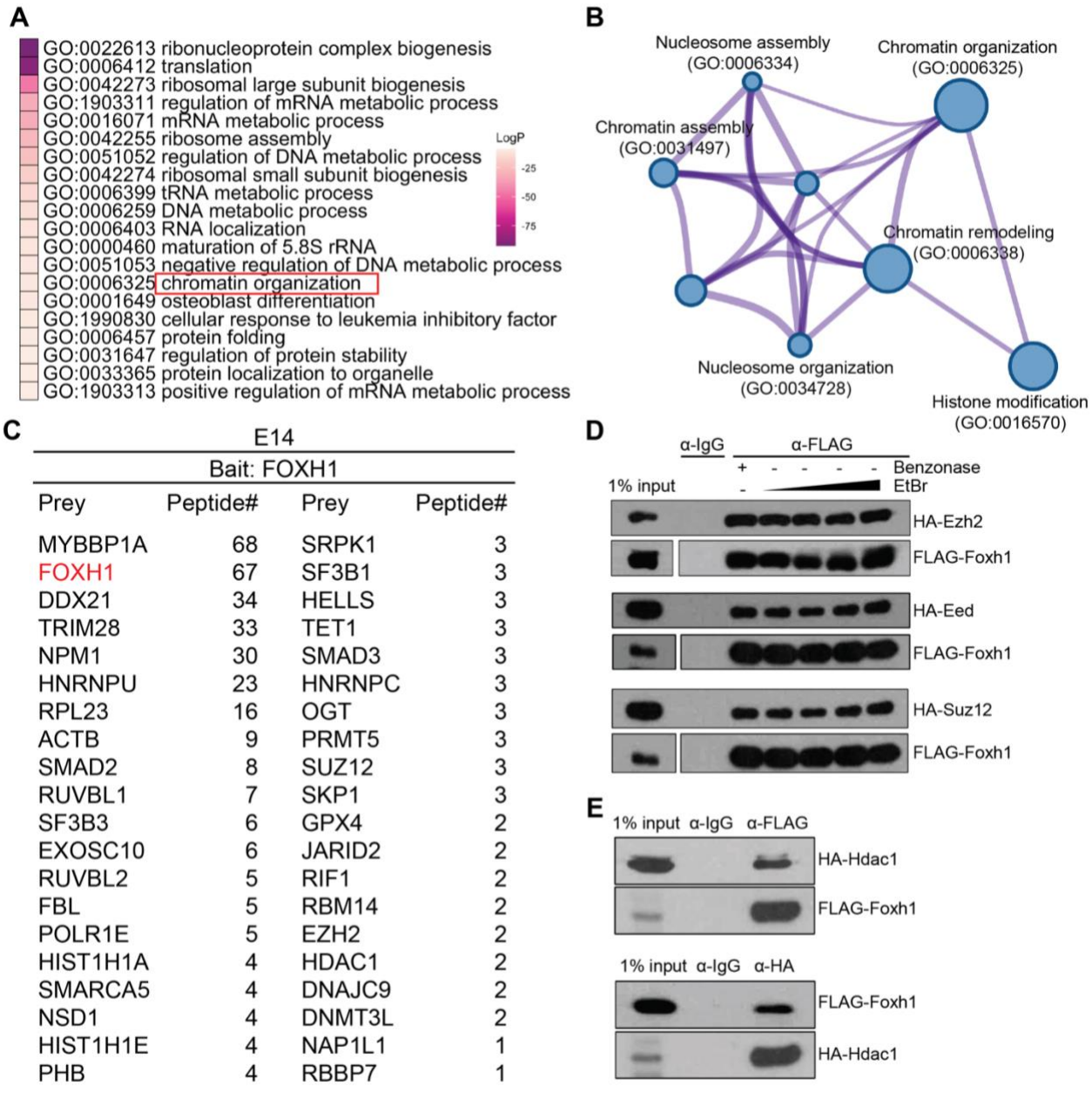

Figure 2.2: Foxh1 interacts with negative epigenetic regulators PRC2 subunits and Hdac1.

(**A**) Gene ontology enrichment analysis of 342 potential putative Foxh1 interacting proteins identified by mass spectrometry. (**B**) Interactome diagram illustrating Foxh1 interacting proteins that function in chromatin regulation. The sizes of circles are modified based on p-values from Metascape. (**C**) The list of proteins belonging to GO term chromatin organization, plus bait protein FOXH1, and known interactants SMAD2 and SMAD3. The numbers indicate the mean

of total peptides detected between three biological replicates. (**D**) Western blot showing the interaction between Foxh1 and PRC2 core subunits Ezh2, Eed, and Suz12 in the lysate from HEK293T cells. EtBr: ethidium bromide. (**E**) Interaction between Foxh1 and Hdac1 in the lysate from HEK293T cells.

Though Foxh1 is shown to interact with Groucho family repressors (Reid et al., 2016; Charney et al., 2017), specific effectors of the repressive action are unknown. Our mass spectrometry results led us to hypothesize that Foxh1 can recruit PRC2 and HDAC-containing complexes to facilitate transcriptional repression. To confirm interactions identified in mESCs are also present in *Xenopus*, we performed co-IP experiments using respective *Xenopus* proteins and found that Foxh1 interacts with all three core subunits (Ezh2, Eed, and Suz12) of PRC2 independent of nucleic acids (Figure 2.2D). The interaction between Foxh1 and Hdac1 was also confirmed (Figure 2.2E). These validated Foxh1 interactions led us to predict that Foxh1 functions as a transcriptional repressor by the recruitment of PRC2 and HDAC-containing complexes in *Xenopus*.

The role of Foxh1 in developing embryonic ectoderm

Foxh1 is a well-characterized co-activator of Nodal signaling; the validated interactions between Foxh1 and repressive epigenetic modifiers raise the question of whether Foxh1 can function as a transcriptional repressor, presumably independent of Nodal signaling. Given that Nodal signaling but not Foxh1 is spatially restricted, we speculate that Foxh1 bears a repressive function in the absence of Nodal signaling within the ectoderm. We, therefore, hypothesize that Foxh1 recruits

repressive epigenetic modifiers such as Hdac1 to repress mesendodermal genes in the developing ectoderm.

To test this hypothesis, we first sought to identify genes regulated by Foxh1 in the ectoderm. One-cell staged embryos were injected with control or Foxh1 morpholino (MO) at the animal cap site, and incubated to the late blastula stage. Animal cap tissues were dissected and then incubated to the early gastrula stage for RNA sequencing (Figure 2.3A). Consistent with the previous study (Chiu et al., 2016), Foxh1 morphants display incomplete blastopore closure and impaired AP axis; these abnormalities are reduced in rescued embryos by ectopic-RNA expression of *foxh1* (Figure 2.3B). Differential gene expression analysis between Foxh1 morphant and control ectoderm tissues revealed 192 upregulated and 384 downregulated genes (Figure 2.3C). Spatial gene expression analysis revealed that differentially regulated genes by Foxh1 in ectoderm does not exhibit any germ layer enrichment (Figure 2.3D). We then tested the function of Foxh1-Hdac1 interaction. The genomic binding profiles of Hdac1 share high similarities to the genomic binding profiles of Foxh1 from mid-blastula to early gastrula, suggesting that Foxh1 and Hdac1 occupies similar genomic regions (Figure 2.3E). A list of 521 Hdac1 direct repressing genes were generated from overlapping Hdac1 bound genes to TSA-induced genes in gastrula ectoderm (Figure 2.3F). Of 192 Foxh1 repressed genes, 12 genes are also suppressed by Hdac1 (Figure 2.3G). A small percentage (6.3%) of genes regulated by Foxh1-Hdac1 suppression suggests that additional mechanisms (such as other TFs recruit Hdac1 in the absence of Foxh1) must be present to inhibit the expression of non-ectodermal genes in the developing ectoderm (Figure 2.3H).

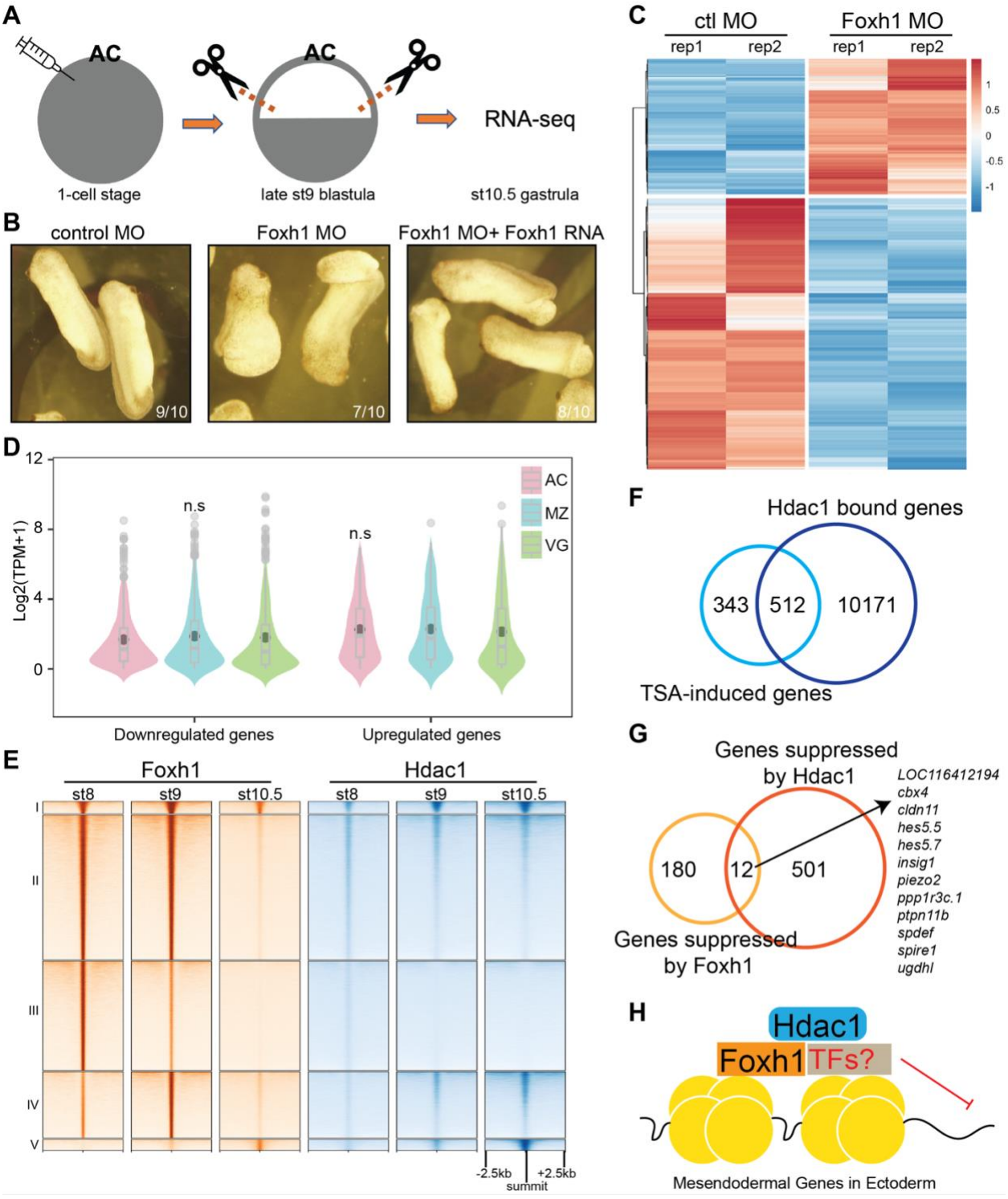

Figure 2.3: Foxh1 participates in mesendodermal gene suppression in the ectoderm.

(**A**) A schematic diagram showing the animal cap (AC, presumptive ectoderm) dissection for embryos. (**B**) Foxh1 morphant embryos display abnormal anterior-posterior axis and incomplete

blastopore closure. (**C**) Heatmap representing differentially expressed genes in Foxh1 morphant versus control ectodermal tissues. (**D**) Violin plot showing the spatial expression levels of genes differentially regulated by Foxh1. AC: animal cap (presumptive ectoderm), MZ: marginal zone (presumptive mesoderm), and VG: vegetal mass (presumptive endoderm). (**E**) Clustered heatmaps depicting the genomic binding signals of Foxh1 and Hdac1 in st8, st9, and st10.5 embryos. (**F**) Venn diagram illustrating Hdac1 directly suppressed genes in the ectoderm. (**G**) Venn diagram depicting genes co-suppressed by Foxh1 and Hdac1 in the ectoderm. (**H**) Model of Foxh1 and other TFs recruiting Hdac1 to suppress mesendodermal genes in the ectoderm.

Discussion

Here we uncover the proteomic interaction network of transcription factor FOXH1 (Table 2.1). In mESCs, FOXH1 interacts with proteins functioning in RNA bioprocesses, ribosomal bioprocesses, and chromatin organization (Figure 2.2A). Interactions of Foxh1 with repressive epigenetic modifiers, PRC2 subunits and Hdac1, are reconstituted with *Xenopus* proteins, underlying a transcriptional repressor function for Foxh1. However, only a small subset of genes is co-repressed by Foxh1 and Hdac1 in the ectodermal lineage of *Xenopus*.

FOXH1 participates in diverse aspects of chromatin regulation in mESCs (Figure 2.2B). In agreement with the known role of FOXH1 as a Nodal signaling co-activator, our mass spectrometry identified both SMAD2 and SMAD3 as FOXH1 interactants supporting the validity of our experiments. Transcription factors bind to specific DNA motifs and recruit epigenetic regulators to facilitate the assembly of basal transcriptional machinery (Chen and Dent, 2013; Zaret, 2020). Our mass spectrometry result revealed several known epigenetic regulators involved in various aspects of chromatin regulations, including nucleosome assembly,

histone modifications, and chromatin organization (Figure 2.2B, 2.2C). Foxh1 binds to the zygotic genome as early as 32~64 cell stages in early *Xenopus* embryogenesis, suggesting it might act as a passive pioneer transcription factor (Charney et al., 2017). The capture of SMARCA5 and HELLS, regulators related to SWI/SNF complex, provides a mechanism where Foxh1 may exploit the activity of SWI/SNF complex to access the highly compacted embryonic genome during such an early stage of development. Foxh1 is reported to interact with Groucho family co-repressors (Reid et al., 2016; Charney et al., 2017). Nodal/Foxh1 inhibits *nodal5* and *nodal6* in the ventral vegetal blastula embryos in *Xenopus* (Kofron et al., 2004). These findings infer a transcriptional repressor role for Foxh1. Indeed, our results elucidate a potential mechanism by which Foxh1 recruits PRC2 (Figure 2.2D) and Hdac1 (Figure 2.2E) to confer repressive chromatin states through H3K27me3 deposition and histone hypoacetylation, thereby suppressing transcription. Together, our Foxh1 protein-protein interaction network suggests that Foxh1 can both positively and negatively regulate transcription through the engagement of epigenetic regulators.

The protein-protein interaction between Foxh1 and negative epigenetic modifiers raises the question of where such an interaction is employed in the physiological context. We propose that Foxh1 acts as a transcriptional repressor in the ectoderm of developing embryos (Figure 2.3H) for the following evidence: (1) Nodal signaling is absent in the ectoderm (Hill et al., 2001) while (2) Foxh1 is ubiquitously expressed (Chiu et al., 2014; Charney et al., 2017) in embryos and (3) interacts with PRC2 subunits and Hdac1 (Figure 2.2D, 2.2E), and (4) Hdac1 genomic binding profiles highly overlap with Foxh1's (Figure 2.3E). Differential gene expression analysis revealed that Foxh1 can both positively and negatively regulate gene expression in the ectoderm (Figure 2.3C). Genes repressed by Foxh1 encompass only a small overlap (~10%) with genes

inhibited by HDAC activity in the ectoderm (Figure 2.3G). These findings support the notion that Foxh1 can function as a transcriptional repressor through Hdac1 activity. However, this mechanism only applies to a small subset of genes, suggesting that additional factors may be present to compensate for the loss of Foxh1 because transcription factors often function in a combinatory fashion (Ravasi et al., 2010). Previous studies in *Xenopus* showed that Foxh1 binds to genomic regions overlapped with Otx1, Vegt, Sox3, and Sox7 (Paraiso et al., 2019; Jansen et al., 2022). It is highly likely that, Sox3, for instance, functions cooperatively with Foxh1 to suppress mesendodermal genes in the ectoderm. Thus, a loss or reduction of Foxh1 alone is not sufficient to relieve this suppression. Additional investigations are needed to further uncover the repressive role of Foxh1.

Materials and Methods

DNA plasmid constructs

To stably express mFOXH1 tagged with SFB tags, mFOXH1 open reading frame (ORF) was obtained from reverse-transcribed cDNA and cloned into the destination vector that contains *C*-terminal triple (S tag-Flag tag-SBP tag, or SFB) tags (Wang et al., 2014). mFOXH1-SFB fusion protein is then cloned into pLV-EF1a-IRES-Puro vector (Addgene, 85132) for lentivirus production.

To test the interaction between Foxh1 and candidate interactants (Ezh2, Eed, Suz12, Hdac1), ORFs of candidate interactants were obtained using reverse-transcribed cDNA and cloned into pCS2+ (modified from Addgene 102860) vector to express *N*-terminal 1X HA tagged fusion proteins. Foxh1 was also cloned in pCS2+ vector for the expression of *N*-terminal 3X FLAG tagged fusion proteins.

Cell culture and Lentiviral transduction

E14 mESC cells were cultivated in KnockOut D-MEM (Thermo Fisher, 10829018) with ES-cell certified FBS (R&D, S10250H), GlutaMAX (Thermo Fisher, 35050061), MEM non-essential amino acids (Thermo Fisher, 11360070), penicillin-streptomycin (Fisher Scientific, 15140122), b-mercaptoethanol (1000X, 3.5ul of 100% b-ME in 1ml D-PBS), and murine LIF (Millipore Sigma, esg1107) on gelatinized culture dishes. Lentivirus containing pLV-EF1a-mFOXH1-IRES-Puro was packaged with psPAX2 (Addgene, 12260) and pMD2.G (Addgene, 12259) and transduced on lowly seeded mESCs for two days. Puromycin (5ug/ml, A1113803) was applied consecutively to select target clones. After 14 days of selection, single clones were picked, and expanded using the dilution method.

HEK293T cells were cultivated in DMEM (Thermo Fisher, 11965118) with FBS (R&D, S11195) and penicillin-streptomycin (Fisher Scientific, 15140122). HEK293T cells were transfected with polyethylenimine (PEI, Polysciences, 23966) accordingly (Longo et al., 2013).

Animal Model and Morpholinos

Xenopus tropicalis embryos were obtained by *in vitro* fertilization according to Ogino et al. (2006) and staged according to Nieuwkoop and Faber (1994). All embryos were cultured in 1/9X Marc's modified Ringers (MMR) at 25 °C. Embryos were dissected at the late blastula stage (6 hpf) and cultured to the early gastrula (7 hpf) for harvest. Animals were raised and maintained following the University of California, Irvine Institutional Animal Care Use Committee (IACUC). Animals used were grown in the laboratory and/or purchased from National *Xenopus* Resource.

For Foxh1 morpholino (MO) experiments, injections were performed at the 1-cell stage in the animal cap region. 22.5 ng of *foxh1* MO (Chiu et al., 2014) and standard control MO (Genetools) were used. To rescue Foxh1 morphants, *foxh1* mRNA is *in vitro* transcribed (Invitrogen, AM1340) from pCS2+-3X FLAG-foxh1 and co-injected with *foxh1* MO at 30pg/embryo.

Immunoprecipitation and co-Immunoprecipitation

For mFOXH1 mass spectrometry (MS), thirty million respective mFOXH1 mESCs were harvested and proceeded as described (Wang et al., 2014). Briefly, cells were lysed in 1X NETN buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 0.5mM EDTA, 0.5% NP40, PICs). The clear lysate was incubated with streptavidin beads (Thermo Scientific, 88816) for 4 hours, washed with 1X NETN five times, and eluted with biotin overnight (2mg/ml in 1X NETN). The clear eluted lysate was incubated with S-protein beads (Millipore-Sigma, 69704-4) for 4 hours, washed with 1X NETN five times, and eluted using 2X SDS loading buffer. Eluate was subjected to 10% SDS-PAGE gel for 10 mins and stained with coommassie-blue. Stained protein bands were cut out and sent to Harvard Taplin MS facility for LC-MS.

Co-immunoprecipitation (co-IP) was performed as following: cells were harvested in 1X cold PBS and lysed in 1X cell lysis buffer (50mM Tris-HCl pH7.5, 100mM NaCl, 1% NP-40, 0.05mM DTT, and PICs). The clear cell lysate was then subjected to IP using anti-FLAG (Millipore-Sigma, F3165), or anti-HA (Millipore-Sigma, 11583816001) with protein A/G magnetic beads (Thermo Scientific, PI88802). Beads were washed five times with lysis buffer, followed by the elution using 2X SDS loading buffer (100mM Tris-HCl pH6.8, 20% glycerol,

0.2% bromophenol blue, 4% SDS, and 5% b-mercaptoethanol). The protein eluate was then analyzed using western blotting.

Western Blotting

Cells were directly harvested in 1X SDS loading buffer (50mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 1% b-mercaptoethanol, 0.02% bromophenol blue), boiled at 95C for 10 mins, and subjected to western blotting using anti-FLAG (Millipore-Sigma, F3165), anti-Tubulin (Sigma, T5168), and anti-HA (Millipore-Sigma, 11583816001).

Gene Expression Analysis and RNA-seq

For RT-qPCR, reverse-transcription assays were performed using Maxima reverse transcriptase (Thermo Fisher EP0741). RT-qPCR was performed using LightCycler 480 SYBR Green I master mix (Roche). For quantification of gene expression, the 2^ -ΔΔCt method was used. *gapdh* was used as a control gene for normalization. Primer sequence information is provided in Table 2.2.

For RNA-seq, total RNA from dissected animal cap explants was extracted using Trizol as described (Amin et al., 2014). mRNA was then isolated using NEBNext PolyA mRNA Magnetic Isolation Module (NEB E7490S). Sequencing libraries were prepared using NEBNext Ultra II RNA library prep kit (NEB E7770S) and sequenced by the Illumina NovaSeq 6000 with 100bp paired-end reads. All experiments were done in 2 biological replicates. All sequencing samples were aligned using STAR v2.7.3a (Dobin et al., 2013) to *Xenopus tropicalis* genome v10.0 (http://www.xenbase.org/, RRID:SCR_003280) to obtain raw read counts. Differentially expressed genes were identified using edgeR v3.36.0 (Robinson et al., 2010) with parameters

greater than 2-fold change and less than 0.05 false discovery rate (FDR, also known as the adjusted p-value) in R v4.1.2 (R Core Team, 2022). Gene ontology analysis was performed on Metascape (Zhou et al., 2019) with default parameters (min overlap= 3, p-value cutoff= 0.01, and min enrichment= 1.5).

Additional Analysis

Mass spectrometry: To obtain the high confidence FOXH1 interactants, proteins detected in each mass spectrometry experiment were overlapped (three biological independent samples: heterogeneous pool cells, monoclonal 1 cells, monoclonal 2 cells), followed by subtracting proteins from the negative control (wildtype E14 mESCs). Spatial expression analysis: Sequencing samples (Blitz et al., 2017) were aligned using STAR v2.7.3a (Dobin et al., 2013) and quantified by RSEM v1.3.3 (Li and Dewey, 2011) to obtain TPM values. TPM values for marginal zones are calculated as the mean TPMs of dorsal, lateral and ventral mariginal zones. ChIP-seq: Sequencing samples (Chiu et al., 2014; Charney et al., 2017; Zhou et al., BioRxiv) were aligned to *Xenopus* tropicalis v10.0 genome (http://www.xenbase.org/, RRID:SCR_003280) using Bowtie2 v2.4.4 (Langmead and Salzberg, 2012). PCR duplicates were removed using Samtools v1.10 (Li et al., 2009). Irreproducibility discovery rate (IDR) analysis (Li et al., 2011) was used to identify high-confidence peaks called by Macs2 v2.7.1 (Zhang et al., 2008) between two biological replicates following ENCODE3 ChIP-seq pipelines (IDR threshold of 0.05) (https://docs.google.com/document/d/1lG_Rd7fnYgRpSIqrIfuVlAz2dW1VaSQThzk836Db99c/edit). Clusters are made by Bedtools v2.29.2 (Quinlan and Hall, 2010) denoting the following: Cluster I: Foxh1 persistent peaks at st8, 9 and 10.5; Cluster II: Foxh1 overlap peaks in any of two stages out of st8, 9 and 10.5; Cluster III: Foxh1 unique peaks

at st8; Cluster IV: Foxh1 unique peaks at st9; Cluster V: Foxh1 unique peaks at st10.5. Clustered heatmaps were generated using DeepTools v3.5.0 (Ramírez et al., 2014).

Table 2.1: 342 high confidence proteins identified from mass spectrometry

MYBBP1A	MTDH	RPS15A	HNRNPA3	SSB	SRSF6	GNAI2	HIST1H1E
TRIM28	ACTB	KPNA2	EBNA1BP2	MCM5	RBM28	JARID2	ABCF2
NCL	PABPC1	PNPLA6	RPS19BP1	ZSCAN4C	DHX37	IMP3	DNAJC21
HNRNPU	RPL8	ATP5C1	HNRNPC	EXOSC10	IARS	DNMT3L	POLR2A
HSPA8	SYNCRIP	EPRS	RFC4	ATP1A1	TCP1	POLR2E	RPF2
HSP90AB1	HNRNPH1	RPS24	ATP2A2	FRG1	RTCB	SPTLC1	MRPS30
HNRNPM	PTBP1	MDN1	NOL6	SKP1	URB2	DUS3L	HDAC1
RPL7	TDH	RPL21	DDX28	RBMXL2	ALYREF	RPLP1	AP2A2
RPS4X	CKAP4	GLYR1	SRSF3	C1QBP	TUBB4B	PISD	LARP7
ATAD3	DDX1	ABCE1	MYEF2	DDX47	RPL34	PPAN	AP2M1
DDX21	UTP20	RUVBL1	PRDX1	ESCO2	NARS	TMEM214	SPTLC2
RPL4	NAT10	RPS7	TARDBP	HSPD1	RPL35	RRP15	CDIPT
TRIM71	RCC2	CAD	POLR1B	FAM98B	VCP	DNAJB6	CCT3
PDCD11	RRP12	RPL26	TRIM25	SLC25A1	SSR1	HSP90B1	TRIP12
FOXH1	HSPA2	RPS26	PES1	RPS27	NSUN2	CANX	FAU
RPS9	RPL23A	RPL14	SGPL1	L1TD1	SMAD2	NOL9	DDX49
LBR	RPL18	RPL18A	VDAC2	NOP53	RPN2	YBX3	MRPL39
GNL3	NOP2	YBX1	RPL22	EXOSC9	NOP16	MRPL15	GTF3C1
RPS2	RPL30	SERBP1	POLR1E	RBM39	TIMM23	CFL1	MRPS26
DDX5	RPL27	RPLP2	AP2B1	DDX10	HNRNPAB	RSL24D1	IGF2BP3
RPL6	DRG1	ASPH	DNAJA3	RPL36-PS3	PCBP3	DDX27	RBBP7
D1PAS1	DNAJA1	CSDE1	TECR	RUVBL2	LYAR	TUBA4A	ERLIN2
NOP56	MCM3	HK2	ILF2	RRP1B	BCAS2	CUL1	PRMT5
RPL7A	RARS	RPL19	GNL3L	RBM19	ARMC10	PTCD3	STAU1
RPL3	RSL1D1	HELLS	RPL17	NOP14	WDR46	RPS19	TEX10
RPS6	DDX17	DARS	CAPRIN1	RFC2	NIFK	FUS	EIF2S3Y
RPL10A	PRPF19	RACK1	RPL10L	DDX56	SRPK1	SMAD3	CEBPZ
RPS11	DHX15	TSR1	PFKP	MRPL21	NAP1L1	DDOST	PCBP1
RPL13	EIF2S2	RPSA	KPNB1	RFC5	AIMP1	GRSF1	NEMF
RPL13A	TUFM	ABCF1	NXF1	SLC25A3	NISCH	G3BP1	POLRMT
IGF2BP1	DDX51	SUZ12	NOP58	RPL27A	RFC1	GPX4	RARS2
RPL23	RPL24	POLR1C	EXOSC8	RBMX	EZH2	XRN2	MSH6
YME1L1	HDLBP	FXR1	ANKRD17	WDR18	DDX52	DDX31	OGT
CDC5L	POP1	RPS20	HNRNPR	PHB	HNRNPD	RPS28	TRMT1L
NPM1	DHX30	SLC25A5	HNRNPA0	GTPBP4	SRPRB	SMARCA5	TET1
RPS18	FBL	TIMM50	RPS4L	CD3EAP	RRBP1	RBM14	PDE12
HNRNPF	RPL35A	HNRNPA1	NVL	AIFM1	CCT4	ZFR	HIST1H1A
RPS16	RPL31	MTCH2	LAS1L	SLC25A13	FKBP8	NSD1	MSH2
RPS13	RPS25	RPL15	RBFOX2	SF3B1	GTPBP1	KARS	RIF1
HSD17B12	HSP90AA1	DHX9	MARS	FTSJ3	RPL37A	PRKCI	LENG8
SPATA5	RPL10	RBM34	ILF3	AFG3L1	DNAJA2	SLC25A4	POLR1A
RPLP0	RPS8	TUBB4A	SF3B3	SENP3	EEF1G	SRSF4	
TMPO	RPS23	RPN1	EIF4A2	CDKAL1	AIMP2	DNAJC9	

Table 2.2: RT-qPCR primer list used for gene expression assays in mESCs

Target genes	Forward primer sequence	Reverse primer sequence
Eomes	ACCGGCACCAAACTGAGA	AAGCTCAAGAAAGGAAACATGC
Foxa2	CCATCAGCCCCACAAAATG	CCAAGCTGCCTGGCATG
Foxh1	ATTATCCGTCAGGTCCAGGC	TAGAGGAAAGGTTGTGGCGG
Nanog	CCAACCCAACTTGGAACAAC	TATGGAGCGGAGCAGCAT
Oct4	TTGCAGCTCAGCCTTAAGAAC	TCATTGTTGTCGGCTTCCCT
Pax6	CACCAGACTCACCTGACACC	ACCGCCCTTGGTTAAAGTC
T	CAGCCCACCTACTGGCTCTA	GAGCCTGGGGTGATGGTA

CHAPTER 3

Histone Deacetylase 1 Maintains Lineage Integrity Through Histone Acetylome Refinement During Early Embryogenesis

Abstract

Histone acetylation is a pivotal epigenetic modification that controls chromatin structure and regulates gene expression. It plays an essential role in modulating zygotic transcription and cell lineage specification of developing embryos. While the outcomes of many inductive signals have been described to require enzymatic activities of histone acetyltransferases and deacetylases (HDACs), the mechanisms by which HDACs confine the utilization of the zygotic genome remain to be elucidated. Here, we show that histone deacetylase 1 (Hdac1) progressively binds to the zygotic genome from mid blastula and onward. The recruitment of Hdac1 to the genome at blastula is instructed maternally. *Cis*-regulatory modules (CRMs) bound by Hdac1 possess epigenetic signatures underlying distinct functions. We highlight a dual function model of Hdac1 where Hdac1 not only represses gene expression by sustaining a histone hypoacetylation state on inactive chromatin, but also maintains gene expression through participating in dynamic histone acetylation-deacetylation cycles on active chromatin. As a result, Hdac1 maintains differential histone acetylation states of bound CRMs between different germ layers and reinforces the transcriptional program underlying cell lineage identities, both in time and space. Taken together, our study reveals a comprehensive role for Hdac1 during early vertebrate embryogenesis.

Keywords: Hdac1, histone acetylation, germ layer, epigenetics, zygotic genome activation, Xenopus, genomics

Introduction

A fundamental question in early development is the mechanism of zygotic genome activation (ZGA), which requires the degradation of maternal mRNAs and the activation of embryonic transcription (Tadros and Lipshitz, 2009). During ZGA, the embryonic genome undergoes a dramatic reprogramming of gene expression, which is also accompanied by remodeling of the embryonic epigenome. Post-translational modifications to histones are a major epigenetic regulation influencing chromatin structure and thus play a central role in ZGA. Histone acetylation appears during the onset of both minor and major ZGA waves in many species. In *Drosophila*, histone acetylation occurs at mitotic cycle 8 on a few early zygotic genes (Li et al., 2014). miR430, the first zygotically active gene, is marked by H3K27ac in 64-cell staged zebrafish embryos (Chan et al., 2019). Genome-wide H3K27ac is detected at mid blastula shortly after the onset of ZGA in *Xenopus* (Gupta et al., 2014). In mice, the zygotic genome is increasingly marked by H3K27ac from immature and metaphase II oocytes to 2-cell-stage embryos (Dahl et al., 2016). Despite these findings, there remain several major questions. How is the interplay of enzymes regulating histone acetylation employed in developing embryos? What is the role of observed histone acetylation on gene expression? How are the spatial and temporal patterns of histone acetylation established during ZGA?

Histone acetylation occurs on the ε-amino group of the lysine residues within *N*-terminal tails of all four core histones (Inoue and Fujimoto, 1969; Seto and Yoshida, 2014). Acetylation is a reversible process that is directly catalyzed by opposing activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs). In addition, HATs and HDACs can also regulate the acetylation of lysine residues on non-histone proteins (Choudhary et al., 2009). Histone

acetylation is often associated with active gene transcription because the acyl groups neutralize the positive charge on the lysine residues, thereby reducing the affinity of histones to DNA (Wang et al., 2000; Anderson et al., 2001); it also serves as a binding platform for bromodomain (BRD) proteins which scaffold and stimulate the transcriptional machinery (Hassan et al., 2007; Filippakopoulos et al., 2012). The balance between HATs and HDACs directly shapes histone acetylation landscapes and subsequently affects transcriptomes.

HDACs are critical epigenetic regulators because they reset chromatin states by returning acetylated lysine residues on histones to the basal state, which can subsequently be subjected to alternative modifications such as methylation. HDACs are grouped into four classes based on phylogenetic conservation. Class I (HDAC1, 2, 3, 8), Class II (HDAC4, 5, 6, 7, 9, 10), and Class IV (HDAC11) HDACs are zinc-dependent and are related to yeast Rpd3, Had1, and Hos3 respectively; Class III (SIRT1, 2, 3, 4, 5, 6, 7) HDACs, also known as Sirtuins, are NAD+-dependent and are related to yeast Sir2 (Gregoretti et al., 2004; Milazzo et al., 2020). HDACs are well-characterized negative regulators of gene expression during development. For example, HDAC1 silences homeotic genes in cooperation with Polycomb group repressors in *Drosophila* (Chang et al., 2001). In zebrafish, Hdac1 represses Notch targets during neurogenesis (Cunliffe, 2004; Yamaguchi et al., 2005). In *Xenopus*, HDAC activity suppresses Vegt-induced ectopic mesoderm in ectoderm lineages (Gao et al., 2016), represses multi-lineage marker genes at blastula (Rao and LaBonne, 2018) and desensitizes dorsal Wnt signaling at late blastula (Esmaeili et al., 2020). Conversely, HDACs can also positively regulate gene expression. For instance, inhibition of HDAC activities rapidly down-regulates some genes in yeast, suggesting an activator function of HDACs (Bernstein et al., 2000). Genetic deletions or pharmacological application of HDAC inhibitors in cell lines results in both up- and down-regulated genes (Reid

et al., 2005; Zupkovitz et al., 2006; Meganathan et al., 2015). Furthermore, genome-wide studies showed that HDACs occupy genomic loci of active genes, and their binding correlates with gene activities (Kurdistani et al., 2002; Wang et al., 2002; Wang et al., 2009; Kidder et al., 2012). These seemingly opposing functions of HDACs raise an important question as to the exact roles of HDACs on chromatin states and transcriptomes in developing embryos.

In this study, we focus on the role of Hdac1 in regulating the zygotic epigenome and transcriptome during *Xenopus* germ layer specification coinciding with ZGA. Current evidence in *Xenopus* as well as in other non-mammalian systems suggests that the early embryonic genome is rather naïve and that major chromatin modifications occur at or after ZGA (Bonn et al., 2012; Vastenhouw et al., 2010; Gupta et al., 2014; van Heeringen et al., 2014; Hontelez et al., 2015). Thus, the system allows us to probe the earliest establishment of histone acetylation and dissect the link between actions of Hdac1, the zygotic histone acetylome, and zygotic transcriptome during the first cell lineage segregation event. Here, we report that the major Hdac1 binding to the embryonic genome occurs at blastula; the binding of Hdac1 during this stage is dependent on maternal factors. We highlight a dual function model for Hdac1. First, Hdac1 keeps inactive chromatin free of histone acetylation, preventing gene misactivation in respective germ layers. Second, Hdac1 participates in dynamic histone acetylation-deacetylation cycles on active chromatin, sustaining the expression of genes that are enriched in respective germ layers. Taken together, our study reveals a coordinated spatial and temporal regulation by Hdac1 during ZGA.

Results

Hdac1 binds to the genome progressively during blastula and onward

In order to identify the major functional candidates of HDACs during the early *Xenopus* embryogenesis, we examined the temporal RNA expression profiles (Owens et al., 2016) of HDAC family members (Class I, Class II, and Class III HDACs) from the zygote to the beginning of the neurula stage in *Xenopus tropicalis*. The RNA expression level of *hdac1* is the highest among all HDACs examined, followed by *hdac2* (Figure 3.2A). Both Hdac1 and Hdac2 proteins are present in the unfertilized egg to the mid gastrula stage and the overall expression levels of Hdac1 and Hdac2 are relatively constant during this period (Figure 3.1A). These data reveal that Hdac1/2 are the major maternally endowed HDACs functioning during this window of development.

Since Hdac1 modulates various aspects of transcriptional regulation and the chromatin landscape, we examined genome-wide Hdac1 binding patterns during the early germ layer development using chromatin immunoprecipitation assays followed by sequencing (ChIP-seq) (Figure 3.2B). A set of high-confidence peaks at each stage were obtained using irreproducibility discovery rate (IDR) analysis (Li et al., 2011) from two biologically independent samples (Figure 3.2C). IDR analysis reveals that Hdac1 binds to 1,340 regions at the mid blastula (st8), 5,136 regions at the late blastula (st9), and 22,681 regions at the early gastrula (st10.5) stages (Figure 3.1B). Overall, a minority of Hdac1 peaks are unique to each of the blastula stages (Cluster a and b) while a majority of Hdac1 peaks are present across multiple stages (Cluster c) and at the early gastrula stage (Cluster d) (Figure 3.1C, 3.2D). These results reveal that Hdac1 is progressively directed to the genome during early development.

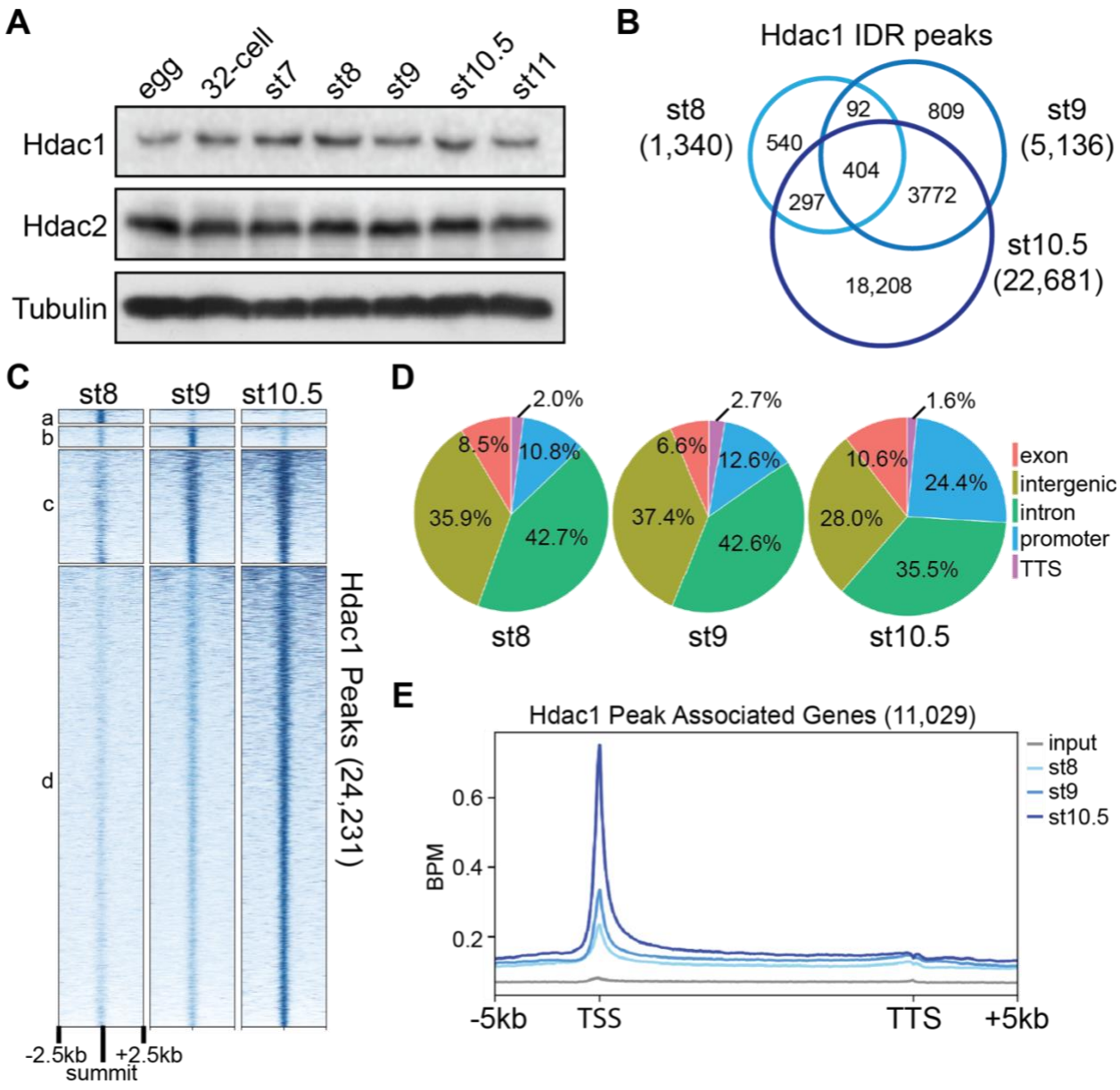

Figure 3.1 Hdac1 binds to genome gradually during early *Xenopus* development

(**A**) Western blot analyses showing Hdac1 and Hdac2 proteins over a time course of early development. β-tubulin is used as a loading control. (**B**) Venn diagram comparing Hdac1 IDR peaks among examined stages. The sum of peaks in st8 and st9 are smaller due to instances where more than one peak from st8 or st9 overlaps the single same st10.5 peak. (**C**) Clustered heatmap showing Hdac1 ChIP-seq signals at each stage over a window of 5kb centered on the

summit of all Hdac1 IDR peaks in descending order. (**D**) Distributions of Hdac1 IDR peaks at each stage across five defined genomic features. The promoter is defined as -1kb to +100bp from TSS (transcription start site) while the TTS (transcription termination site) is defined as -100bp to +1kb from TTS. (**E**) Distributions of Hdac1 ChIP-seq signals within the intervals of 5 kb upstream of gene model 5' ends, gene bodies (normalized for length), and 5 kb downstream of gene model 3' ends at each stage. The signal of st9 input DNA ChIP-seq is used as a negative control. Y-axis values represent reads quantified by bins per million (BPM) at a bin size of 50 bp.

To investigate the differences of Hdac1 genomic occupancy across early stages, we examined various genomic features of regions bound by Hdac1 at each stage. The majority of Hdac1 peaks are found in either intergenic or intronic regions; a minor fraction of Hdac1 peaks is present at exons or transcriptional termination sites. A notable observation is the increased Hdac1 binding to an increased amount of promoter regions over developmental times (Figure 3.1D). We then analyzed the distribution of Hdac1 across 3 stages along the gene bodies of genes bound by Hdac1. A higher enrichment of Hdac1 binding is located near the promoters of genes as development proceeds (Figure 3.1E). The increased binding of Hdac1 around promoters infers that Hdac1 is involved in the epigenetic regulation of these genes during ZGA.

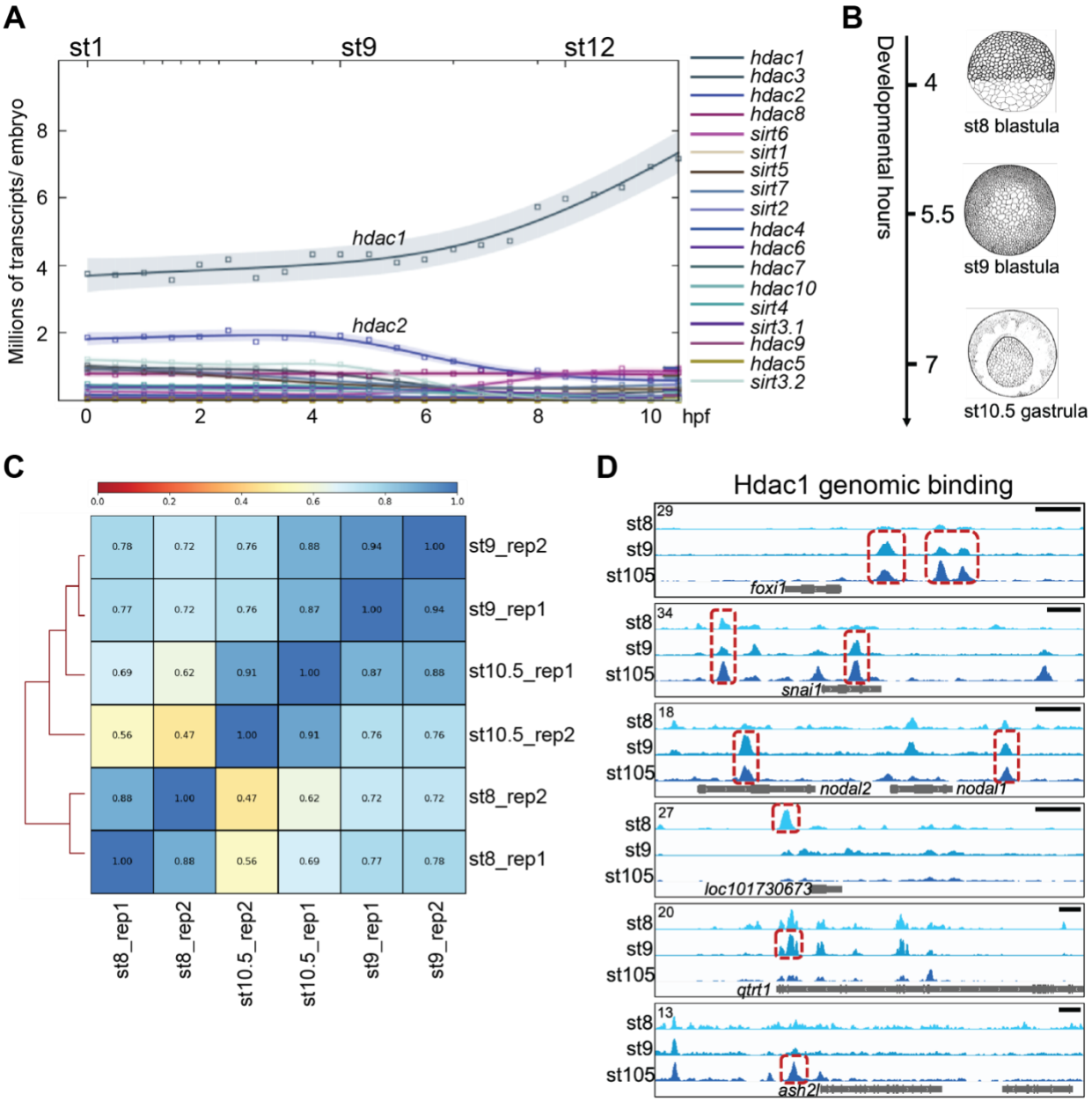

Figure 3.2 Hdac1 genomic occupancy in early *Xenopus* development

(A) Time-course expression levels of HDACs revealed by ribo-depleted RNA-seq (rdRNA-seq) in *Xenopus tropicalis*. (**B**) Graphic illustration of examined stages during early *Xenopus* development: st8 mid blastula, st9 late blastula, and st10.5 early gastrula (Nieuwkoop and Faber, 1994). (**C**) Pearson correlation analyses of intra- and inter-stage Hdac1 ChIP-seq samples. (**D**) Genome browser visualization of Hdac1peaks on *foxi1*, *snai1*, *nodal1*, *nodal2*, *loc101730673*,

qtrt1, and *ash2l*. Red boxes represent IDR peaks passing IDR threshold (see Materials and Methods) on listed genes (not all IDR peaks are boxed). Y-axis values represent scaled track height. Black bars denote an interval of 2kb.

Blastula Hdac1 binding is maternally instructed

Since Hdac1 binds progressively to the genome during ZGA, we investigated the contribution of zygotic factors to the recruitment of Hdac1 at the late blastula stage. We injected α-amanitin, which blocks both transcription initiation and elongation (Chafin et al., 1995), to block embryonic transcription (Figure 3.4A). A high Pearson correlation (0.93) between α-amanitin-injected and control embryos (CT) is observed at all Hdac1 IDR peaks (Figure 3.3A), which is higher than the Pearson correlation (0.87) representing the batch effect when CT is compared to a different batch of same staged embryos (WT). There are no significant signal differences between α-amanitin-injected and control embryos at all Hdac1 IDR peaks (Figure 3.4B). We conclude that Hdac1 binding is independent of zygotic transcription at the late blastula stage, suggesting the importance of maternal factors in Hdac1 recruitment.

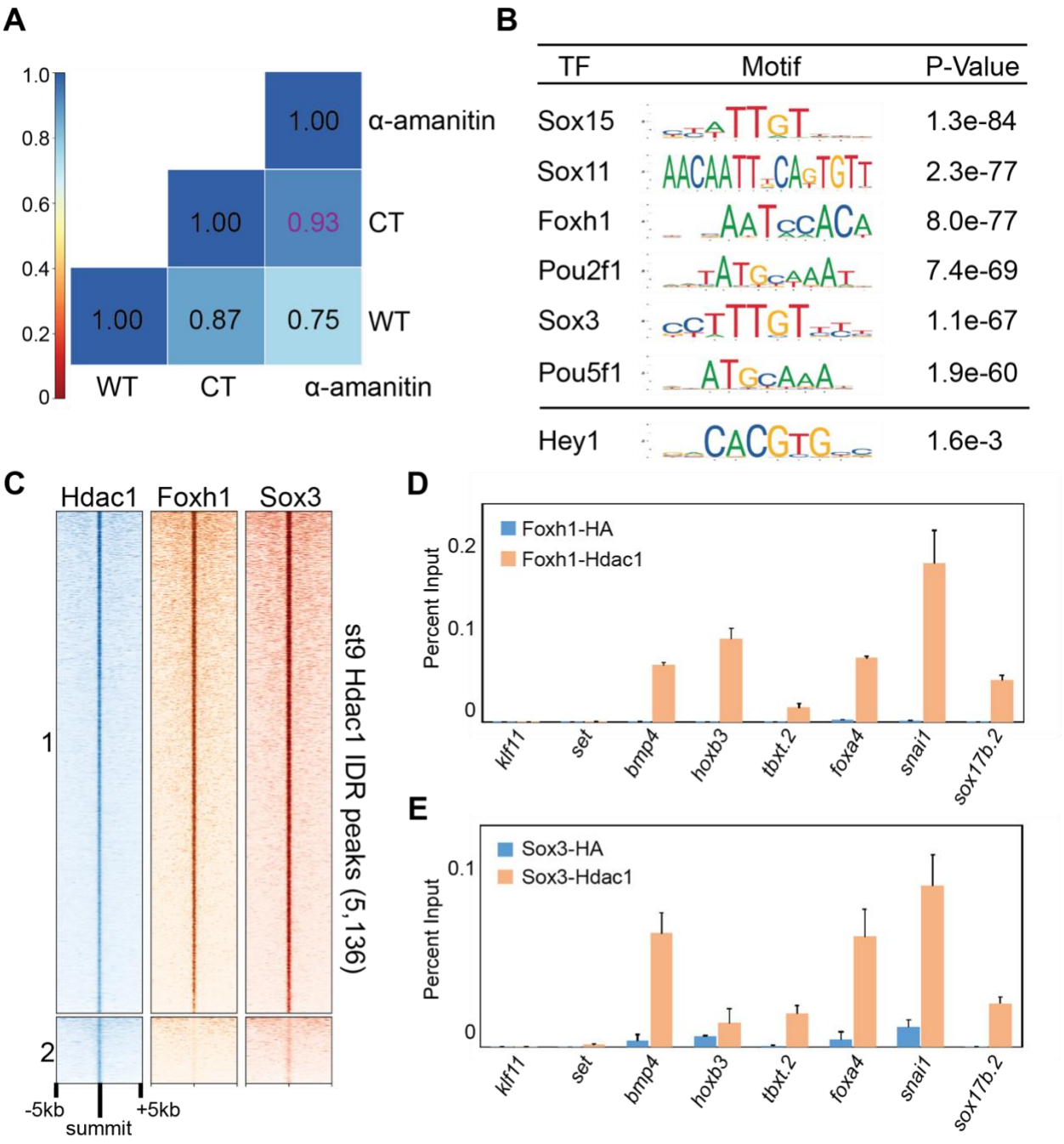

Figure 3.3 Maternal factors instruct Hdac1 recruitment during blastula stages

(**A**) Pairwise Pearson correlation analyses comparing ChIP-seq signals of st9 Hdac1 IDR peaks between listed samples. WT denotes samples from time-course experiments; α-Amanitin denotes α-amanitin injected embryos and CT denotes batch-matched control uninjected embryos. (**B**) Motif analyses of st9 Hdac1 peaks (500bp centered on IDR peak summit). Motif sequence to the

corresponding factor is retrieved from JASPAR. Hey1 is an example of TF motif with low significance. (**C**) Clustered heatmap depicting st9 Foxh1 and Sox3 ChIP-seq signals in a window of 10kb centered on st9 Hdac1 IDR peaks with descending order. (**D-E**) St9 sequential ChIP-qPCR analyses for (**D**) Foxh1 and Hdac1 co-bound regions and (**E**) Sox3 and Hdac1 co-bound regions. anti-HA is used as a negative control. The error bars represent the variation from two technical replicates.

To identify maternal factors that facilitate the recruitment of Hdac1 to the genome, we performed a *de novo* motif search of the DNA sequences of 5,136 st9 Hdac1peaks. Sox, Foxh1, and Pou motifs are found to be the most frequent maternal TF motifs (Figure 3.3B). We thus compared the genomic binding profiles of Hdac1 to two maternal TFs, Foxh1 (Charney et al., 2017) and Sox3. A majority of Hdac1 bound regions (Cluster 1) overlaps with both Foxh1 and Sox3 bound regions while only a small fraction of Hdac1 bound regions (Cluster 2) overlaps the binding of either Foxh1, or Sox3, or neither (Figure 3.3C). More than 80% of Hdac1 peaks overlap with Foxh1 or Sox3 peaks (Figure 3.4C). A positive correlation between Hdac1 binding and Foxh1/Sox3 binding is observed at all Hdac1 IDR peaks (Figure 3.4D). We noted frequent overlapping binding of Hdac1 with Foxh1/ Sox3 (Figure 3.4E), and highly enriched signals of Hdac1, Foxh1, and Sox3 present around promoters of genes (Figure 3.4F). All these observations suggest a role for Foxh1 and Sox3 in Hdac1 recruitment. We also confirmed the co-occupancy of Hdac1 with each of Foxh1 and Sox3 TFs on the same DNA molecules using sequential ChIP-qPCR (Figure 3.3D, 3.3E, 3.4G). Taken together, we propose that maternal TFs such as Foxh1 and Sox3 facilitate Hdac1 recruitment.

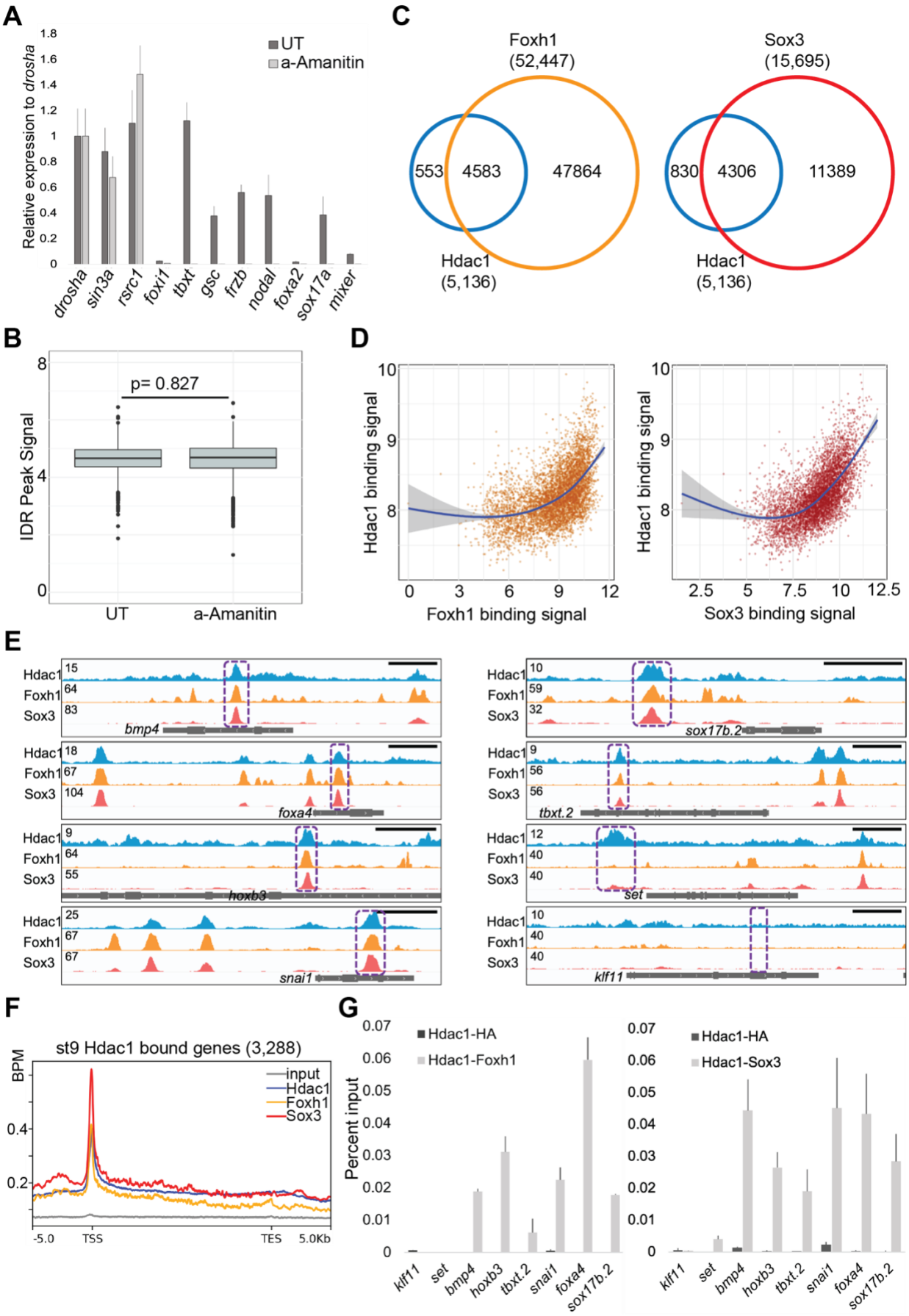
A
Relative expression to drosha
UT
a-Amanitin
drosha
sin3a
rsrc1
foxi1
tbxt
gsc
frzb
nodal
foxa2
sox17a
mixer
C
Foxh1
(52,447)
553
4583
47864
Hdac1
(5,136)
Sox3
(15,695)
830
4306
11389
Hdac1
(5,136)
B
IDR Peak Signal
p= 0.827
UT
a-Amanitin
D
Hdac1 binding signal
Foxh1 binding signal
Sox3 binding signal
E
Hdac1
Foxh1
Sox3
bmp4
foxa4
hoxb3
snai1
sox17b.2
tbxt.2
set
klf11
F
st9 Hdac1 bound genes (3,288)
BPM
input
Hdac1
Foxh1
Sox3
-5.0
TSS
TES
5.0Kb
G
Percent input
Hdac1-HA
Hdac1-Foxh1
Hdac1-Sox3
klf11
set
bmp4
hoxb3
tbxt.2
snai1
foxa4
sox17b.2

Figure 3.4 Hdac1 binding to the genome is instructed maternally

(**A**) RT-qPCR analysis of maternal (maternal-zygotic) genes (*drosha*, *sin3a*, and *rsrc1*) and zygotic genes (*foxi1*, *tbxt*, *gsc*, *frzb*, *nodal*, *foxa2*, *sox17a*, and *mixer*) showing a high efficacy of α-amanitin blocking zygotic transcription. The error bars represent the variation from 2 technical replicates. (**B**) Box plot illustrating Hdac1 signal enrichment at st9 Hdac1 IDR peaks between uninjected and α-amanitin injected embryos. Y-axis values represent log2(reads to IDR peaks/ (peak size* read depth at IDR peaks) +1). *p*-Value is calculated by two-tailed Student's *t*-test. (**C**) Venn diagram showing the overlap between st9 IDR peaks of Foxh1 and Hdac1; Sox3 and Hdac1. (**D**) Scatter plots representing a positive correlation between binding signals of Foxh1 and Hdac1; Sox3 and Hdac1. The binding signal is calculated as log2(reads mapped to st9 Hdac1 IDR peaks per kb +1). Blue lines depict linear regression curves generated from a generalized additive model. Grey areas denote a 95% confidence interval. (**E**) Genome browser visualization of Hdac1, Foxh1, and Sox3 signals on *bmp4*, *foxa4*, *hoxb3*, *snai1*, *sox17b.2*, *tbxt.2*, *set*, and *klf11* genes. Purple boxes represent genomic regions examined in sequential ChIP-qPCR experiments. Y-axis values represent scaled track height. Black bars denote an interval of 2kb. (**F**) Distributions of Hdac1, Foxh1, and Sox3 ChIP-seq signals within the intervals of 5 kb upstream of gene model 5' ends, gene bodies (normalized for length), and 5 kb downstream of gene model 3' ends. The signal of st9 input DNA ChIP-seq is used as a negative control. Y-axis values represent reads quantified by bins per million (BPM) at a bin size of 50 bp. (**G**) Reciprocal sequential ChIP-qPCR analyses to Figure 3.3D& E assessing Foxh1 and Hdac1 co-bound regions; Sox3 and Hdac1 co-bound regions. anti-HA is used as a negative control. The error bars represent the variation from 2 technical replicates.

Hdac1 binds to genomic regions with distinct epigenetic signatures

To further characterize regions bound by Hdac1 across early germ layer development, we examined epigenetic signatures (Gupta et al., 2014; Hontelez et al., 2015; Charney et al., 2017) on Hdac1 peaks across various stages. A high signal density of Ep300 (a HAT) is observed in Hdac1 peaks (Cluster b, c, and d) from the late blastula and onward where RNA polymerase II signals also emerge (Figure 3.6A). This indicates that Hdac1 and Ep300 share similar binding profiles on transcriptionally active genes. We next surveyed several histone acetylation modifications on regions bound by Hdac1 (Figure 3.5A). Consistent with the overlapping binding of Ep300, Hdac1 peaks display a moderate level of H3K9ac (Cluster c and d), high levels of H3K18ac (Cluster b, c, and d), H3K27ac (Cluster b, c, and d), and pan-H3 lysine acetylation (pan-H3Kac) (Cluster b, c, and d). We then examined several histone methylation modifications that are associated with gene activation (Figure 3.6B). H3K4me1, a primed enhancer mark (Creyghton et al., 2010), displays a moderate level of signal density at Hdac1 bound regions (Cluster c and d); H3K4me3, an active promoter mark (Heintzman et al., 2007), displays a high level of signal density at Hdac1 bound regions (Cluster c and d); H3K36me3, a transcription elongation mark (Kolasinska-Zwierz et al., 2009), displays minimal signal density at any Hdac1 bound regions. Taken together, these observations reveal that Hdac1 binds to genomic regions with active epigenetic signatures.

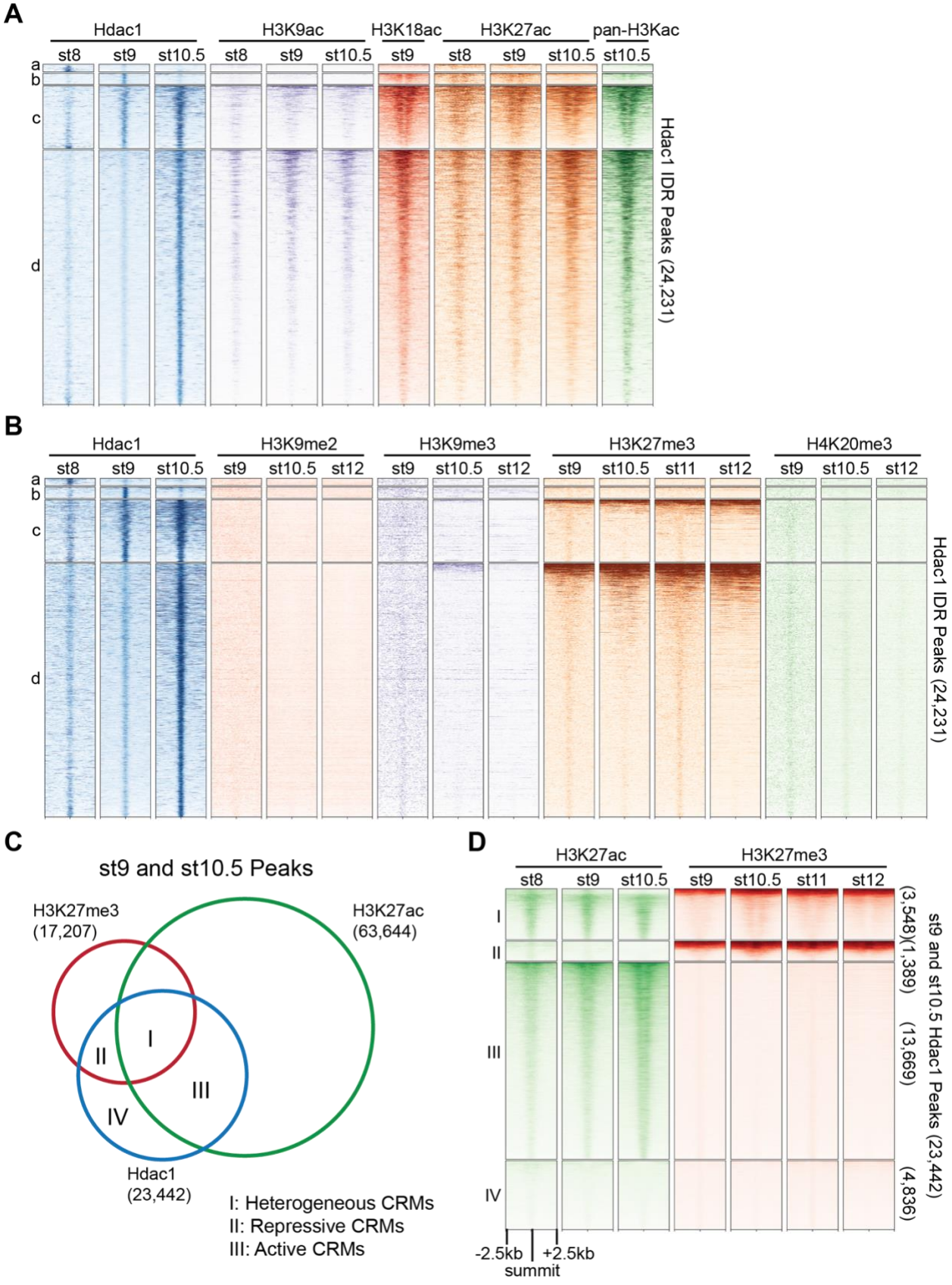
A
Hdac1
H3K9ac
H3K18ac
H3K27ac
pan-H3Kac
st8
st9
st10.5
a
b
c
d
Hdac1 IDR Peaks (24,231)
B
Hdac1
H3K9me2
H3K9me3
H3K27me3
H4K20me3
st11
st12
Hdac1 IDR Peaks (24,231)
C
st9 and st10.5 Peaks
H3K27me3
(17,207)
H3K27ac
(63,644)
I
II
III
IV
Hdac1
(23,442)
I: Heterogeneous CRMs
II: Repressive CRMs
III: Active CRMs
D
H3K27ac
H3K27me3
(3,548)
(1,389)
(13,669)
(4,836)
st9 and st10.5 Hdac1 Peaks (23,442)
-2.5kb
summit
+2.5kb

Figure 3.5 Hdac1 binds to CRMs containing functionally distinct histone modifications

(**A, B**) Clustered heatmaps showing signals from several stages of various (**A**) histone acetylation marks and (**B**) repressive histone methylation marks on Hdac1 peaks. Each cluster corresponds to the same regions present in Figure 1C. (**C**) Venn diagram illustrating Hdac1 peaks overlapping with H3K27me3 and H3K27ac peaks from both st9 and st10.5 combined. (**D**) Clustered heatmaps depicting signals of H3K27me3 and H3K27ac across multiple stages on st9 and st10.5 Hdac1 combined peaks. Clusters denote the same genomic regions in Venn diagram regions demarcated in Figure 3.5C. In panels A, B and D, numbers on the right side indicate the total number of regions in each cluster. The signals are shown in a window of 5kb centered on the summits of Hdac1 peaks presented in descending order of track signal intensities within each cluster.

Hdac1 resets the state of acetylated histones by removing acetyl groups, thus can facilitate the formation of repressive chromatin. Here we analyzed several histone methylation modifications associated with gene repression (Figure 3.5B). Hdac1 binds to genomic regions mostly free of H3K9me2, H3K9me3, or H4K20me3. All three modifications are known to mark constitutive heterochromatin denoting gene-poor areas consisting of tandem repeats (Richards and Elgin, 2002). In addition, a fraction of Hdac1 bound regions (Cluster c and d) display a strong signal density of H3K27me3 (Figure 3.5B), which is known to mark facultative heterochromatin consisting of developmental-cue silenced genes (Trojer and Reinberg, 2007). These observations suggest that Hdac1 binds to facultative heterochromatic regions facilitating the repression of genes instructed by developmental programs.

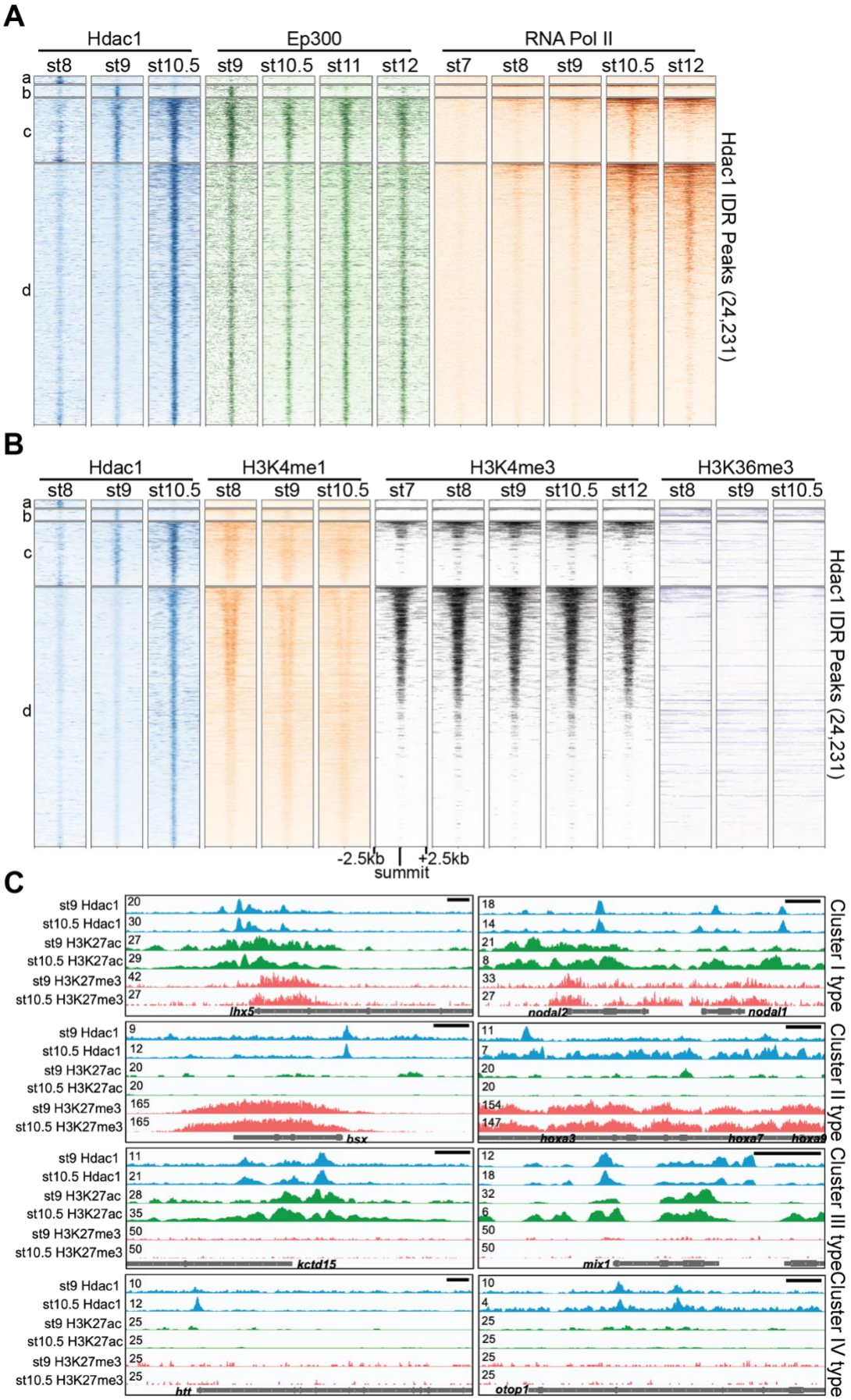
A
Hdac1
st8
st9
st10.5
Ep300
st9
st10.5
st11
st12
RNA Pol II
st7
st8
st9
st10.5
st12
a
b
c
d
Hdac1 IDR Peaks (24,231)
B
Hdac1
st8
st9
st10.5
H3K4me1
st8
st9
st10.5
H3K4me3
st7
st8
st9
st10.5
st12
H3K36me3
st8
st9
st10.5
Hdac1 IDR Peaks (24,231)
-2.5kb
summit
+2.5kb
C
st9 Hdac1
st10.5 Hdac1
st9 H3K27ac
st10.5 H3K27ac
st9 H3K27me3
st10.5 H3K27me3
lhx5
nodal2
nodal1
bsx
hoxa3
hoxa7
hoxa9
kctd15
mix1
btf
otop1
Cluster I type
Cluster II type
Cluster III type
Cluster IV type

Figure 3.6 Hdac1 bound CRMs are marked with distinct epigenetic signatures

(**A, B**) Clustered heatmaps showing signals across several stages of (**A**) Ep300, RNA polymerase II, and (**B**) various active histone methylation marks on Hdac1 peaks. The signals are shown in a window of 5kb centered on the summits of Hdac1 peaks. The signal density within each cluster is ordered by RNA polymerase II in (**A**) and by all samples in (**B**) by descending order. Each cluster corresponds to the same region in Figure 3.1C. (**C**) Genome browser visualization of Hdac1, H3K27ac, and H3K27me3 signals on genes representing Hdac1 CRM clusters: Cluster I (lhx*5*, *nodal1*, *nodal2*), Cluster II (*kctd15*, *mix1*), Cluster III (*bsx*, *hoxa3*, *hoxa7*, *hoxa9*) and Cluster IV (*htt*, *otop1)*. Y-axis values represent scaled track height. Black bars denote an interval of 2kb.

Since Hdac1 peaks (Cluster b, c, d, but not a) are marked by both active and repressive epigenetic signatures, we wonder whether Hdac1 functions differently at epigenetically distinct genomic loci. We compared peaks between two functionally opposing histone modifications H3K27ac and H3K27me3 to Hdac1 peaks, then subdivided these Hdac1 peaks into 4 clusters representing functionally distinct CRM types (Figure 3.5C, 3.6C). Cluster I denotes 3,548 Hdac1 peaks marked by both H3K27ac and H3K27me3 (Figure 3.5D). Given that both H3K27ac and H3K27me3 are subjected to the same lysine residue, we speculate that these regions are differentially marked in space due to heterogeneous cell populations present in the whole embryo. Hdac1 Cluster I peaks are referred to as heterogeneous CRMs. Cluster II denotes 1,389 Hdac1 peaks marked by only H3K27me3 indicating that these regions are associated with inactive developmental genes (Figure 3.5D). Hdac1 Cluster II peaks represent repressive CRMs. Cluster III denotes 13,669 Hdac1 peaks marked by only H3K27ac suggesting that these CRMs

are associated with active genes (Figure 3.5D). Hdac1 Cluster III peaks are considered as active CRMs. Cluster IV denotes 4,836 Hdac1 peaks with neither H3K27ac nor H3K27me3 modifications (Figure 3.5D). Together, we show that different Hdac1 bound regions are subjected to functionally distinct epigenetic modifications suggesting that Hdac1 is differentially utilized at different CRMs.

HDAC activity is required for differential germ layer histone acetylomes

The major function of HDACs is to catalyze the removal of acetyl groups from histones. We hypothesize that Hdac1 differentially regulates histone acetylation of four different Hdac1 CRM clusters (Clusters I-IV in Figure 3.5C). To test this hypothesis, we treated embryos continuously with a widely used pan-HDAC inhibitor, Trichostatin A (TSA) (Yoshida et al., 1990) beginning at the 4-cell stage and followed the development up to tailbud stages. Embryos treated with TSA are developmentally arrested at gastrula (Figure 3.7A). We observed the presence of dorsal blastopore lip suggesting that the progression but not the initiation of gastrulation is defective. A Class I HDAC inhibitor, Valproic acid (VPA) (Göttlicher et al., 2001), also produces a similar phenotype (Figure 3.8A). To investigate whether HDAC inhibition leads to changes in overall histone acetylome, we surveyed six well-known histone acetylation modifications between TSA- and solvent-treated embryos by western blot. Indeed, drastically increased levels of H3K9ac, H3K18ac, and H3K27ac are observed in TSA-treated embryos (Figure 3.7B; Rao and LaBonne, 2018) while H3K14ac, H3K56ac, and H4K20ac are not detected during this stage of development. These data indicate that HDAC activity is required to maintain the proper level of histone acetylation during gastrulation.

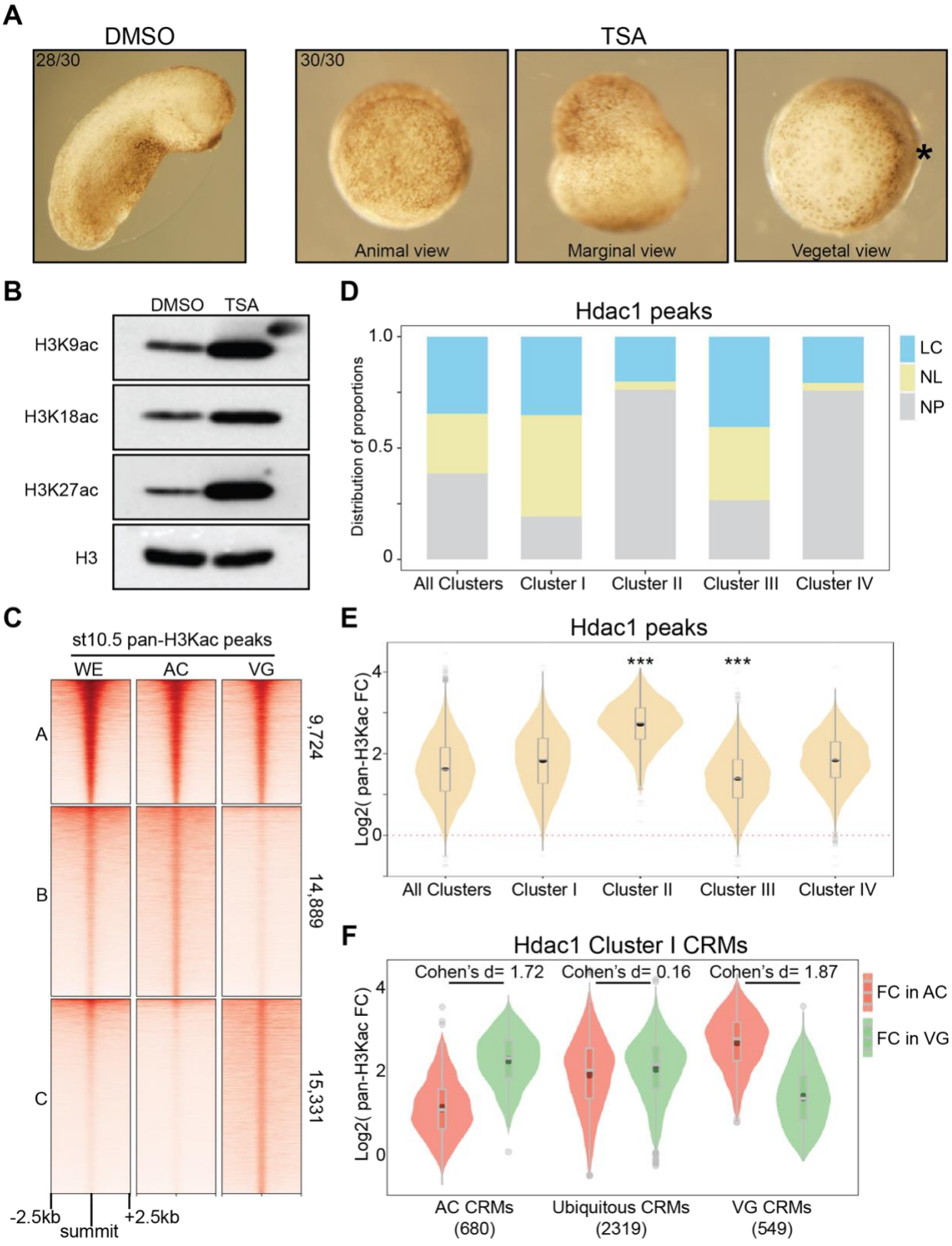
A
DMSO
TSA
28/30
30/30
Animal view
Marginal view
Vegetal view
*
B
DMSO
TSA
H3K9ac
H3K18ac
H3K27ac
H3
C
st10.5 pan-H3Kac peaks
WE
AC
VG
A
B
C
9,724
14,889
15,331
-2.5kb
summit
+2.5kb
D
Hdac1 peaks
Distribution of proportions
1.0
0.5
0
LC
NL
NP
All Clusters
Cluster I
Cluster II
Cluster III
Cluster IV
E
Hdac1 peaks
Log2(pan-H3Kac FC)
4
2
0
***
***
All Clusters
Cluster I
Cluster II
Cluster III
Cluster IV
F
Hdac1 Cluster I CRMs
Cohen's d= 1.72
Cohen's d= 0.16
Cohen's d= 1.87
Log2(pan-H3Kac FC)
4
2
0
FC in AC
FC in VG
AC CRMs
(680)
Ubiquitous CRMs
(2319)
VG CRMs
(549)

Figure 3.7 Hdac1 maintains differential H3 acetylomes between germ layers

(**A**) Embryos treated with 100uM TSA displaying gastrulation defects 24 hours post-fertilization. Asterisk denotes the dorsal side containing the early blastopore lip. (**B**) Western blot analyses showing various histone acetylation modifications affected by HDAC inhibition. anti-H3 is used as a loading control. (**C**) Clustered heatmap depicting signals of pan-H3Kac at st10.5 in whole embryos (WE), animal cap cells (AC), and vegetal mass cells (VG). The signals are shown in a window of 5kb centered on the summits of combined AC and VG peaks presented in descending order of track signal intensities within each cluster. (**D**) Stacked bar graph representing proportions of localized versus non-localized pan-H3Kac signals found at Hdac1 CRM clusters in Figure 3C. A Hdac1 peak is considered to exhibit localized pan-H3Kac if it overlaps with either AC- or VG-specific pan-H3Kac peaks (Cluster B and Cluster C in Figure 3.7C); a Hdac1 peak is considered to exhibit non-localized pan-H3Kac if it overlaps with pan-H3Kac peaks shared between AC and VG (Cluster A in Figure 3.7C). LC: localized; NL: non-localized; NP: not overlap with any pan-H3Kac peak (**E**) Fold changes (FC) of pan-H3Kac signals at Hdac1 CRM clusters (clusters in Figure 3.5C). Red dotted line denotes the level of zero. *** denotes $p<$ 0.001 (Student's *t*-test). (**F**) Fold changes (FC) of pan-H3Kac signals in Cluster I of Hdac1 CRM clusters (clusters in Figure 3.5C) categorized spatially (described in Supplemental Methods).

Given that Hdac1 CRM clusters (Clusters I-IV in Figure 3.5C) are subjected to both active and repressive epigenetic modifications presumably in different germ layers, we compared the general status of H3 acetylome (pan-H3Kac) between two distinct germ layers, the animal cap (AC, presumptive ectoderm) and the vegetal mass (VG, presumptive endoderm). We performed pan-H3Kac ChIP-seq at early gastrula (st10.5) on dissected AC and VG explants from

embryos treated with either TSA or solvent control (Figure 3.8B). A comparison of high-confidence peaks from each explant reveals that a majority (~ 75%) of pan-H3Kac are specifically marked in ectodermal (~ 37%) and endodermal (~ 38%) germ layers (Figure 3.7C, 3.8C). To correlate the differential pan-H3Kac states and the differential gene expression profiles between two germ layers, we assigned high enrichment pan-H3Kac peaks (~ top 30%) to the nearest genes within 10kb and compared the expression levels of these genes in each germ layer (Blitz et al., 2017). Indeed, the expression level of genes enriched with AC-specific pan-H3Kac is higher in AC than VG (Figure 3.8D), and the expression level of genes enriched with VG-specific pan-H3Kac is higher in VG than AC (Figure 3.8E). Well-known genes with germ-layer specific expression exhibit localized pan-H3Kac signals between germ layers (Figure 3.8F). These results illustrate that the histone acetylation profile generally coincides with the animally and vegetally localized expression of transcripts.

To uncover the role of Hdac1 in regulating histone acetylation states of CRM clusters (Clusters I-IV in Figure 3.5C), we first examined the general distribution of pan-H3Kac in these Hdac1 CRM clusters. Consistent with H3K27 modifications, we observed that the majority of both heterogeneous CRMs (Cluster I, both H3K27ac and H3K27me3) and active CRMs (Cluster III, only H3K27ac), but not repressive CRMs (Cluster II, only H3K27me3) are marked by pan-H3Kac; near half of pan-H3Kac marks on heterogeneous or active CRMs are localized either animally or vegetally (Figure 3.7D). The localized pan-H3Kac marks on active CRMs suggests that these CRMs are regionally active in specific germ layers. We then quantitatively examined the levels of pan-H3Kac signals on each CRM within clusters I-IV (Figure 3.5C) after TSA treatment using a ChIP spike-in strategy (Egan et al., 2016). A global increase of pan-H3Kac signals across all Hdac1 CRM clusters is observed (Figure 3.8G), which is consistent with the

result by western blot (Figure 3.7B). We found that CRMs of Hdac1 clusters I-IV respond differently upon HDAC inhibition: repressive CRMs (Cluster II, only H3K27me3) show a significantly higher acquisition (~ 5-fold increase) of pan-H3Kac signals, while active CRMs (Cluster III, only H3K27ac) show the lowest acquisition (~ 3-fold increase) of pan-H3Kac signals when compared to other clusters (Figure 3.7E). Collectively, these results reveal that repressive CRMs are prone to loss of HDAC activity resulting in histone hyperacetylation, while active CRMs gain a modest increase of histone acetylation upon HDAC inhibition.

Lastly, we explored how Hdac1 CRM clusters (Clusters I-IV in Figure 3.5C) differentially respond to HDAC inhibition in specific germ layers. For heterogeneous CRMs (Cluster I, both H3K27ac and H3K27me3), we found that the fold increases of pan-H3Kac signals after TSA treatment are negatively correlated with the levels of endogenous pan-H3Kac signals in both AC and VG explants (Figure 3.8H). Next, heterogeneous CRMs were subdivided into 3 spatial categories: pan-H3Kac enriched animally (AC CRMs), pan-H3Kac enriched vegetally (VG CRMs), and ubiquitous CRMs. We then calculated the fold changes of pan-H3Kac signals after TSA treatment for each CRM within each spatial category. Interestingly, germ-layer specific AC and VG CRMs have a much greater effect size (1.72 for AC CRMs and 1.87 for VG CRMs) when compared to ubiquitous CRMs (0.16) upon HDAC inhibition (Figure 3.7F). This indicates that AC CRMs acquire a drastic increase of pan-H3Kac signals in VG but not in AC in response to the loss of HDAC activity; likewise, VG CRMs acquire a drastic increase of pan-H3Kac signals in AC but not in VG upon the loss of HDAC activity. A similar trend is also observed on active CRMs (Cluster III, only H3K27ac) emphasizing germ layer-specific functions of these CRMs (Figure 3.8I, 3.8J). Taken together, these data demonstrate that

Hdac1 maintains differential H3 acetylation states between germ layers through its catalytic activity.

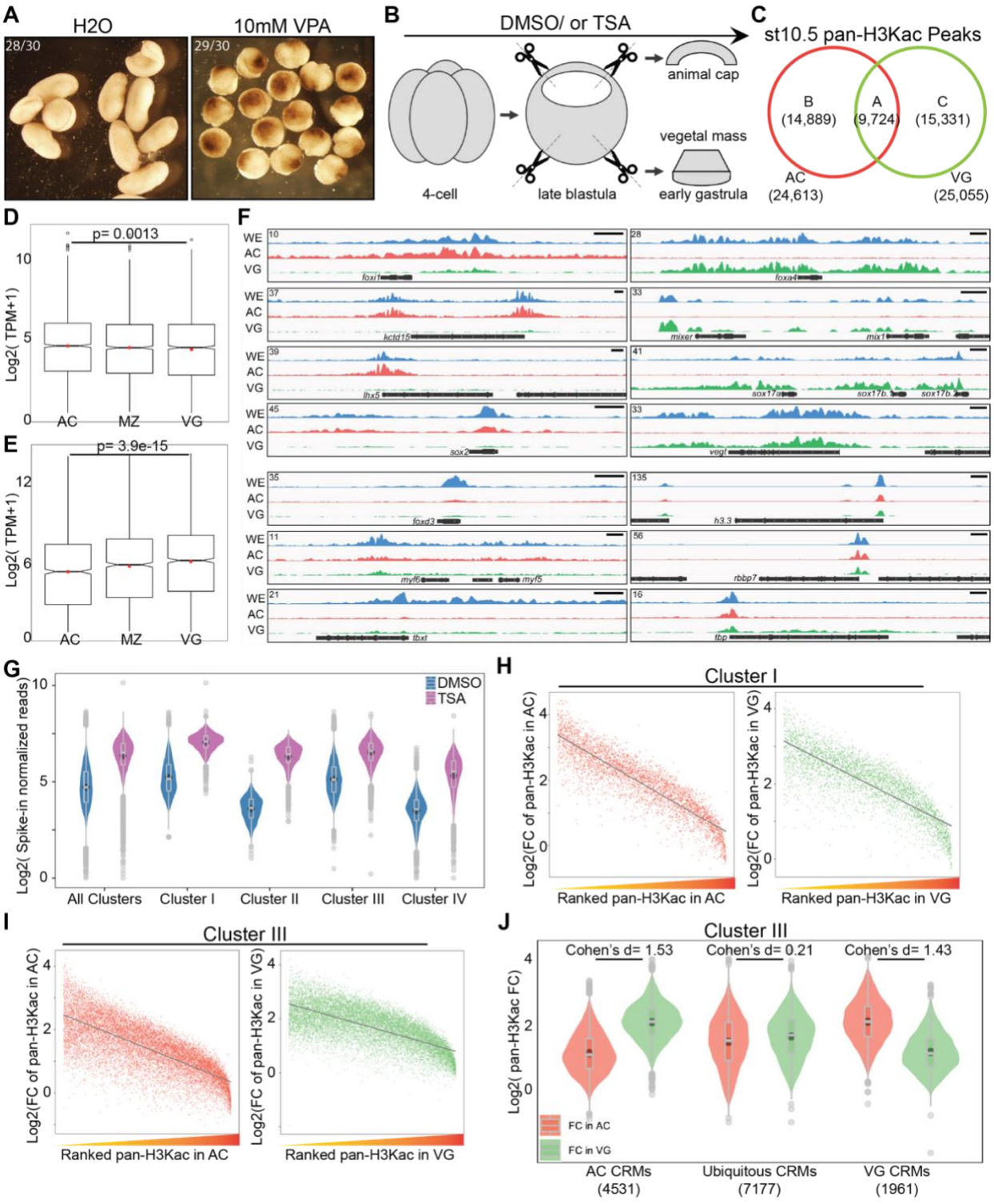

Figure 3.8. Germ-layer specific H3 acetylomes require HDAC activity

(**A**) Embryos treated with 10mM valproic acid (VPA) display gastrulation defects 24 hours post-fertilization. (**B**) A schematic diagram of dissection experiments to separate animal cap tissues (AC) and vegetal mass tissues (VG) for RNA-seq or ChIP-seq. Late blastula stage explants were cultured ~1- 1.5hrs to the early gastrula stage before being harvested for analyses. (**C**) Venn diagram comparing pan-H3Kac peaks in AC and VG. (**D, E**) Transcripts per million (TPM) expression of genes associated with (**D**) AC specific pan-H3Kac and (**E**) VG specific pan-H3Kac peaks across three germ layers. *p*-values are calculated by Student's *t*-test. (**F**) Genome browser visualization of pan-H3Kac signals at known ectodermal (*foxi1*, *kctd15*, *lhx5*, *sox2*), mesodermal (*foxd3*, *myf5*, *myf6*, *tbxt*), endodermal (*foxa4*, *mixer*, *mix1*, *sox17a*, *sox17b.1*, *sox17b.2*, *vegt*) and "housekeeping" (*h3.3*, *rbbp7*, *tbp*) genes in whole embryos, AC, and VG. Y-axis values represent scaled track height. Black bars denote an interval of 2kb. (**G**) Spike-in normalized pan-H3Kac signals across Hdac1 CRM clusters (clusters in Figure 3.5C) in DMSO- or TSA-treated samples. Signals of pan-H3Kac are combined from experiments done in AC and VG. (**H, I**) Scatter plots representing a negative correlation between fold changes of pan-H3Kac signals and ranked endogenous pan-H3Kac signals in AC or VG at both (**H**) Cluster I and (**I**) Cluster III Hdac1 CRM clusters (clusters in Figure 3.5C). Grey lines depict linear regression curves generated from a linear model. Both x- and y-axes' values are log2 transformed. (**J**) Fold changes (FC) of pan-H3Kac signals at Cluster III of Hdac1 CRM clusters categorized spatially (described in Methods).

HDAC activity modulates developmental genes between germ layers

Histone deacetylation is generally associated with transcriptional repression; hence, HDACs are considered as transcriptional corepressors. To determine how Hdac1 CRM clusters I-III (Figure 3.5C) influence the activities of their corresponding genes, each CRM within a cluster was assigned to the nearest gene located within 10 kb and then subdivided these genes into 3 classes (Figure 3.9A). Class 1 denotes 3,104 genes that have CRMs with mixed marks of H3K27ac and/or H3K27me3 suggesting that these genes are differentially expressed in different germ layers. Class 2 denotes 629 genes whose CRMs are marked with only H3K27me3 indicating that these genes may be repressed. Class 3 denotes 5,913 genes whose CRMs are marked with only H3K27ac inferring that these genes are active in various germ layers. We first investigated the temporal expression pattern (Owens et al., 2016) of each gene class. To exclude the interference posed by residual maternal transcripts at these early stages, we examined the expression patterns of exclusively zygotic genes in each class and found that Class 1 and Class 3 genes are gradually activated after ZGA while Class 2 genes remain mostly silent even at late gastrula (Figure 3.9B). To assess whether Class 2 genes remain inactive throughout the development, we extended our temporal expression analysis through to tailbud stage 26 (23 hpf). Class 1 and Class 3 genes are continuously active after ZGA, whereas Class 2 genes are gradually activated from early neurula and onward but not significantly before (Figure 3.9C, Figure 3.10A). These data suggest that Hdac1 regulates both transcriptionally active and silent genes at gastrulation.

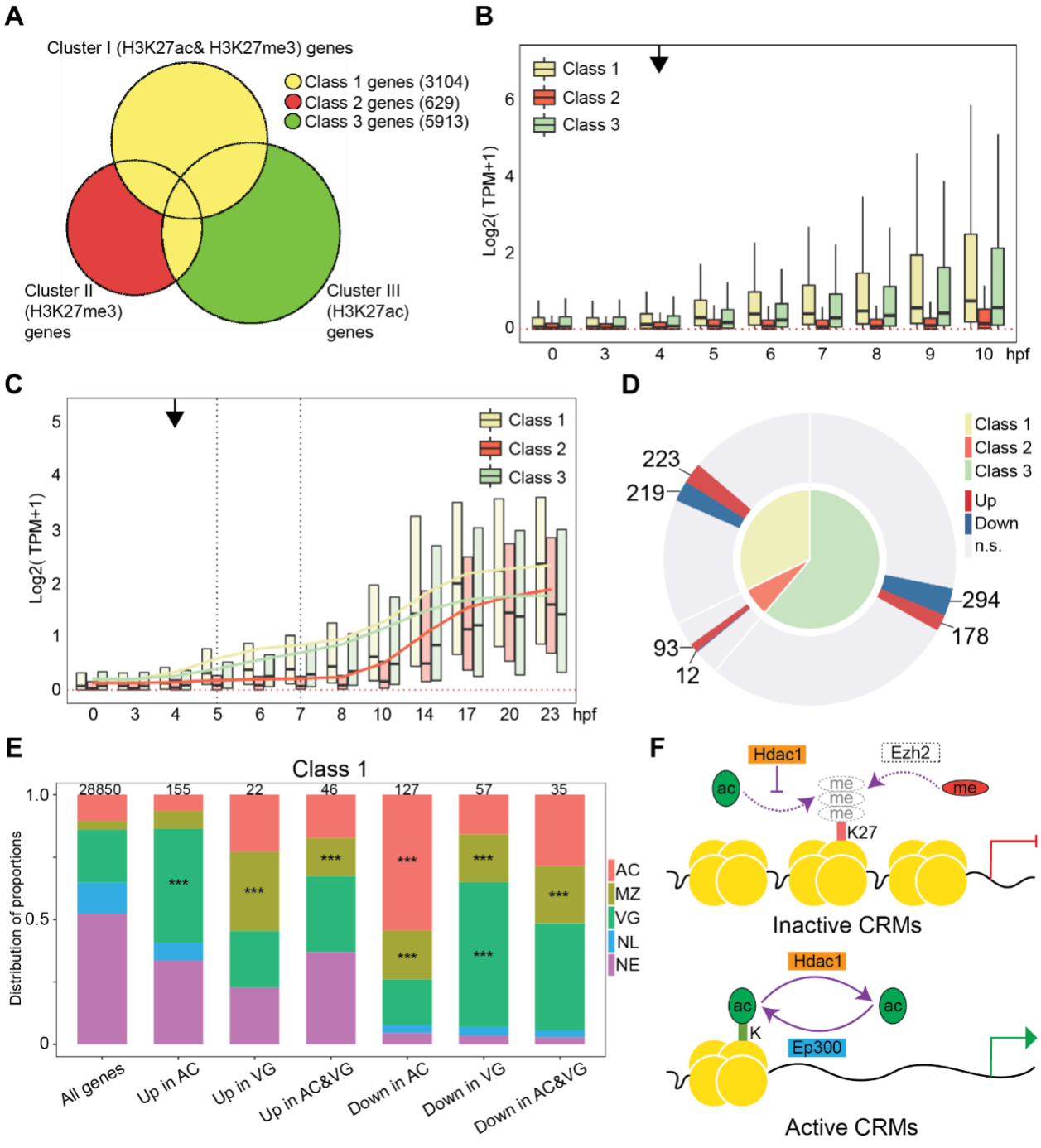

Figure 3.9 Hdac1 regulates germ layer transcriptomes both in time and space

(**A**) Venn diagram comparing genes associated with Hdac1 CRM clusters (Clusters I, II, and III in Figure 3.5C). Class 1 are genes closest to Cluster I peaks plus overlapped genes in both Class 2 and 3; Class 2 are unique genes closest to Cluster II peaks; Class 3 are unique genes closest to

Cluster III peaks. (**B**) Time-course TPM expression of zygotic genes in Classes 1, 2, and 3 up to 10 hours post-fertilization (hpf, late st12.5). Red dotted line denotes zero. Black arrow denotes the onset of ZGA. (**C**) Time-course TPM expression of zygotic genes in Classes 1, 2, and 3 up to 23 hpf (tailbud st26). Black vertical dotted lines denote the time window when Hdac1 binding is examined. Trend lines for each class are generated by connecting mean TPM values at each time point. (**D**) Nested pie chart illustrating the proportions of genes affected by TSA in each class. n.s.: not significant. (**E**) Spatial expression pattern of Class 1 genes affected by TSA in proportions. The total number of genes in each category is listed at the top of each bar. Only *** denoting $p < 0.001$ (Fisher's exact test) is shown. AC: animal cap, presumptive ectoderm; MZ: marginal zone, presumptive mesoderm; VG: vegetal mass, presumptive endoderm; NL: non-localized genes; NE: non-expressed genes. Parameters of spatial categories are detailed in Supplemental Methods. (**F**) Model of Hdac1 functioning at both inactive and active CRMs: on inactive CRMs, H3K27 residue is maintained as unmodified by Hdac1, which may be subjected to H3K27me3-mediated suppression; on active CRMs, the state of histone acetylation is dynamically modulated by Ep300-and-Hdac1-mediated acetylation-deacetylation cycles.

To understand how HDAC activity affects the expression of nearby genes, we performed RNA sequencing (RNA-seq) using early gastrula AC or VG explants treated with or without TSA (Figure 3.10B). Differential gene expression analyses revealed a greater number of differentially regulated genes in AC (1401 genes) than in VG (471 genes) suggesting that the ectodermal lineage is more susceptible to the loss of HDAC activity (Figure 3.10B). Many genes known to function during the germ layer development are affected when HDAC activity is inhibited. Genes affected by TSA treatment are affected similarly by VPA treatment indicating

that identified genes are not due to non-specific effects of TSA (Figure 3.10C). We found that HDACs have many overlapping and non-overlapping gene targets between AC and VG suggesting that HDACs confer both specific and ubiquitous activities in different germ layers (Figure 3.10D). Gene ontology analyses revealed that genes affected by HDAC inhibition primarily function in early embryonic development such as pattern specification, cell fate commitment, regionalization, tissue morphogenesis and many other processes. (Figure 3.10E, 3.10F).

Integrity of germ-layer specific transcriptomes is maintained by Hdac1 activity temporally and spatially

We attempt to correlate how the expression of different classes of Hdac1 bound genes (Classes 1-3 in Figure 3.9A) is affected by HDAC inhibition. Differentially regulated genes in Class 2 (whose CRMs are only marked by H3K27me3) are primarily (>85%) activated when HDAC activity is inhibited (Figure 3.9D). Given that Class 2 genes are not transcribed until after gastrulation normally (Figure 3.9C), premature activation of these genes at early gastrula upon HDAC inhibition supports the idea that Hdac1 suppresses the transcription of these genes temporally. This is consistent with observed histone hyperacetylation at surrounding repressive CRMs upon HDAC inhibition (Figure 3.9E). Furthermore, we observed that differentially regulated genes in Class 1 (whose CRMs are marked by a mixture of H3K27ac and/or H3K27me3) and Class 3 (whose CRMs are only marked by H3K27ac) can either be up- or down-regulated upon HDAC inhibition (Figure 3.9D). We speculate that HDAC activity is spatially employed by genes in Class 1 and 3.

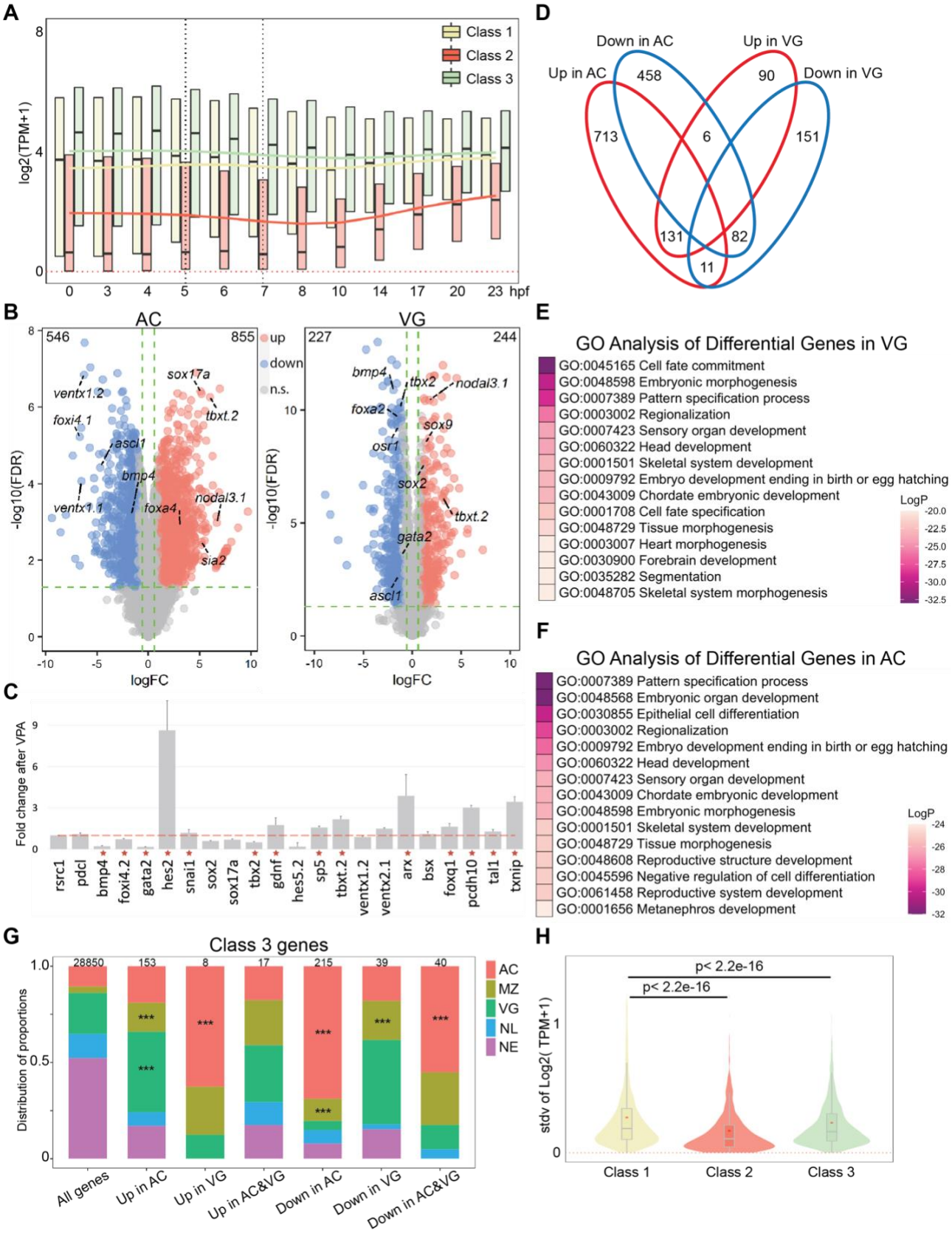
A
Class 1
Class 2
Class 3
log2(TPM+1)
hpf
B
AC
VG
up
down
n.s.
-log10(FDR)
logFC
C
Fold change after VPA
rsrc1
pdcl
bmp4
foxi4.2
gata2
hes2
snai1
sox2
sox17a
tbx2
gdnf
hes5.2
sp5
tbxt.2
ventx1.2
ventx2.1
arx
bsx
foxq1
pcdh10
tal1
txnip
D
Down in AC
Up in VG
Up in AC
Down in VG
458
90
713
6
151
131
82
11
E
GO Analysis of Differential Genes in VG
GO:0045165 Cell fate commitment
GO:0048598 Embryonic morphogenesis
GO:0007389 Pattern specification process
GO:0003002 Regionalization
GO:0007423 Sensory organ development
GO:0060322 Head development
GO:0001501 Skeletal system development
GO:0009792 Embryo development ending in birth or egg hatching
GO:0043009 Chordate embryonic development
GO:0001708 Cell fate specification
GO:0048729 Tissue morphogenesis
GO:0003007 Heart morphogenesis
GO:0030900 Forebrain development
GO:0035282 Segmentation
GO:0048705 Skeletal system morphogenesis
LogP
F
GO Analysis of Differential Genes in AC
GO:0007389 Pattern specification process
GO:0048568 Embryonic organ development
GO:0030855 Epithelial cell differentiation
GO:0003002 Regionalization
GO:0009792 Embryo development ending in birth or egg hatching
GO:0060322 Head development
GO:0007423 Sensory organ development
GO:0043009 Chordate embryonic development
GO:0048598 Embryonic morphogenesis
GO:0001501 Skeletal system development
GO:0048729 Tissue morphogenesis
GO:0048608 Reproductive structure development
GO:0045596 Negative regulation of cell differentiation
GO:0061458 Reproductive system development
GO:0001656 Metanephros development
LogP
G
Class 3 genes
Distribution of proportions
AC
MZ
VG
NL
NE
All genes
Up in AC
Up in VG
Up in AC&VG
Down in AC
Down in VG
Down in AC&VG
H
p< 2.2e-16
p< 2.2e-16
stdv of Log2(TPM+1)
Class 1
Class 2
Class 3

Figure 3.10 The integrity of germ layer transcriptomes requires HDAC activity

(**A**) Time-course TPM expression of all genes in Classes 1, 2, and 3 up to 23 hpf (hours post fertilization, tailbud st26). Red dotted line denotes zero. Black dotted vertical lines denote the time window when Hdac1 binding is examined. Trend lines for each class are generated by connecting the mean TPM values at each time point. (**B**) Differentially expressed genes after TSA treatment in either animal cap cells (AC) or vegetal mass cells (VG). The total number of differentially expressed genes in each category is listed on the top two corners. X-axis represents log2 transformed gene expression fold changes after TSA treatment, y-axis represents log10 transformed false discovery rate. n.s.: not significant. (**C**) RT-qPCR analysis examining the effect of VPA on TSA-responsive genes in whole embryos. *rsrc1* and *pdcl* are control genes; *bmp4*, *snai1*, *foxi4.2*, *gata2*, *hes2*, *sox2*, *sox17a*, and *tbx2* are Class 1 genes; *gdnf*, *hes5.2*, *sp5*, *txbt.2*, *ventx1.2*, and *ventx2.1* are Class 2 genes; *arx*, *bsx*, *foxq1*, *pcdh10*, *tal1,* and *txnip* are Class 3 genes. Red stars denote genes regulated in the same direction by both VPA and TSA treatments. The error bars represent the variation from 2 technical replicates. (**D**) Venn diagram comparing the genes differentially regulated by HDAC activity in AC and VG germ layers. (**E, F**) Gene ontology enrichment analyses of genes in (**E**) AC (963 genes) or (**F**) VG (334 genes) that are differentially expressed upon HDAC inhibition. The numbers of genes used for gene ontology enrichment analyses are smaller because the analyses only included genes with matched gene synonym to *Homo sapiens*. (**G**) Spatial expression patterns of Class 3 genes affected by TSA in proportions. The total number of genes in each category is listed at the top of each bar. Only *** denoting $p < 0.001$ (Fisher's exact test) is shown. AC: animal cap, presumptive ectoderm; MZ: marginal zone, presumptive mesoderm; VG: vegetal mass, presumptive endoderm; NL: non-localized genes; NE: non-expressed genes. Parameters of

spatial categories are detailed in Supplemental Methods. (**H**) The standard deviation of TPM expression between five germ layer explants (Blitz et al. 2017) for Class 1, 2, and 3 genes. *p*-values are calculated using Student's *t*-test.

To gain insights into the mechanism of spatial gene regulation by Hdac1, we analyzed the spatial expression patterns of all genes at early gastrula (Blitz et al., 2017) and subdivided genes into 5 categories (AC, MZ, and VG: genes enriched in ectoderm, mesoderm, and endoderm lineages, respectively. NL: uniformly expressed genes, and NE: lowly or not expressed genes). We then examined the endogenous spatial expression patterns of Class 1 genes affected by HDAC inhibition. Compared to spatial transcriptomic profiles of wide-type embryos, genes normally expressed in endoderm are significantly up-regulated in AC, and genes normally expressed in mesoderm are significantly up-regulated in VG upon HDAC inhibition (Figure 3.9E). This finding reveals that Hdac1 prevents misactivation of genes in respective germ layers. We highlight that misactivated genes are highly correlated with histone hyperacetylation of surrounding CRMs. For example, VG CRMs are hyperacetylated in AC upon HDAC inhibition (VG CRMs in Figure 3.7F), which in turn leads to the misactivation of endodermal genes in AC. Surprisingly, we found that genes normally expressed in ectoderm and mesoderm are significantly down-regulated in AC, and genes normally expressed in mesoderm and endoderm are significantly down-regulated in VG upon HDAC inhibition (Figure 3.9E). Contrary to the repressive role of Hdac1, this finding suggests that Hdac1 positively influences the transcription of active genes in each germ layer. We propose that the state of histone acetylation on CRMs associated with active genes is dynamic, and disruption to such equilibrium impairs normal transcription activities. Based on the quantitative analysis of pan-H3Kac profiles upon TSA

treatment, we found that Hdac1 bound CRMs gain increased levels of pan-H3Kac globally including on already acetylated CRMs (Figure 3.8G, 3.7F). This TSA-induced excessive histone acetylation on active CRMs may contribute to the attenuated expression of associated active genes in respective germ layers. A similar trend is observed among TSA-responsive Class 3 genes (Figure 3.10G). These data demonstrate a spatial regulation where Hdac1 not only prevents aberrant activation of silent genes, but also sustains proper gene expression levels in each germ layer.

Since both Class 1 (whose CRMs are marked by a mixture of H3K27ac and/or H3K27me3) and Class 3 (whose CRMs are only marked by H3K27ac) genes show germ-layer specific pan-H3Kac signals and transcriptomic profiles, we sought to examine the differences between these two classes. Class 1 genes display a higher variability in expression levels between different germ layers compared to Class 3 genes (Figure 3.10H). This indicates that the expression of genes undergoing active H3K27me3 suppression is more spatially restricted; thus, their functions may be more intimately associated with germ layer determination. Altogether, these results show that Hdac1 maintains the integrity of germ layer genes both in time and space during gastrulation.

Discussion

Here we defined a critical role for Hdac1 during early *Xenopus* embryogenesis. Our findings demonstrate a close link between homeostasis of histone acetylation and transcriptional activities in developing embryos. Progressive binding of Hdac1 to the genome shapes the zygotic histone acetylome, thereby reinforcing a proper germ-layer specific transcriptome, both in time and space. Thus, Hdac1 is an essential epigenetic regulator in the control of embryonic cell

identity and lineage. We propose that TFs inducing differentiation programs exploit the activity of Hdac1 to confine the expression of zygotic genes.

Gradual binding of Hdac1 coincides with ZGA

The genome-wide binding of Hdac1 begins at mid blastula and gradually accumulates at thousands of loci. Such progressive binding of Hdac1 around ZGA raises the question of whether Hdac1 recruitment requires zygotic factors during these stages. Surprisingly, ChIP-seq analyses of Hdac1 from α-amanitin-injected embryos show that zygotic transcription is dispensable for Hdac1 recruitment to the genome, at the very least, up to late blastula (Figure 3.3A, 3.4B). Recent work in yeast shows that active transcription is required to shape histone acetylation patterns largely due to a direct role of RNAPII in recruitment and activation of H4 HATs but not HDACs (Martin et al., 2021). Our result agrees with the notion that Hdac1 recruitment is not directed by on-going transcription. We, therefore, speculate that maternal factors instruct early Hdac1 recruitment. Our results reveal a positive binding correlation between Hdac1 and maternal Foxh1/Sox3, inferring the importance of these two maternal TFs in Hdac1 early recruitment. Though maternal TFs such as Foxh1, Vegt, and Otx1 are shown to bind the genome as early as the 32~64-cell stage (Charney et al., 2017; Paraiso et al., 2019), the genome-wide binding of Hdac1 begins at blastula and is not significantly widespread until early gastrula (Figure 3.1C). It is also known that Hdac1 containing complexes bind to pre-existing epigenetic marks such as DNA methylation and H3K4me3 (Wade et al., 1999; Lee et al., 2018). A previous study in *Xenopus* demonstrates that H3K4me3 is primarily instructed maternally until late gastrula (Hontelez et al., 2015). Therefore, maternally instructed histone modifications may also play a role in Hdac1 recruitment.

Hdac1 functions differently on active versus inactive CRMs

One perplexing finding is that Hdac1 occupies both active and repressive genomic loci (Figure 3.5A, 3.5B). Hdac1 binds to repressive genomic regions that are facultative but not constitutive heterochromatin. Though histone modifications underlying constitutive heterochromatin have been shown to regulate developmental genes (Riddle et al., 2011; Wang et al., 2018; Methot et al., 2021), the profile of H3K27me3 is largely different (~90% non-overlapping peaks) from profiles of H3K9me2, 3, and H4K20me3 in early *Xenopus* development (data not shown, partially shown in van Kruijsbergen et al., 2017). These observations suggest that Hdac1-mediated suppression is largely dictated by developmental programs. In contrary to commonly accepted repressor function of Hdac1, the binding of Hdac1 at active genomic regions is surprising. However, our findings are consistent with previous studies in yeast and mammalian cell culture suggesting a dynamic equilibrium of histone acetylation at active loci (Kurdistani et al., 2002; Wang et al., 2002; Wang et al., 2009; Kidder et al., 2012). Hence, we hypothesize that Hdac1 has different functions on active versus inactive CRMs in early embryos.

To test the *in vivo* function of HDACs, we blocked their endogenous activity using an inhibitor and quantitatively examined the changes of general H3 acetylation (pan-H3Kac) upon HDAC inhibition. First, an increase of pan-H3Kac is observed across Hdac1-bound CRMs, including active CRMs, consistent with the canonical enzymatic activity of Hdac1 (Figure 3.8G). Second, repressive CRMs marked by H3K27me3 undergo drastic H3 hyperacetylation compared to active CRMs marked by H3K27ac suggesting a HDAC-activity dependent suppression of repressive CRMs (Figure 3.7E). Third, germ-layer specific pan-H3Kac profiles are disrupted indicating the importance of Hdac1 in defining spatial patterns of histone acetylation (Figure 3.7F, 3.8J). Based on these results, we propose a dual function model for Hdac1. On the one

hand, Hdac1 prevents histone acetylation at inactive CRMs, thereby preserving H3K27 as unacetylated (Figure 3.9F, inactive CRMs model). Interestingly, H3K27me3 is not always imposed on inactive CRMs. For instance, active CRMs (Cluster III, only H3K27ac) are spatially modified with pan-H3Kac but are not subjected to H3K27me3 (Figure 3.3D, 3.7D, 3.8I, 3.8J). This suggests that HDAC-mediated histone deacetylation and Polycomb-mediated histone methylation are not coupled at inactive CRMs. On the other hand, Hdac1 participates in dynamic histone acetylation-deacetylation cycles at active CRMs (Figure 3.9F, active CRMs model). Although we did not directly test the co-binding of HATs and HDACs, CRMs may be simultaneously bound since (1) the binding profiles of Ep300 and Hdac1 mostly overlap (Figure 3.6A), and (2) pan-H3Kac signals increase at all Hdac1 peaks including active CRMs, upon HDAC inhibition (Figure 3.8G). Presumably, active CRMs are maintained in a state of dynamic equilibrium. This model is in accordance with a previous study demonstrating that HATs and HDACs simultaneously participate in histone acetylation cycles, which initiate and reset chromatin between rounds of transcription (Wang et al., 2009).

Hdac1 safeguards misactivation of genes both in time and space

We attempted to correlate the activity of CRMs with the transcriptional activity of potential target genes. Genes associated with repressive CRMs (H3K27me3 only) are mainly inactive until neurula (Figure 3.9B, 3.9C). More than 85% of these genes are prematurely activated upon HDAC inhibition suggesting that Hdac1 maintains the state of histone hypoacetylation on repressive CRMs, thereby preventing premature expression of genes. Moreover, genes associated with active (H3K27ac only) and heterogeneous (both H3K27ac and H3K27me3) CRMs are misactivated in different germ layers when HDAC activity is blocked

(Figure 3.9E, 3.10G). This indicates that Hdac1 safeguards differential histone acetylation states in each germ layer (Figure 3.7F, 3.8J), resulting in a proper spatial transcriptome. We did not directly address whether hypoacetylated heterogeneous CRMs are subjected to H3K27me3. However, a previous study showed that H3K27me3 is spatially deposited in line with spatial patterns of gene expression at late gastrula (Akkers et al., 2009). We predict that heterogenous CRMs are differentially marked by opposing H3K27me3 or acetylation in different germ layers. In summary, Hdac1 preserves the histone hypoacetylation state of inactive CRMs resulting in gene suppression both in time and space, which follows the canonical transcriptional corepressor role of Hdac1.

Cyclical histone acetylation sustains germ layer gene transcription

Our study reveals an unexpected role for Hdac1 in sustaining active gene expression during germ layer formation. Within both ectoderm and endoderm, the expression of TSA down-regulated genes associated with either active (H3K27ac only) or heterogeneous (H3K27ac and H3K27me3) CRMs are enriched in their respective germ layers (Figure 3.9E, 3.10G). This suggests a paradoxical activator role for Hdac1, which is also reported in previous studies (Vidal and Gaber, 1991; Zupkovitz et al., 2006; Baltus et al., 2009; Hughes et al., 2014; Rao and LaBonne, 2018). We speculate that utilization of HDAC activities at active genomic loci is a general mechanism as seen in examples of *Xenopus* ectoderm and endoderm lineages which deploy distinct gene regulatory networks. Based on our findings, we propose that a dynamic equilibrium between acetylation and deacetylation is essential to sustain gene transcription. The function of HDACs on active genomic regions has been elucidated in several contexts. In yeast, cotranscriptional methylation (H3K36me3, H3K4me2) recruits HDAC containing complexes

(Rpd3S, Set3C) to suppress intragenic transcription and delay induction of genes that overlap non-coding RNAs (Carrozza et al., 2005; Keogh et al., 2005; Li et al., 2007; Kim and Buratowski, 2009; Kim et al., 2012; Heo at al., 2021). Genetic deletion of Set3C affects transcript levels only in altered growth conditions (Lenstra et al., 2011), consistent with the notion that cyclical histone acetylation act as a mechanism to regulate dynamics of RNA induction and fidelity of transcription. Studies in metazoans show a conserved mechanism. HDAC1 can be targeted by p300 to transcribing genes through a direct interaction (Simone et al., 2004). Simultaneous binding of both HATs and HDACs at active genomic regions is shown in T cells (Wang et al., 2009). Inhibition of both DNA methyltransferases and HDACs induces cryptic transcription in lung cancer cells (Brocks et al., 2017). Down-regulated genes upon HDAC inhibition exhibit high levels of cryptic transcripts during mouse cardiogenesis (Milstone et al., 2020). These findings suggest a role for HDAC activity in transcriptional fidelity.

Why does excessive histone acetylation due to disruption of cyclical histone acetylation lead to reduced transcription instead of elevated transcription? We observed that active (H3K27ac only) and heterogeneous (both H3K27ac and H3K27me3) CRMs are excessively acetylated (Figure 3.7E, 3.8G) following HDAC inhibition, which results in the reduced expression of CRM-associated genes within their respective germ layers (Figure 3.9E, 3.10G). This leads us to speculate that excessive histone acetylation interferes the activity of the transcriptional machinery. A recent study shows that excessive histone acetylation on chromatin induced by inhibition of HDAC1, 2, and 3 leads to increased aberrant contacts and reduced native contacts between super-enhancer loops (Gryder et al., 2019). This suggests that the excessive histone acetylation impairs active transcription by altering chromatin interactions. Alternatively, excessive histone acetylation can alter the binding of acetyl-histone readers. H4

polyacetylation induced by HDAC inhibition is shown to be preferentially bound by BRD proteins (such as BRD4), thereby sequestering these factors away from active genes (Slaughter et al., 2021). Therefore, HDACs safeguards the function of normal acetyl-histone readers. Further investigation of cyclical histone acetylation regulating developmental programs is needed in the context of germ layer specification.

Materials and Methods

Animal Model and Subject Details

Xenopus tropicalis embryos were obtained by *in vitro* fertilization according to Ogino et al. (2006) and staged according to Nieuwkoop and Faber (1994). All embryos were cultured in 1/9X Marc's modified Ringers (MMR) at 25 °C. For HDAC inhibition, 4-cell stage embryos were immersed in 1/9X MMR containing 1). 100nM TSA (Esmaeili et al., 2020) or DMSO; or 2). 10mM VPA (Rao and LaBonne. 2018) or H_2O. For α-amanitin injection, each 1-cell stage embryo was injected with 6pg of α-amanitin (Hontelez et al., 2015). For embryo dissection, embryos were dissected at the late blastula stage (6 hpf), and explants were cultured to the early gastrula (7 hpf). Animals were raised and maintained following the University of California, Irvine Institutional Animal Care Use Committee (IACUC). Animals used were raised in the laboratory and/or purchased from the National *Xenopus* Resource (RRID: SCR_013731).

Western Blotting

Embryos were homogenized in 1X RIPA (50mM Tris-HCl pH7.6, 1% NP40, 0.25% Na-Deoxy-cholate, 150mM NaCl, 1mM EDTA, 0.1% SDS, 0.5mM DTT) with protease inhibitors (Roche cOmplete) and centrifuged twice at 14,000 rpm. The supernatant was then subjected to western

blotting using anti-HDAC1 (Cell Signaling, 34589S), anti-HDAC2 (Genetex, GTX109642), and anti-Tubulin (Sigma, T5168). For histone modifications, acid-extracted histone lysate was prepared accordingly (Shechter et al., 2007) and subjected to western blotting using anti-H3K9ac (Cell Signaling, 9649), H3K14ac (Cell Signaling, 7627), H3K18ac (Cell Signaling, 13998), H3K27ac (Cell Signaling, 8173), H3K56ac (Cell Signaling, 4243), and H4K20ac (Active Motif, 61531).

Chromatin Immunoprecipitation (ChIP)

ChIP protocol was performed as described (Chiu et al., 2014). Antibodies used for ChIP were anti-HDAC1 (Cell Signaling, 34589S, 1:100), anti-H3K18ac (Cell Signaling, 13998, 1:100), and anti-Sox3(Jansen et al., 2021, 2ug). ChIP-seq libraries were constructed using the NEBNext Ultra II DNA Kit (NEB, E7645).

For sequential ChIP, the first round of ChIP was performed as described and eluted in 1X TE containing 1% SDS at 37 °C for 30 minutes. The eluate was diluted with 1X RIPA (without SDS) 10 times and subjected to the second round of ChIP as described (Desvoyes et al., 2018). Real-time quantitative PCR (RT-qPCR) was performed using Power SYBR Green PCR master mix (Roche) to quantify the DNA recovery compared to ChIP input DNA at one embryo equivalency (percent input). The error among technical replicates was calculated using the rule of error propagation. Primer sequence information is provided in Table S1.

For dissected pan-H3Kac ChIP, spike-in chromatin (Active Motif, 53083) was added to the chromatin of dissected tissues at a ratio of 1:35. Mixed chromatin was then subjected to ChIP with 5ug anti-panH3Kac (Active Motif, 39139) and 1ug of anti-H2Av (Active Motif, 61686) and followed as described.

All experiments were performed in two independent biological replicates unless noted. Sequencing was performed using the Illumina NovaSeq 6000 and 100bp single-end reads or 100bp paired-end reads were obtained. Computational analysis of ChIP-seq is described in the Supplemental Methods.

RNA-seq and Gene Expression Assays

Total RNA from dissected tissues was extracted using Trizol as described (Amin et al., 2014). mRNA was then isolated using NEBNext PolyA mRNA Magnetic Isolation Module (NEB E7490S). Sequencing libraries were prepared using NEBNext Ultra II RNA library prep kit (NEB E7770S) and sequenced by the Illumina NovaSeq 6000 with 100bp paired end reads. All experiments were performed in two independent biological replicates. Computational analysis of RNA-seq is described in the Supplemental Methods.

For RT-qPCR, reverse-transcription assay was performed using Maxima reverse transcriptase (Thermo Fisher EP0741). RT-qPCR was performed using LightCycler 480 SYBR Green I master mix (Roche). For quantification of gene expression, the 2^ -ΔΔCt method was used. The expression level of *drosha* or *rsrc1* was used for normalization. The error among technical replicates was calculated using the rule of error propagation. Primer sequence information is provided in Table S3.

Data Accessibility

Raw and processed RNA-seq and ChIP-seq datasets generated for this study are available at NCBI Gene Expression Omnibus using the accession GSE198378.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Drs. Kyoko Yokomori (University of California, Irvine), Yongsheng Shi (University of California, Irvine), and current Cho lab members for insightful comments on this study. We thank the Genomic High Throughput Facility at the University of California, Irvine for sequencing services. We also thank the Research Cyberinfrastructure Center at the University of California, Irvine for the ongoing support of High Performance Community Computing Clusters. This work is supported by NIH R01GM126395, R35GM139617 and NSF 1755214 to K.W.Y.C..

Author contributions

J.J.Z., J.S.C., H.H. and W.W. did wet bench experiments; J.J.Z. performed bioinformatic analyses; J.J.Z. and K.W.Y.C. contributed to the design of the study; J.J.Z., I.L.B., K.W.Y.C. wrote the manuscript.

Supplemental Methods

ChIP-seq Analysis

All sequencing data were aligned to *Xenopus tropicalis* v10.0 genome (http://www.xenbase.org/, RRID:SCR_003280) using Bowtie2 v2.4.4 (Langmead and Salzberg, 2012). PCR duplicates were removed using Samtools v1.10 (Li et al., 2009). ChIP-seq signals were visualized using IGV v2.11.3 (Robinson et al., 2011) after concatenating two biological replicates when available. Irreproducibility discovery rate (IDR) analysis (Li et al., 2011) was used to identify high-confidence peaks called by Macs2 v2.7.1 (Zhang et al., 2008) against the stage-matched input (Charney et al., 2017) between two biological replicates according to ENCODE3 ChIP-seq pipelines (IDR threshold of 0.05) (https://docs.google.com/document/d/1lG_Rd7fnYgRpSIqrIfuVlAz2dW1VaSQThzk836Db99c/edit).

For dissected pan-H3Kac ChIP, **all second replicates** were downsampled to 25% by random sampling to a comparable sequencing depth for later analysis. *Drosophila* H2Av peaks are generated from overlapped peaks with two published S2 cell samples (Weber et al., 2014; Tettey et al., 2019). Normalization factors were then calculated based on reads that mapped to *Drosophila* H2Av peaks for each ChIP-seq sample as used in Egan et al., 2016. When replicates or dissected samples are merged for analyses, normalization factors are calculated as the sum of $Ratio_{frog}$ *$Ratio_{fly}$. Detailed normalization factors used are listed in Table S2.

RNA-seq Analysis

All sequencing samples were aligned using STAR v2.7.3a (Dobin et al., 2013) to *Xenopus tropicalis* genome v10.0 (http://www.xenbase.org/, RRID:SCR_003280) to obtain raw read

counts. RSEM v1.3.3 (Li and Dewey, 2011) was used to calculate expression values in transcripts per million (TPM). Differentially expressed genes were identified using edgeR v3.36.0 (Robinson et al., 2010) with parameters, greater than 2-fold change and less than 0.05 false discovery rate (FDR, also known as the adjusted p-value), in R v4.1.2 (R Core Team, 2021). Metascape (Zhou et al., 2019) was used to perform gene ontology enrichment analyses with default parameters (min overlap= 3, p-value cutoff= 0.01, and min enrichment= 1.5).

Additional Bioinformatics and Statistical Analyses

Samtools v1.10 (Li et al., 2009) was used to convert between SAM and BAM files. DeepTools v3.5.0 (Ramírez et al., 2014) was used to generate: (1) ChIP-seq signal track (bigwig files) normalized by reads per genomic content (-RPGC) at the bin size of 1bp; (2) heatmaps around peak summits normalized by Bins Per Million mapped reads (-BPM) at the bin size of 50bps; (3) signal profile along the gene bodies normalized by -BPM at the bin size of 50bps; (4) Pearson correlation between ChIP-seq samples at peaks. Homer v4.10 (Heinz et al., 2010) was used to annotate genomic features of ChIP peaks. Bedtools v2.29.2 (Quinlan and Hall, 2010) was used to determine peak overlaps among ChIP-seq peaks and obtain counts of reads at ChIP-seq peaks. CentriMo (Bailey and Machanick, 2012) was used to perform local motif enrichment analysis. Welch Two Sample t-test in R v4.1.2 was used to determine the statistical significance between groups. Cohen's d (effect size) was calculated using lsr v0.5.2 package in R v4.1.2 (R Core Team, 2021). Time-course gene expression was obtained from ribosomal RNA-depleted RNA-seq data (Owens et al., 2016). The expression in TPM was calculated as outlined above. Spatial gene expression at early gastrula was obtained from RNA-seq of 5 dissected tissues (Blitz et al., 2017) consisting of the animal cap (ectoderm), the dorsal marginal zone (dorsal mesoderm), the

lateral marginal zone (lateral mesoderm), the ventral marginal zone (ventral mesoderm), and the vegetal mass (endoderm). The expression in TPM was obtained as outlined above. Fisher's Exact test (alternative = "greater") in R v4.1.2 (R Core Team, 2021) was used to determine the significance of the proportional enrichment between groups.

Categorical Analyses

Spatial categorization of CRMs is defined as below: AC CRMs represent CRMs whose pan-H3Kac signals in AC is 2-fold higher than in VG; VG CRMs represent CRMs whose pan-H3Kac signals in VG is 2-fold higher than in AC; ubiquitous CRMs represents the remaining CRMs whose pan-H3Kac signals do not exceed 2-fold enrichment in either of the two examined germ layers. For temporal gene expression analysis, (strictly) zygotic genes are determined by removing genes whose expression levels are greater than 1 TPM during the first 2 hours post-fertilization. Spatial categorization of genes at early gastrula stage: the average TPM between 3 dissected mesoderm tissues (dorsal, marginal, and lateral marginal zones) was used to represent the expression of mesoderm. Genes with the expression in any dissected tissue less than 1 TPM were considered not expressed. Genes with the coefficient of variance of TPM less than 0.1 (10%) were considered evenly expressed. The remaining genes with localized expression were assigned to a germ layer based on the maximum TPM.

Supplemental Table 3.1: Primer Sequences for sequential ChIP-qPCR

Primer targets	Forward Sequence	Reverse Sequence
bmp4 intron	TGCTAATAAGCGTGCCTTTG	AAGGAGGACCAGCCTATTCA
foxa4 intron	ATCAGCACTTTTGTGTCCTT	ACTGTTTGGGGTTGGAATG
hoxb3 intron	GTAGATCTGTCAGGGAGAAA	GCCATTGATTGAGATCCGTA
klf11 exon	TCCCATTTCTGTTCCTGTGT	GGAGAATAGGCTTTACTGTATTA
set 3'-down	TGCAGCCCTGTTGTATAAAG	CAGATTATGATTGGCTGCCG
snai1 intron	CGCACCATTAGCAATTCATGA	GGAAGGGCAATCTAAACAAG
sox17b.2 5'-up	AGCACCTAGGATAGTGTCAG	TTGAGAGGCTGTTATAGGCA
tbxt.2 intron	ATTGGATTGGAACACTGGGA	AAGGTCTATTGACTCTGTTTCTC

Supplemental Table 3.2: Normalization Factors of ChIP spike-in strategy. Table listing all spike-in normalization factors. Reads mapped to *Xenopus tropicalis* (frog) for all second replicate samples are already **downsampled to 25%**. Red numbers indicate the base ratio line for each column. The ratios for reads mapped to H2Av peaks boxed in yellow are used for all later calculations; reads mapped to *Drosophila melanogaster* (fly) serves as a control to show the highly consistent ratios between whole-genome and H2Av peak mapped reads. The ratios for reads between replicate-merged samples are boxed in blue (corresponding to ratios used in Figure 4F, S4J), which is calculated as sum of $Ratio_{frog}$ *$Ratio_{fly}$ between two replicates. The ratios for reads between explant-merged samples are boxed in magenta (corresponding to ratios used in Figure S4G), which is calculated as sum of $Ratio_{frog}$ *$Ratio_{fly}$ between two replicates followed by two explants.

Sample	Reads mapped to frog	Ratio (frog)	Reads mapped to fly	Ratio (fly)	Reads mapped to H2Av	Ratio (H2Av)	Ratio of merged reps	Ratio of merged explants
AC_DMSO_rep1	43417738	1.17	2817756	4.09	615449	3.18	18.07	33.57
AC_TSA_rep1	127763576	3.43	1175034	1.71	229326	1.18	12.49	24.66
VG_DMSO_rep1	41215108	1.11	2749022	3.99	657611	3.40	15.49	
VG_TSA_rep1	124387262	3.34	688352	1	193553	1	12.16	
AC_DMSO_rep2	47280904	1.27	10230400	14.86	2187348	11.30		
AC_TSA_rep2	86344294	2.32	3320496	4.82	703883	3.64		
VG_DMSO_rep2	37237950	1	9530640	13.85	2268171	11.72		
VG_TSA_rep2	101907192	2.74	2557466	3.72	622410	3.22		

Supplemental Table 3.3: Primer Sequences for RT-qPCR

Primer target genes	Forward Sequence	Reverse Sequence
arx	GTACCGCACCACCTTTACCA	GCATCGCTAGCTCCTCTCTG
bmp4	ATGGAGTACCCGGAGAGACC	TCCATTTTCTGCTGTGCTTG
bsx	CATCTGGGTTGCCAGTACCA	GAGTTGCGAGTCGGAGAACA
drosha	TTACAGACCGCTGTTTGCTG	CAATTCGAGAGGGAGTTTCG
foxa2	AGGACACGAAGCCACAGACT	CGAGGTGCTCATAGCAGACA
foxi1	TTCCACCCTCAGACAATGCC	AGCCACAGGTAGGGATTTGC
foxi4.2	ACACCTCTCCGGCTAATCCT	AAAGGACGGGGGCATGTAAG
foxq1	GCGACTTAACAGGGTCACCA	GGGAGAAGATCCTTGGCTGG
frzb	GGCTGTGAGCCCATTCTCAT	GGGGAAGTCTGGCATCGAAT
gata2	AGGGTGTTTTACAGCCAAGC	AGTCTTCCCGCTCTCCAAC
gdnf	AGGTGAAGCAGGCTTGTTGT	TGGCCATCCAAGCAAGTCAA
gsc	CAGAGAAAGGGGACGAACAG	GATCCCACATCATGGCACT
hes2	ATCAGGAGGCCAGTGAGCTA	TCCAGCTTGGAGTATCGGGA
hes5.2	CCTAAGAGATGGCACCCAGC	TCTCCACAACTGGCTTTCGT
mixer	TTCCAAGAGGCCAGCTTAAA	GGGGAGACCATTTCTTGCTT
nodal (nodal4)	AATGCCTATCGGTGTGAAGG	AACTCATCCTGATGGGAACG
pcdh10	CAGAGCGCCTTACAAGCCT	GTTGGCACAACATTGCCTACA
pdcl	CAAAACCATCAAGTTCTCTGGA	CCTCATTTTGTTCGGTTTCAA
rsrc1	TCGAACTGATTGAAAGCGACT	AACAAATCGATGGTGCCTGT
sin3a	GGGGCTGGACTAGAACTAGGA	ATACGGCCACACCAAGTCAG
snai1	GCCAGCAAGAAACCCAACTA	TGTCCCAGACAAGAGGTGTG
sox2	TAGAGGGGGACAAACTGTGG	TTTGGCCTTCCTGCTGTTAC
sox17a	CTGGAAGGCTCTGACTCTGG	CTGGAAGGCTCTGACTCTGG
sp5	GCTTTCCTACAGGACAGGACCC	CCAGAAGAGCCAAGGGTGAG
tal1	GAGGTTGGGGACAGACATCG	CCCACTCAGCTCCACAGTTC
tbx2	GAGAATGTTTCCCCCATTCA	CCGCAGCCACTATATCCATT
tbxt	GCCCCCAGTTGTGTTTACAT	GCCTCCACCATTCATTTTGT
tbxt.2	TTACCCCAGCCCCTATACCC	TGGAACACTGGGAGACATGC
txnip	AAGCCTTCCATCTTGGGGTG	GCCTATGACAAGGGGCAAGT
ventx1.2	AGAAGCTGGCAACGTCTCTC	GCTGTGCTGGTCTTGTA
ventx2.1	TTGCTACACAGGGACACAGC	GCCTGAGTCAGTGCTAGTGC

CHAPTER 4

High-throughput Identification of Functional *cis*-Regulatory Modules in *Xenopus* Embryos using STARR-seq

Introduction

Cellular differentiation, beginning at fertilization, is programmed by a hard-wired genetic control system through the activity of a subset of transcription factors (TFs), which sit high in the regulatory hierarchy. These key TFs coordinate a series of gene regulatory cascades through combinatorial interactions with *cis*-regulatory modules (CRMs, also known as enhancers). These interactions are fundamental molecular languages that instruct cell fate specification and determination. Dissecting these interactions *in vivo* is challenging because it is difficult to clearly discern whether the genomic occupancy of TFs is solely based on the property of TFs alone or restricted due to pre-existing epigenetic modifications in complex tissues or differentiated cells. In *Xenopus*, germ layer specification (the delineation of ectoderm, mesoderm, endoderm) is the first cell lineage commitment event in development, which depends on unequally distributed maternal determinants present in the egg before fertilization (Heasman, 2006). Prior to zygotic gene activation, cleavage stage chromatin is free of permissive H3K4me3 marks, repressive H3K27me3 marks, and H3K4me1 modifications (Akkers et al., 2009; Gupta et al., 2014; van Heeringen et al., 2014). Therefore, this is the ideal stage to understand the *cis*-regulatory logic without discerning the pre-existing chromatin restrictions.

Genome-wide studies of *Xenopus* maternal TFs (e.g., Foxh1, Vegt, Otx1, Sox7) critical for endoderm formation show that thousands of genomic loci were co-bound by many of these TFs since the blastula stage (Charney et al., 2016; Paraiso et al., 2019). Subsequently, these co-

bound loci gain epigenetic modifications (H3K4me1, H3K27ac) that are well-known to mark enhancers (Creyghton et al., 2010). In addition, many other genomic loci also gain H3K4me1 and/or H3K27ac marks with little or no binding of these known maternal TFs. These observations raise the question of which genomic loci (CRMs) are actual functional CRMs used during the first stage of cell differentiation. Many developmental enhancers that have been described in the past utilized reporter assays that assess the ability of candidate CRMs to drive the expression of different reporter genes (Watabe et al., 1995; Osada et al., 2000; Jensen et al., 2022). Thousands of developmental enhancers were systematically characterized with the reporter system *in vivo* using transgenic flies (Kvon et al., 2014). Yet, no systematic and high-throughput analysis is done to address functional CRMs in early *Xenopus* embryo development. To address this question, we adopt a high-throughput approach, known as self-transcribing active regulatory region sequencing (STARR-seq), to identify functional CRMs *in vivo*. STARR-seq is a method that measures the level of transcripts that are linked to underlying CRM DNA sequences (Arnold et al., 2013). Specifically, CRMs are cloned individually downstream of a minimal promoter within a DNA plasmid; active CRMs will be transcribed as a part of transcripts in given cells. CRM transcripts themselves serve as unique barcodes and can be deconvoluted by RNA sequencing (RNA-seq).

Here, we developed a protocol to perform STARR-seq in *Xenopus* embryos. We show that CRMs identified from integrating ChIP-seq data sets of TFs and epigenetic signatures exhibit episomal reporter activities that are primarily dependent on TF motifs within. Our STARR-seq protocol effectively captures RNAs transcribed from injected reporter plasmids; however, this protocol still needs further optimizations in order to reveal the physiological functions of CRMs.

Result

CRMs identified from overlapping ChIP-seq signals show enhancer activity

In early *Xenopus* embryogenesis, zygotic genes are activated through precisely controlled developmental programs instructed by maternal TFs. Given that a TF motif is usually 6 base pairs in length on average, such a TF motif within the entire genome is highly frequent statistically. However, only a small subset of a particular TF motif is utilized by the corresponding TF, as shown in genome-wide maternal TF studies (Chiu et al., 2014; Charney et al., 2017; Paraiso et al., 2019). This raises the question of how maternal TFs select particular CRMs to be functional. This question cannot be easily addressed by studying the behaviors of TFs alone, as numerous genomic metadata analyses suggest that the specificity, robustness, and adaptivity of transcriptome rely on distinct combinations of TFs occupying specific genomic loci in a context-specific manner.

To facilitate our findings on maternal TF behaviors, we preselected candidate functional CRMs individually, mutated TF binding motifs in the CRM, and examined the changes in reporter activities between reporter genes with either wild type or mutated CRMs. We hypothesize that preselected CRMs exhibit germ-layer specific activities and their activities depend on the presence of TF motifs. We selected 38 candidate CRMs that harbor maternal endoderm TF motifs, well-known enhancer histone modifications, and Ep300 binding (Hontelez et al., 2015; Chiu et al., 2014; Charney et al., 2017; Paraiso et al., 2019). These CRMs are cloned individually into a luciferase reporter construct whose expression is driven by the *goosecoid* minimal promoter (Watabe et al., 1995). Out of 38 CRMs we tested, most showed a preferential activation in endoderm lineage with variable levels (Figure 4.1A). In addition, we mutated the

TF motifs within the CRMs and showed that the activity of a CRM is mainly dependent on the presence of TF motifs (Figure 4.1B-D). These results demonstrated that CRMs predicted from integrating different ChIP-seq data are functionally active in developing embryos.

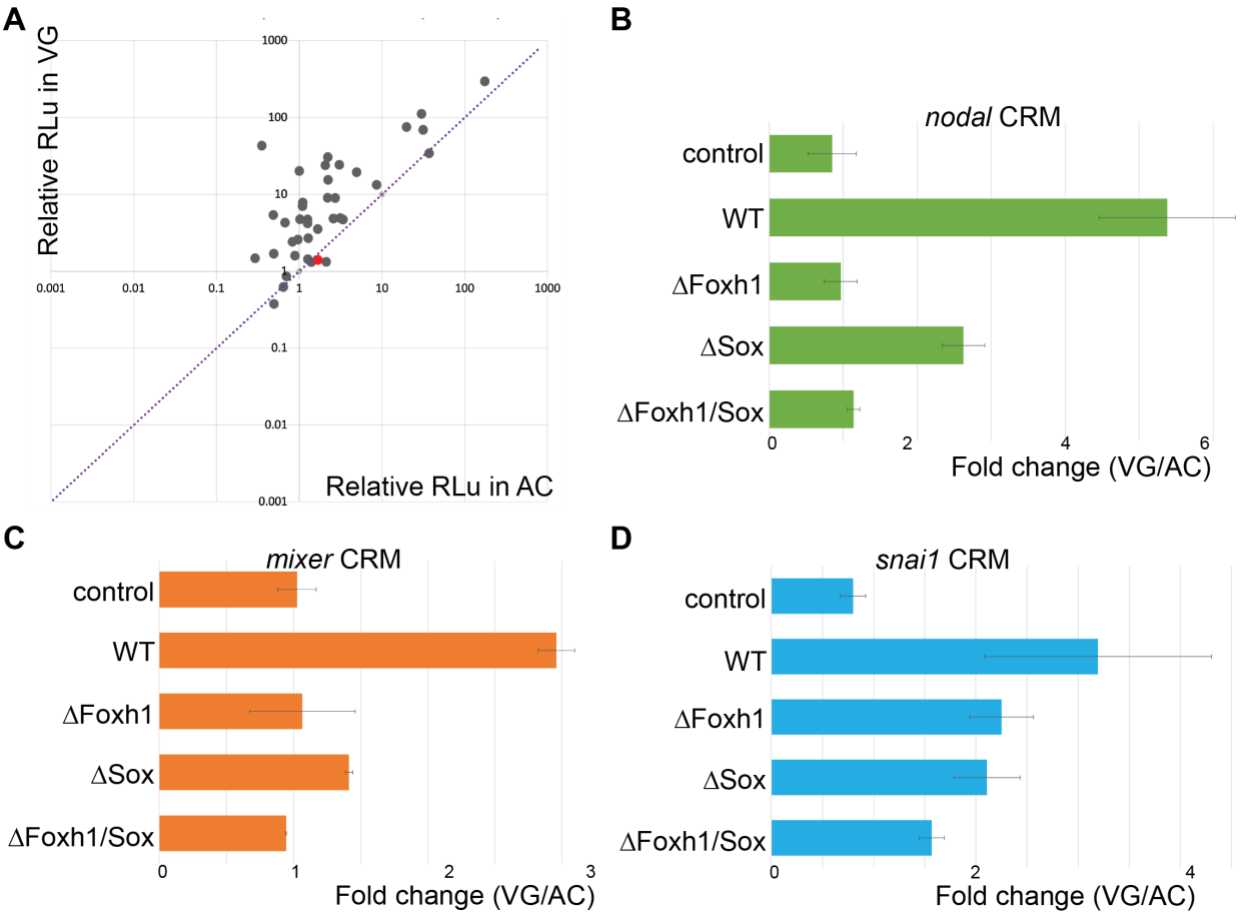

Figure 4.1: Candidate CRMs identified exhibit reporter activities.

(**A**) Relative luciferase activities (RLu) of 38 tested CRMs (*nodalX2, mix1X4, snai1, frzb, nodal1, fgf20X2, nodal6X4, sia1, gsc, mixerX2, sebox, sox17b.1, fam184a, fst, mesbp, stox1, tcf7, grhl3, bmp7.2, hhex, admp, map7d3, pcdh8.2X3, pdk4, serpinf2, sfrp2, slc12a3.1*). The red dot denotes empty vector used as a negative control; the purple line denotes where the reporter activity is the same between animal cap (AC, presumptive ectoderm) and vegetal mass (VG, presumptive endoderm). (**B-D**) Fold changes of RLu in VG versus AC of CRMs associated with

gene (**B**) *nodal*, (**C**) *mixer*, and (**D**) *snai1* when TF motifs are mutated. Error bars denote standard error from 2 different biological replicates.

STARR-seq primarily captures transcripts derived from candidate CRMs

Though the luciferase assay system is the gold standard method to assess the functionality of CRMs, it is limited by its low-throughput. We sought to identify functional CRMs in a high-throughput fashion using the recently developed STARR-seq (Arnold et al., 2013; Barakat et al., 2018). Applying STARR-seq directly in *Xenopus* embryos is difficult. First, *Xenopus tropicalis* has a genome size of 1.7 billion base pairs, which is ten times larger than the genome size of *Drosophila*. A tremendous high sample size will be required if the entire genome were to be assessed. Second, though *Xenopus* embryos are large, and microinjection of plasmid DNA is a routine technique, the amount of plasmid that an embryo can tolerate is 20pg (Watabe et al., 1995), which is much less than the amount used in cell culture transfection. To limit the number of testing CRMs, we decided to assess the regulatory function of regions bound by Ep300. Ep300 is a well-characterized transcription activator that correlates to active CRMs. During early *Xenopus* embryogenesis, Ep300 binds to many genomic loci at the blastula stage (stage 9) and onward (Hontelez et al., 2015). These observations infer that Ep300 bound regions in early development represent potentially active CRMs for zygotic gene activation. To date, functional validation of these CRMs has never been performed. Here, we propose to perform a large-scale functional validation of Ep300 bound regions using STARR-seq (Figure 4.2A).

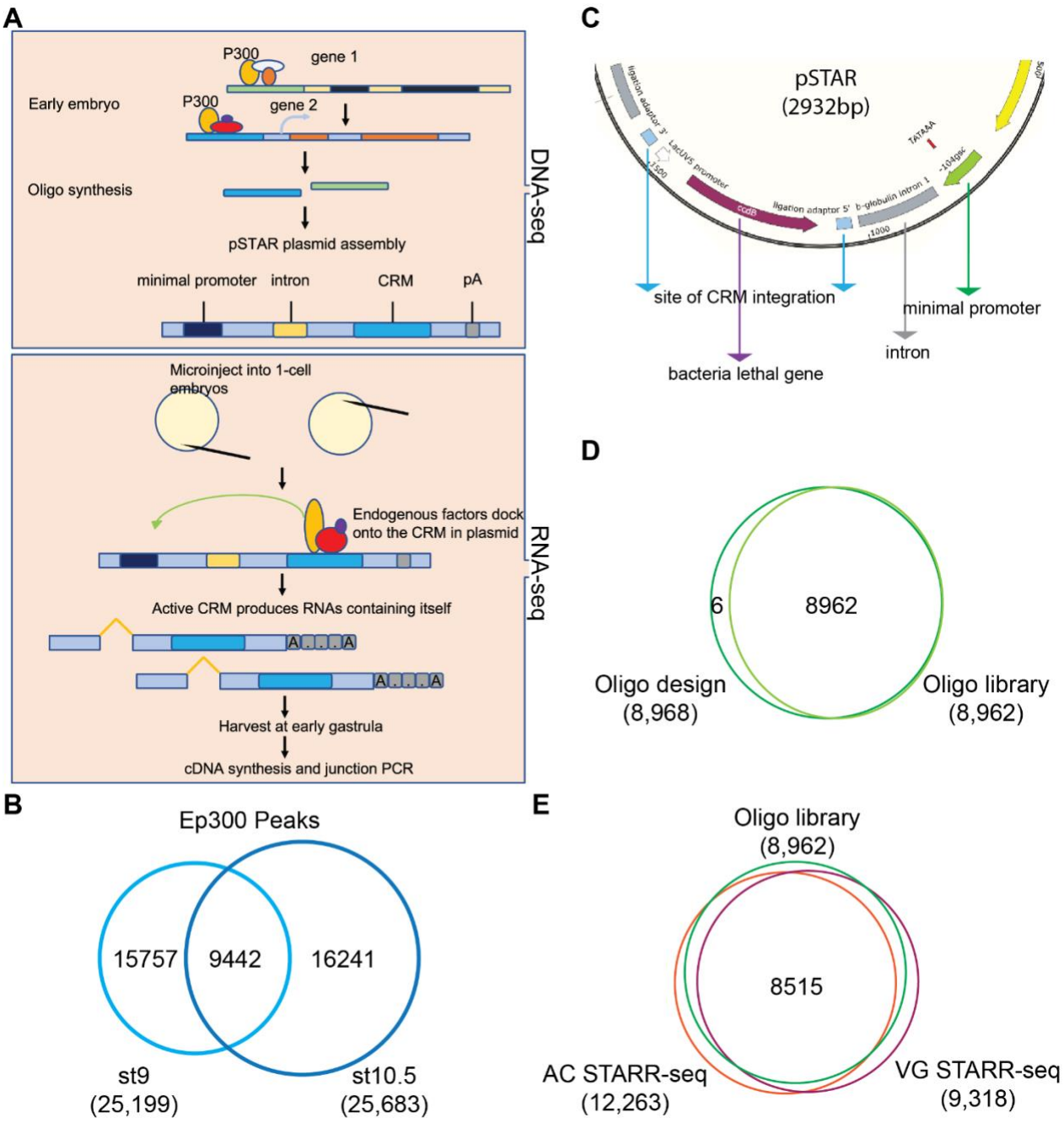

Figure 4.2: STARR-seq selectively captures transcripts derived from plasmids containing CRM oligos.

(**A**) A schematic diagram showing the workflow of STARR-seq using Ep300 peak oligos. (**B**) Venn diagram depicting Ep300 overlapping peaks between st9 and st10.5 embryos. (**C**) A map

illustrating the pSTAR plasmid designed for STARR-seq in frog. Highlighted features include a *gsc* minimal promoter, b-globulin intron one for mature mRNA selection, sites for CRM insertion, and *ccdB* lethal gene cassette for empty vector elimination. (**D**) Venn diagram depicting the efficiency of oligo plasmid library assembly. (**E**) Venn diagram showing the capture numbers of transcripts derived from the oligo plasmid library in STARR-seq experiments.

We analyzed previously published Ep300 ChIP-seq data of blastula (stage 9) and early gastrula (stage 10.5) *Xenopus tropicalis* embryos (Hontelez et al., 2015). About 9100 persistent Ep300 peaks were determined by a minimum of 50% overlap between peaks called in stage 9 and stage 10.5 (Figure 4.2B). Two hundred bp synthetic oligonucleotides (100 bps on each side of the absolute peak summit) representing these overlapping Ep300 peaks were then synthesized and cloned into pSTAR plasmid (Figure 4.2C). The resulting plasmid library was sequenced to examine the efficiency of the cloning strategy (Figure 4.2A, DNA-seq). We injected 20 pg of Ep300 plasmid library into either animal cap (presumptive ectoderm) or vegetal mass (presumptive endoderm) of a 1-cell stage *Xenopus tropicalis* embryo in a total of 150 embryos. Then embryos were harvested at the early gastrula stage (stage 10.5) and isolated RNAs for sequencing (Figure 4.2A, RNA-seq).

To demonstrate that CRMs are effectively assembled into pSTAR plasmids, we sequenced the plasmid library generated and found that nearly all oligos (>99.9%) designed are integrated into the pSTAR backbone (Figure 4.2D). This suggests our protocol of generating a complex plasmid library consisting of candidate CRMs is highly effective. We then assessed if all CRMs in the plasmid library were transcribed *in vivo*. A majority of CRM oligo-derived

transcripts (> 95%) were detected in STARR-seq experiments performed in either AC or VG tissues (Figure 4.2E), indicating that *Xenopus* embryos can transcribe CRMs from nearly all injected candidate CRMs.

CRM activities by STARR-seq do not correlate with gene activities

Ep300 is a well-known acetyltransferase catalyzing histone acetylation. We hypothesize that CRM activities measured by STARR-seq coincide with nearby gene activities in each examined germ layer. To test this hypothesis, we first calculated the activity of each CRM measured by STARR-seq using CPP (counts per plasmid) method (Barakat et al., 2018). Reads mapped to each CRM were counted and normalized using DESeq2 (Love et al., 2014), followed by normalization to input plasmid library counts. We then identified active CRMs measured by STARR-seq using a change-point method by ranking CPP values for all CRMs. We found that a drastic elevation of activities distinguishes active CRMs from the rest of CRMs with basal activities (Figure 4.3A, 4.3B). Next, active CRMs identified from change-point methods are overlapped, and each CRM from three overlapping sections is assigned to the nearest genes within 10kb (Figure 4.3C). The spatial expression patterns (Blitz et al., 2017) of each overlapping section did not show any germ-layer specificity, suggesting that differential CRM activities measured by STARR-seq do not correlate with the germ-layer specific gene expression patterns (Figure 4.3D-F). We speculate that CRMs are under-sampled in embryos such that CRMs with relatively low activities may have skewed the data. Hence, we investigated whether the most active CRMs measured by STARR-seq correlate with gene activities. The top 100 active CRMs ranked by CPP values were examined the same way described above. We found that genes associated with highly active CRMs fail to show preferential activities in respective

germ layers (Figure 4.3G, 4.3H). The inconsistency between CRM and gene activities may be partly explained by the inappropriate gene assignment based on distance vicinity. To overcome this potential issue, we examined the correlation of VG CRM activities measured by two different approaches. The linear correlation between STARR-seq measurements and luciferase assay measurements is poor (Figure 4.3I). These analyses suggest that this current version of STARR-seq method is sub-optimal in revealing the functionality of CRMs *in vivo*.

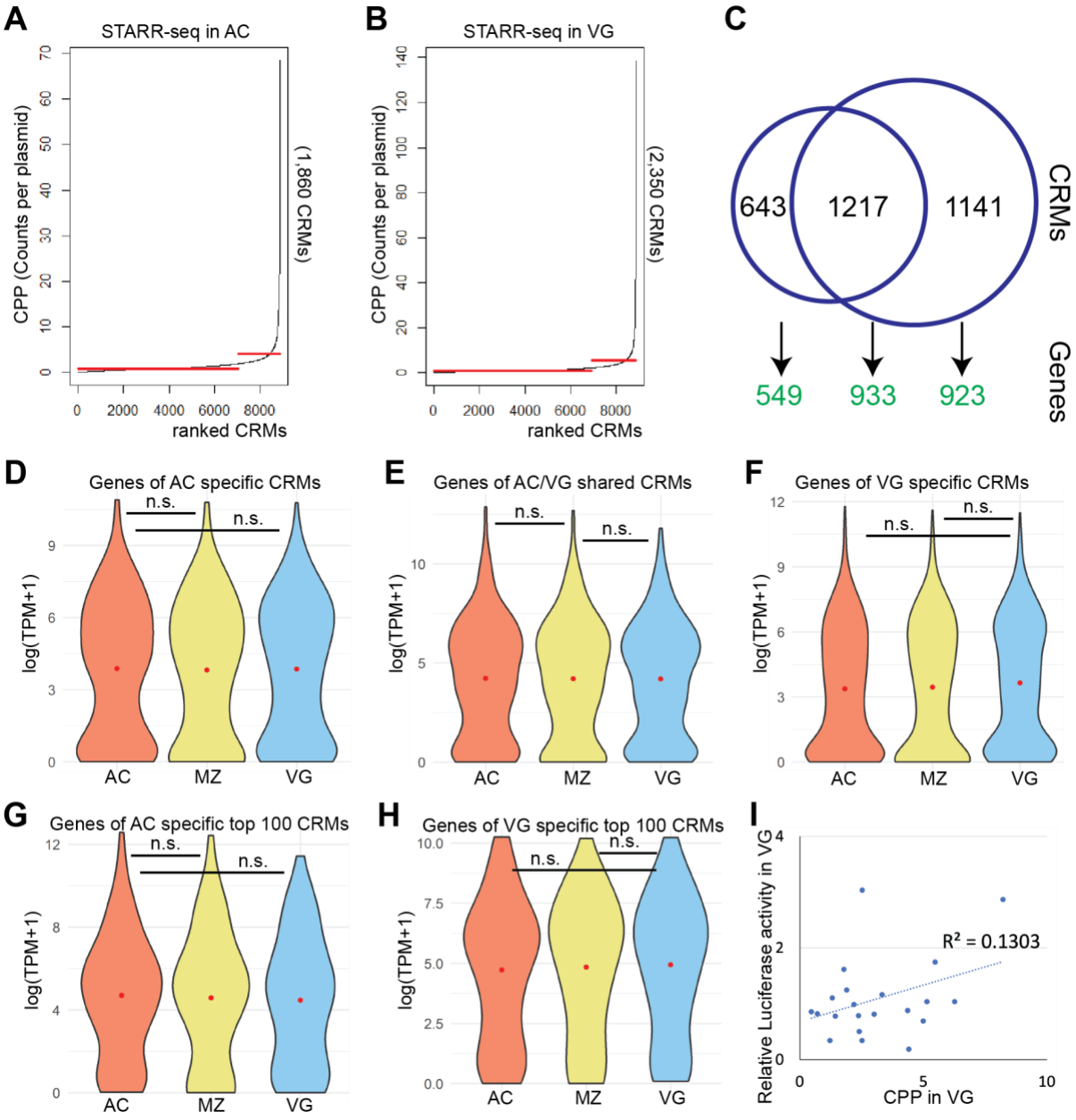

Figure 4.3: STARR-seq measured CRM activities fail to correlate spatial gene expression patterns.

(**A, B**) Change-point method identifying active CRMs in (**A**) animal cap tissues (AC) and (**B**) vegetal mass tissues (VG). Red horizontal lines depict the position where a drastic shift in value occurs. (**C**) Venn diagram showing the overlap of active CRMs in AV and VG. Green numbers denote genes associated with each overlapping sections. (**D-F**) Violin plots showing the TPM values of genes associated with (**D**) AC-specific CRMs, (**E**) AC/VG-shared CRMs, and (**F**) VG-specific CRMs in the ectoderm (AC), mesoderm (MZ), and endoderm (VG). Red dots denote average TPM values. *P*-values are calculated by Student's t-tests. (**G, H**) Violin plots showing the TPM value of genes associated with (**G**) AC-specific CRMs and (**H**) VG-specific CRMs using the top 100 active CRMs. Red dots denote average TPM values. *P*-values are calculated by Student's t-tests. (**I**) Linear correlation of known VG CRM (*nodalX2, mix1, snai1, frzb, fgf20, nodal6, gsc, mixerX2, sebox, sox17b.1, fst, mesbp, stox1, bmp7.2, hhex, admp, pcdh8.2X2, slc12a3.1*) activities measured by STARR-seq and luciferase assays. The blue dotted line depicts a linear trendline. Both x-axis and y-axis values are log-transformed.

Discussion

Here, we developed a modified STARR-seq protocol and applied this method in early *Xenopus* embryos. A plasmid library encompassing a complexity of 9000 synthetic CRM oligos was successfully constructed (Figure 4.2D). Embryos injected with this plasmid library generate transcripts containing CRM sequences from the library. This protocol strongly discriminates against endogenous transcripts and favors episomal transcripts (Figure 4.2E). However, the

activities of CRMs measured from STARR-seq correlate poorly with both endogenous gene activities (Figure 4.3D-H) and the reporter activity of individual CRM in embryos (Figure 4.3I).

We speculate that the current difficulty of applying STARR-seq to detect functional CRMs in embryos originates from the complexity of testing CRM library. The STARR-seq approach has been successful in cell culture-based systems. Plasmids are transfected into cells evenly, where each cell receives a relatively equal but small number of plasmids. This allows the population of transfected cells to utilize the received plasmids fully and transcribe plasmid-derived RNAs based on the activities of CRMs. A collection of millions of such cells, each expressing a small subset of the plasmid library, covers the full complexity of the plasmid library. However, the usage of the plasmid library in embryos after DNA microinjection is distinctly different from that of the transfected cells. First, the initial amount of plasmid library DNA injected in a 1-cell staged *Xenopus* embryo is 20pg, which encompasses the entire plasmid library complexity. This pool of plasmids is then subdivided into each blastomere after rounds of cell divisions. However, a gastrula *Xenopus* embryos have less than 10,000 cells, which leads to a much greater library-complexity-to-cell-number ratio than that of the cell culture. This could lead to a potential sub-sampling issue because the plasmid complexity tolerated by embryos is unclear. Second, injected plasmids do not freely diffuse but are primarily retained at the injection site (Jonas et al., 1989). They are passively distributed into daughter cells through cell divisions. This leads to an uneven distribution of plasmids across different cells within an embryo. Although this issue may be resolved by increasing the number of injected embryos, the appropriate sampling size (number) of embryos to overcome this difficulty is unknown. Moreover, the sampling size will vary based on the complexity of the plasmid library used.

Further experiments are needed to identify the appropriate degrees of the plasmid library complexity and the optimal sampling size of injected embryos.

Method

Protocol development: STARR-seq in Xenopus Embryos (modified from Neumayr et al., 2019)

Plasmid library

1. Resuspend synthetic oligos (ordered from Twist Bioscience) in 20ng/ul concentration using 1X TE (10mM Tris-HCl pH8.0; 1mM EDTA) and aliquot to store in -80C.

2. Using following primers to PCR amplify oligos (adaptor synthesized within) for 10 cycles:

oligo_for acacgacgctcttccgatc*t

oligo_rev gacgtgtgctcttccgatc*t

oligo(20ng)	1ul
10uM oligo_for	0.75 ul
10uM oligo_rev	0.75 ul
ddH2O	10ul
2X KAPA MM (Roche 7958927001)	12.5ul

3. Purify PCR amplicons using AMPure XP beads (Beckman A63882) at 1.8X and elute in 20ul of 0.1X TE; check the size and concentration of amplicons using TapeStation; these are assembly ready inserts.

4. PCR amplify the pSTAR-intron vector for 20 cycles with following primers:

pSTAR_HiFivector_for	agatcggaagagcacacgtc*t
pSTAR_HiFivector_rev	agatcggaagagcgtcgtgt*a
pSTAR-intron1 DNA (0.1ng)	1 ul

10uM pSTAR_HiFi_vector for	1.5 ul
10uM pSTAR_HiFi_vector rev	1.5 ul
ddH2O	21 ul
2X KAPA MM	25 ul

5. Gel-purify the pSTAR-intron1 vector and quantify; this is assembly-ready vectors.

6. Assemble 2 NEB HiFi Assembly reaction:

pSTAR-intron1 vector (50ng)	1 ul
oligo amplicon inserts (1:3)	2 ul
ddH2O	7 ul
2X NEB HiFi MM (NEB, E2621S)	10 ul

Incubate at 50C for 1 hr.

7. Pool two HiFi reactions together and add 80ul of 1X TE to a final volume of 100ul, add 10ul of 3M NaOAc, 2ug of glycogen (Thermo Scientific, R0561), and 250ul of 100% EtOH; incubate at -20C overnight to precipitate assembled DNA.

8. Pellet DNA by centrifuge at 14,000 rpm at 4C for 30 mins; wash the pellet with fresh 80% EtOH twice. Air-dry the DNA pellet in the cabinet for at least 15 mins.

9. Resuspend DNA pellet in 12ul of 0.1X TE, and run 6 ul of DNA on the gel to determine the assembly efficiency.

10. Split the remaining 6ul of assembled DNA into 3 tubes at 2ul/ tube. Add 21ul of electrocompetent cells MegaX DH10b (Invitrogen, C640003) to each tube on ice, and transfer the mixture to a pre-chilled 1.0 mm electroporation cuvette (BioRad, 1652089).

11. Electroporate the DH10b cells using 2.0 kV, 200 Ω, and 25uF conditions. Immediately add 1 ml of prewarmed (37C) recovery media to the cuvette. Gently resuspend the bacteria and transfer them to a 15ml round bottom inoculation tube.

12. Close the tube lid and shake at 220 rpm for 1 hr at 37C.

13. Plate a dilution of 1:1000 and 1:10000 to determine the transformation efficiency.

14. Add all three 1-ml recovered cultures to 200ml of freshly prepared SOB media (Tryptone 20g, Yeast extract 5g, NaCl 0.5g, 10 mL of 250mM KCl in a total of 1000ml by H2O, pH7.0), shake at 220rpm overnight (~ 12 hrs) at 37C.

15. Maxi-prep (Qiagen 12162) plasmid DNA from the overnight culture, run 500ng of plasmid on the gel, and examine the quality of the plasmid prep; if there is genomic DNA contamination in the plasmid prep, use gel purification or CsCl precipitation.

16. To generate sequencing ready oligo plasmid library, PCR 9 cycles:

pSTAR oligo library plasmid (100ng)	1 ul
10uM illumina index 5 primer	1.5 ul
10uM illumina index 7 primer	1.5 ul
ddH2O	21 ul
2X KAPA MM	25 ul

17. Purify the PCR amplicon using AMPure XP beads: using double size selection.

18. Quantify the sequencing-ready plasmid library using TapeStation.

<u>RNA extraction (100 embryos per reaction)</u>

1. Inject 20pg of plasmid DNA (2nL of 10ng/ul solution) in each embryo for each condition.

2. Harvest embryos in 100 embryos/ 1 ml Trizol (Invitrogen 15596026).

3. Homogenize embryos with vigorous vortex.

4. Incubate homogenized samples for 5 mins at RT to proceed, or store in -80C.

5. Centrifuge at 12,000x g at 4C to separate DNA and cellular debris (jelly-like pellet) and transfer clear supernatant to a new 1.5ml tube.

6. Add 100ul of 1-bromo-3-chloropropane (Sigma B9673-200mL) per 1 ml of Trizol used. Incubate for 3 mins at RT with 15s vigorous vortex/ min.

7. Centrifuge samples at 12,000x g for 15 mins at 4C.

8. Transfer the upper aqueous layer to a new non-stick 1.5ml tube as much as possible.

9. Precipitate RNA by mixing with 0.7X volume of isopropanol to the aqueous phase, incubate at RT for 10 mins, and centrifuge at 12,000x g for 20 mins at 4C; a white jelly-like pellet should be observed.

10. Remove the supernatant and wash the pellet with 1ml ice-cold 80% ethanol twice.

11. Air dry the pellet for 20 mins in a bio-cabinet.

<u>mRNA Isolation (NEB #S1550)</u>

1. Dissolve the RNA pellet in 100ul of Nuclease-free water (Sigma W4502-1L) by pipetting up and down gently, then set it on ice for 10 mins with an occasional vortex to completely dissolve the pellet.

2. Aliquot 100ul of full resuspended equilibrated dT beads to a clean non-stick 1.5ml tube (Ambion AM12450), add 200ul of Lysis/Binding buffer, vortex briefly, and agitate for 2 mins. Leave the beads in Lysis/Binding buffer until use.

2. Add 400ul of Lysis/Binding buffer to dissolved RNA to yield a total of 500ul. Remove the Lysis/Binding buffer from the beads, and transfer 500ul of RNA/buffer mixture to the beads.

3. Place the beads mixture on the agitator for 10 mins at RT, then place the tube onto the magnetic rack to pull the beads, remove and discard the supernatant.

4. Add 500ul of Wash buffer 1 to the beads, and carefully resuspend the beads without vortex. Incubate with agitation for 1 min; then place the tube onto the magnetic rack to pull the beads, remove and discard wash solution.

5. Repeat once.

6. Add 500ul of Wash buffer 2 to the beads, carefully resuspend the beads without vortex. Incubate with agitation for 1 min; then place the tube onto the magnetic rack to pull the beads, remove and discard wash solution.

7. Repeat once.

8. Add 500ul of Low Salt Buffer to the beads, carefully resuspend the beads without vortex. Incubate with agitation for 1 min; then place the tube onto the magnetic rack to pull the beads, remove and discard wash solution.

9. Add 88ul of elution buffer and vortex gently to resuspend beads, incubate at 50C for 2 mins. Separate beads with the magnetic stand, and transfer 85ul of mRNA to a clean, sterile non-stick 1.5 ml tube. Use the remaining mRNA to determine the concentration of isolated mRNA with a nanodrop.

<u>DNase (NEB M0303S) Treatment</u>

1. Prepare the following DNase reaction mixture:

Isolated mRNA (< 200ngul):	85 ul
10X DNase buffer	10 ul
NEB RNase free DNase	2.5 ul

NEB RNase inhibitor (NEB M0314S, 40U/ul) 2.5 ul

Gently tap the tube to mix

2. Incubate at 37C in a water bath for 30 mins.

3. Add 900 ul of Trizol and incubate mRNA for 5 mins at RT to proceed or store at -80C.

4. Add 100ul of 1-bromo-3-chloropropane/ 1 ml of Trizol used. Incubate for 3 mins at RT with 15s vigorous vortex/ min.

5. Centrifuge samples at 12,000x g for 15 mins at 4C.

6. Transfer the upper aqueous layer to a new non-stick 1.5ml tube as much as possible.

7. Precipitate RNA by mixing with 0.7X volume of isopropanol to the aqueous phase volume, add 2ug of glycogen, and then incubate at RT for 10 mins.

8. Pellet mRNA by centrifuging at 12,000x g for 20 mins at 4C; a white jelly-like pellet should be observed.

9. Remove the supernatant and wash the pellet with 1ml ice-cold 80% ethanol twice.

10. Air dry the pellet for 20 mins in a bio-cabinet.

11. Resuspend DNase treated mRNA in 25ul of DNase/RNase free water.

12. Use 1 ul of mRNA to determine the concentration with nanodrop or TapeStation.

<u>cDNA Synthesis</u>

1. Prepare RT reaction on ice with the following:

plasmid specific cDNA primer gactctagAgactggagttca*g

RNA 12 ul (1/2 of each mRNA from the previous step)

2uM plasmid specific cDNA primer 1.5 ul (fresh from -80 aliquot)

10mM dNTPs (Invitrogen 10297018) 1 ul

8. Incubate the sample at 65C for 5 mins and immediately put it on ice for 3 mins.

9. Add the following to RT mixture on ice:

NEB RNase inhibitor	0.5 ul (20U)
Maxima 5X First strand buffer (Thermo Scientific EP0742)	4 ul
SuperScript II RT (Invitrogen 18064022)	1 ul

10. Mix by pipetting up and down without forming bubbles.

11. Spin down the sample and use the following thermo cycles:

	42C	60 mins
	50C	2 mins
10 cycles	42C	2 mins
	70C	15 mins
	4C	hold

RNaseA Treatment

1. Pool RT reactions together from each condition to a clean PCR tube.

2. Add 0.3ul of boiled RNaseA (NEB T3018L) to pooled 60ul of RT reaction.

3. Incubate at 37C for 1 hr.

cDNA cleanup with RNAClean XP Beads

1. Add 1.8X volume of RNAClean XP beads (Beckman A66514) to 1 volume of cDNA.

2. Mix thoroughly by vortex and incubate at RT for 15 mins.

3. Transfer to magnetic stand and incubate at RT for 10 mins to allow bead separation.

4. Discard the supernatant and wash with 200ul of freshly prepared 80% EtOH twice, incubating with 80% EtOH for 2 mins each time.

5. Dry the beads at RT for 5-10 mins; elute as soon as you see beads start to crack (look at beads after 5 mins).

6. Elute cDNA with 42ul of 0.1X TE buffer, transfer 40ul of a clean, sterile non-stick tube.

Junction PCR Amplification

1. Prepare the following PCR mix on ice to a final volume of 50ul:

junction PCR fwd primer	ATCCGTAGAAATCTGCTGG*G
plasmid specific cDNA primer	gactctagAgactggagttca*g
cDNA	20 ul
2X KAPA HiFi HotStart MasterMix	25 ul
10uM junction PCR fwd primer	1.5 ul
10uM plasmid specific cDNA rev primer	1.5 ul
ddH2O	2 ul

2. Perform PCR using following PCR conditions

	98C	3 mins
	98C	20s
~12 cycles	65C	15s
	72C	20 sec
	72C	5 mins
	4C	hold

3. Pool all junction PCR amplicons and purify with AmpPure beads, elute in 25ul 0.1X TE

4. Run 1ul of elution to determine the concentration of junction PCR amplicon and calculate how many cycles should run for sequencing-ready PCR. Do not use less than 5 cycles of PCR reactions for sequencing-ready PCR.

Sequencing ready PCR Amplification

1. Prepare the following PCR mix on ice to a final volume of 50ul:

Junction PCR amplicons	10 ul
2X KAPA HiFi HotStart MasterMix	25 ul
10uM illumina index 5 primer	1.5 ul
10uM illumina index 7 primer	1.5 ul
ddH2O	12 ul

2. Perform PCR using following PCR conditions

	98C	3 mins
	98C	20s
5-9 cycles	65C	15s
	72C	20 sec
	72C	5 mins
	4C	hold

3. Clean PCR product using AMPure XP beads, and elute in 20ul of 0.1X TE, examine the concentration of this sequencing ready sample using TapeStation. The sequencing library should have a concentration of about 2ng/ul.

Animal Model and Subjects

Xenopus tropicalis embryos were obtained by *in vitro* fertilization according to Ogino et al. (2006) and staged according to Nieuwkoop and Faber (1994). All embryos were cultured in 1/9X Marc's modified Ringers (MMR) at 25 °C. For plasmid injection, 20pg of plasmid was injected at either the animal cap site or vegetal mass site of the 1-cell staged embryos. Then injected embryos were cultured to st10.5 early gastrula and harvested for RNA extraction as described above. Animals were raised and maintained following the University of California, Irvine Institutional Animal Care Use Committee (IACUC). Animals used were grown in the laboratory and/or purchased from National *Xenopus* Resource.

Plasmids and Reporter Assays

Each candidate CRM was cloned into pGL3 (Promega) with a minimum *gsc* promoter (Watabe et al., 1995) between HindIII individually using PCR. For mutant CRMs, base substitutions (Foxh1 motifs AATMHACA were changed to AAGMHAAA and Sox17 motifs ACAAWRG were changed to ATAGWRG) were introduced by site-directed mutagenesis PCR. Primers for PCR are listed in Table 4.1.

For dual luciferase reporter assays, 20 pg of a CRM plasmid and 2pg of pRL-SV40 plasmid (Promega) were co-injected into either the animal or vegetal site of the 1-2 cell stage embryos. Embryos were then incubated to and harvested at early gastrula st10.5. Ten embryos were lysed in 100ul of passive lysis buffer (Promega), and 10ul of clear lysate was used to assay for reporter activities. pGL3 without CRM cloned was used as a negative control.

Computational Analysis

Sequenced samples were aligned to *Xenopus tropicalis* genome v10.0 (http://www.xenbase.org/, RRID:SCR_003280) using Bowtie2 (Langmead and Salzberg, 2012). After PCR duplicate removal, bam files were converted to bed files. Raw counts to each CRM (oligo sequence) were calculated against the oligo bed file using Bedtools (Quinlan and Hall, 2010) and then normalized in DESeq2 (Love et al., 2014) to minimize distortion by sequencing depth. Values of counts per plasmid (CPP) is calculated by normalizing counts from STARR-seq samples to counts from the plasmid library sample. The threshold for active CRMs is defined by the changepoint of the curve plotted by CPP ranks versus CPP values using the *changepoint* package in R (R core team, 2022). The active CRMs were annotated by assigning them to their nearest genes within 10kb. For spatial gene expression analysis (Blitz et al., 2017), all sequencing samples were aligned using STAR v2.7.3a (Dobin et al., 2013) to *Xenopus tropicalis* genome v10.0 (http://www.xenbase.org/, RRID:SCR_003280) to obtain raw read counts. RSEM v1.3.3 (Li and Dewey, 2011) was used to calculate expression values in transcripts per million (TPM). The average TPM between 3 dissected mesoderm tissues (dorsal, marginal, and lateral marginal zones) was used to represent the expression of mesoderm (Blitz et al., 2017). Student's t-test is used to determine the statistical significance between examined groups.

Table 4.1: CRM PCR primer list

CRMs	Forward sequence	Reverse sequence
admp	atatatatatatatactaacagtatatcttgcccaaag	aagtaaacttgcaacttaaaaaattaaattttatttc
bmp7.2	aatgaatgaatgtgagggct	acactgttatgtagtgtacctgctt
fam184a-like	ttctgttttgtttggtttccc	tttttgcacctattgttctaagg
fgf20 (1)	ctgtttcggcatgtagttgcctt	gagaatgtttcagaaagggtggg
fgf20 (2)	actgatttatcagcctgggtgt	aatgtgcagggcacacgg
frzb	accctgctttagaatacacatgagg	tggcactctagtctgggagaa
fst	tagctgctaaagctgcattatt	aaattggccgacccttttc
grhl3	ttaattgcagtaccaccaacag	tcagtactggcccccctgc
gsc	gaggtgcaagaatccttgtgt	ttctcaatcatttattcatggttgt
hhex	ttaaccttattacattttcaggcc	tttagagtgtgtctgtgtgtgttt
map7d3	agttttccttccaccaaagaaaa	agcttgcctgtatgggat
mesbp	atgcttgcatttcaaacattg	gtgtaatgtttgccctatccc
mix1 (1)	cctggagagaggggcagacta	tctctactgtctccctcgtgt
mix1 (2)	ggcatctacacctacccagg	atagtggatgggcagatacag
mix1 (3)	attactgggtgttagagagccag	ggggatttgcaaggcattcc
mix1 (4)	tggggcaataaaaacaatg	gccttttcattttaggggtta
mixer (1)	gcagacacttctccccaata	gaaagtagtgagtcatgtgacag
mixer (2)	tatataaaagcggcggca	aatgtatttcagttccaaacatg
nodal/nodal 4 (1)	ggtcagacacatttttgagca	tccacatcagcacactcctt
nodal/ nodal4 (2)	gtgaatctgtgagtgtttgatg	attacctgagtgatctcaagcctt
nodal1	agttttgcctaagatgaagggaa	gcacaaagacaacatgagccc
nodal6 (1)	tttgatgccagtaacaggtaac	agtatacactaaatgaagttagttaagata
nodal6 (2)	tactaagatttcacttcatggtgg	gcagcaagttaaggcataaca
nodal6 (3)	ttacacacatcttcagtttattcac	acttcatgccctgagaatgcc
nodal6 (4)	tcactaaatagaaatacaactgtgc	gagacagtcagatactgattaca
pcdh8.2 (1)	aaatctctttcatattcagccgg	tgagttgttttatgcaatatatttttttatagaggc
pcdh8.2 (2)	acctaaagtcacatcccatcag	ttgatgacatcaagaaaggtactctaatc
pcdh8.2 (3)	ggtgcagtgaatggcttattc	caccttagtgccttcataattgg
pdk4	agaactaaaactgttataagaatttctaatttttaataaatatt	gtaaagttgcactgctttatttttacac
sebox	atgtcccattgcagccagtt	gcacttcaaacagctggaga
serpinf2	agaaatggtgcaccactg	tcaaaatcatgcactgaaggatcaa
sfrp2	aatggagaaaagtgtggtataagc	acactgctacttttttaagacagat
sia1	caatgatttcagtgtggcagg	gtccgactcaagttgctcct
slc12a3.1	cttttccagattcaaaaaatattaacaatctgg	caccttgattaaaaatccatggtcc
snai1	tcagacaccaatctcacttttttgc	tgtctgccctttacagccaa
sox17b.1	acagaacacaatcaattagcata	ccatgtgtcaaccgaaactgc
stox1	taacagctggagggttaaaag	gaaaaagcccccatttactg
tcf7	gcttgccctaccctgcct	caaattgccttaaaagagtagtg

CHAPTER 5

Conclusions and Outlook

My dissertation focused on several aspects regulating the zygotic genome activation, which dictates the delineation of three primary germ layers, during early *Xenopus* development. First, I studied a repressive role of Foxh1 for suppressing mesendodermal genes in the ectoderm through proteomic and transcriptomic analyses of maternal transcription factor Foxh1. Second, genome-wide investigation of epigenetic modifier Hdac1 revealed a dual function model of Hdac1 where it functions both in gene repression and activation through modulating histone acetylome. Lastly, I modified the existing cell-culture-based STARR-seq method to develop a high-throughput CRM assessment assay in *Xenopus* embryos. This chapter discusses these major highlights of my dissertation work and their implications, and provide future prospects to further address the mechanisms governing ZGA.

Foxh1 interacts with both positive and negative epigenetic modifiers

To investigate the protein interaction network of Foxh1 in *Xenopus* embryos directly is challenging. First, early staged embryos are loaded with abundant yolks that interfere with the immunoprecipitation of endogenous protein complexes (Conlon et al., 2012). Second, mass spectrometry detection is sensitive, requiring an availability of highly specific antibody to minimize the non-specific interaction with the protein of interest. To overcome these limitations of *early Xenopus* embryos in the proteomic study, we used mouse embryonic stem cells, which are pluripotent like *Xenopus* blastula embryos. Ectopic Foxh1 with triple tags is integrated into the genome of E14 cells as a transgene, and the immunoprecipitation is carried out using two

antibodies against two tags simultaneously (Wang et al., 2014). This approach allows us to identify protein interaction network of Foxh1 efficiently in E14 cells, followed by the validation experiments using *Xenopus* proteins.

Our results revealed 371 candidate interacting proteins to FOXH1 in mESC (Table 2.1). These proteins are involved in diverse cellular processes, including RNA biogenesis, ribosome biogenesis, chromatin regulation, etc. Given that Foxh1 is a maternal transcription factor that imposes transcriptional regulation on zygotic genes, the candidate interactants related to chromatin regulation may elucidate the mechanisms of such regulation. We highlight that Foxh1 can interact with both positive and negative chromatin regulators (Figure 2.2A). Foxh1 acts as a Nodal signaling co-activator for the transcriptional network of mesendodermal genes via interaction with Smads, presumably accompanied by epigenetic regulations. For example, Foxh1 may interact with SWI/SNF complexes (which include identified interactants SMARCA5 and HELLS) to facilitate the nucleosome repositioning during transcription. In addition, histone methyltransferase complexes depositing active histone methylation (which include identified interactants NSD1) may be recruited by Foxh1 during transcription. The activator role of Foxh1 on chromatin may serve as a mechanism by which Foxh1 acts as a passive pioneer transcription factor (Charney et al., 2017) to induce the competence of the zygotic genome for ZGA. However, these interactions identified by mass spectrometry need further validations in embryos. For instance, the native genomic loci of these interactions as well as the exact role of Foxh1 in their genomic engagement require additional examinations.

In a study examining genes regulated by Foxh1, more than half of the genes are independent of Nodal signaling, indicating a role for Foxh1 other than the co-activator of Nodal signaling (Chiu et al., 2014). The spatial expression profiles of Foxh1 regulated genes

independent of Nodal signaling are relatively ubiquitous, like the distribution of Foxh1 transcripts (Chiu et al., 2014; Charney et al., 2017). These findings suggest that Foxh1 modulates genes in all germ layers, which is different from the mesendodermal restricted role of Nodal signaling. Our result revealed that Foxh1 interacts with repressive chromatin regulators such as Polycomb repressive complexes, Hdac-containing complexes, and DNA methylation complexes (Figure 2.2A). The identified interaction between Foxh1 and negative chromatin regulators led us to speculate that Foxh1 is involved in gene repression through epigenetic mechanisms. The repressive role of Foxh1 has been elucidated in previous studies where Foxh1 inhibits the expression of *nodal5* and *nodal6* in the ventral vegetal mass of embryos through Nodal signaling (Kofron et al., 2004). Interestingly, Foxh1 was shown to recruit repressive transcriptional cofactor Tle4 to regulate zygotic gene expression (Reid et al., 2016; Charney et al., 2017). We propose that Foxh1 interacts with these negative chromatin regulators to form a repressive chromatin state primed for gene repression.

The repressive role of maternal Foxh1 in developing Xenopus ectoderm

The genome-wide binding profile of Foxh1 reveals that Foxh1 persistently binds to genomic loci of mesendodermal genes during early development (Charney et al., 2017). This binding model fits the role of Foxh1 as a co-activator for Nodal signaling, where mesendodermal genes are expressed in the marginal zone and the vegetal mass of embryos (Hill et al., 2001). However, Foxh1 is uniformly expressed in all germ layers, which raises the question as to the function of Foxh1 in the ectoderm. The protein interaction network of Foxh1 allows us to infer that Foxh1 functions to suppress mesendodermal genes through its association with negative chromatin regulators in the ectoderm. Co-immunoprecipitation experiments showed that Foxh1 interacts

with PRC2 core subunits and Hdac1 through protein-protein interaction (Figure 2.2B-2.2H). Genome-wide binding profiles also showed that Foxh1 occupies a majority of Hdac1 bound regions (Figure 3.3C, 3.4C). These results support the hypothesis that Foxh1 interacts with negative epigenetic modifiers in the mesendodermal gene loci.

To address the functional outcome of Foxh1-Hdac1 interaction, transcriptomic analyses were conducted in the ectoderm from embryos depleted of Foxh1 or HDAC activity. While inhibition of HDAC activity in the ectoderm upregulates several known mesendodermal genes (Figure 3.10B), depletion of Foxh1 in the ectoderm does not lead to upregulation of well-known mesendodermal genes (Figure 2.3C). Overlapping genes repressed by Foxh1 and Hdac1 in ectoderm revealed a small subset of genes (~10%) that can be suppressed by Foxh1 through its interaction with Hdac1. These observations suggest several possibilities. One is that the utilization of HDAC activity by Foxh1 in gene repression is not a primary mechanism. Foxh1 may mainly exploit the activity of PRC2 complex for H3K27me3 mediated gene repression and Hdac1 only regulates a small subset of genes. The functional role of Foxh1-PRC2 interaction requires further investigations. Alternatively, loss of Foxh1 is insufficient to fully abolish the recruitment of Hdac1 because Hdac1 can be direct to target loci by the presence of other TFs. This is consistent with the finding that motifs of Sox, Pou, and T-box TFs are also enriched within the genomic regions bound by Foxh1 (Figure 3.3B, Chiu et al., 2014; Charney et al., 2017; Paraiso et al., 2019; Gentsch et al., 2019). Since the combinatorial interaction between these TFs regulates the output of mesendodermal genes; hence, the lack of one transcription factor, in this case, Foxh1, may be insufficient to reduce significant transcriptional changes. In the future, ectodermal transcriptome will need to be examined when additional TFs are depleted along with Foxh1. This experiment will inform us about the robustness of transcriptional

networks specifying the ectoderm and the roles of different TFs in gene suppression through cooperative binding.

Hdac1 binds to genomic loci pervasively from ZGA and onward

Our ChIP-seq analyses showed that genome-wide binding of Hdac1 coincides with ZGA (Figure 4.1C); however, the zygotic transcription is dispensable for the genomic recruitment of Hdac1(Figure 4.3A). We speculate that transcriptional regulators engage zygotic genomic loci in sequential order. Maternal TFs such as Foxh1 (Charney et al., 2017), Vegt, and Otx1 (Paraiso et al., 2019) bind to the genome as early as 32~64 cell stages. The binding of maternal TFs might act as a prerequisite for binding subsequent epigenetic modifiers such as Hdac1. By comparing binding profiles of Hdac1 with Foxh1 and Sox3's, we found that Hdac1 and Foxh1/Sox3 are positively correlated in binding patterns and strength (Figure 4.3C, Figure 4.4C-F). The observed co-binding of Hdac1 and Foxh1/Sox3 on the same DNA sequences (Figure 4.3D-E, Figure 4.4G) further supports the speculation that maternal TFs direct the genomic recruitment of Hdac1. It is shown that H3K4me1 requires maternal TFs Foxh1, Otx1, and Vegt for its deposition at endodermal gene loci (Paraiso et al., 2019; Paraiso et al., preprint). It is tempting to presume a general mechanism by which maternal TFs bind to the zygotic genome early in developmental time and subsequently recruit epigenetic modifiers to remodel the zygotic epigenome.

Though the binding of Hdac1 is highly correlated with the binding of maternal TFs Foxh1 and Sox3, a causal link, whether maternal TFs such as Foxh1 and/ or Sox3 are required for the genomic recruitment of Hdac1, has not been established. However, properly addressing this question is challenging because morphant or knockout embryos for Foxh1 are often developmentally delayed, and this affects interpretation of the data. It may be needed to perform

the genome-wide binding study of Hdac1 at closely spaced time intervals using both wild type and Foxh1 mutant embryos, and carefully examine the changes of Hdac1 binding density between pairwise samples. Another useful experiment is to perform time-course ChIP-mass spectrometry (Wang et al., 2013; Ji et al., 2015) on maternal TFs like Foxh1 to determine the assembly of transcriptional complexes on the genome. For instance, ChIP-mass spectrometry on Foxh1 at the early blastula, late blastula, and early gastrula stages may reveal a temporally controlled assembly of protein complexes associated with Foxh1 on the genome. A potential challenge to this experiment is to discern specific interactants over non-specific interactions to Foxh1 because mass spectrometry is highly sensitive. Proper controls such as ChIP-mass spectrometry using a non-specific IgG antibody can exclude some non-specific interactants due to the antibody quality. Introduction of Foxh1 mutant lacking DNA-binding domain into Foxh1 null embryos may identify Foxh1 interacting proteins through a transcription-independent regulation. At present, it is unclear how many embryos are required to perform the proposed ChIP-mass spectrometry.

A dual function model of Hdac1 at inactive versus active CRMs

Genomic loci bound by Hdac1 are marked with distinct epigenetic signatures suggesting the differential functions of these CRMs. Repressive histone modification H3K27me3 is enriched at a subset of Hdac1 peaks indicating that Hdac1 mediates gene repression (Figure 3.5B). In contrast, various active histone modifications, including H3K27ac and H3K4me3, are also present at Hdac1 peaks suggesting that Hdac1 functions at active gene loci (Figure 3.5A, 3.6B). Thus, the binding of Hdac1 to both inactive and active genomic loci led us to hypothesize that Hdac1 carries out different functions on these two types of CRMs in early embryos.

To understand the functional outcome of Hdac1 occupancy, we quantitively examined the changes in histone acetylation (pan-H3Kac) and transcriptomes upon inhibiting the enzymatic activity of HDACs in two different germ layers. We observed a global histone hyperacetylation on all Hdac1 peaks indicating that Hdac1 removes histone acetylation at bound regions. Yet, the transcriptomic changes are different on genes associated with inactive versus active Hdac1 peaks. At inactive CRMs, inhibition of HDAC activity results in a drastic histone hyperacetylation (Figure 3.7E, 3.7F, 3.8G, 3.8J), leading to the aberrant activation of nearby genes (Figure 3.9B-E, 3.10G). Such repressive regulation by Hdac1 is imposed on inactive genes both temporally and spatially. Thus, one function model of Hdac1 is to safeguard the misactivation of developmental genes both in time and space by preserving a hypoacetylation state of CRMs (Figure 3.9F, inactive CRM model).

Unexpectedly, our data revealed that the gene activity requires a dynamic equilibrium between histone acetylation and deacetylation. We found an anti-correlation between histone acetylation and gene expression levels, where histone acetylation is elevated on active CRMs (Figure 3.7E, 3.7F, 3.8G, 3.8J), yet the expression levels of genes associated with active CRMs are decreased (Figure 3.9E, 3.10G). This positive transcriptional regulation by Hdac1 is observed in both ectoderm and endoderm germ layers. Hence, the other function model of Hdac1 is to maintain gene activities through sustaining histone acetylation-deacetylation cycles on active CRMs (Figure 3.9F, active CRM model). These results highlight that the functional outcome of Hdac1 activity is dependent on the context of genomic loci, and influences gene expression differently.

While the simultaneous binding of Ep300 and Hdac1 for histone acetylation cycles is demonstrated in cell culture studies (Wang et al., 2009; Kidder et al., 2012), this first occurrence

of this cyclical histone acetylation phenomenon in embryos at active genes is described here. The binding profiles of Hdac1 and Ep300 are well correlated (Figure 3.6A), yet the direct co-occupancy of both factors on the same CRMs is not examined. The sequential ChIP-qPCR can be used to determine whether Ep300 and Hdac1 bind to the same genomic loci simultaneously. Upon HDAC inhibition, we observed a global histone hyperacetylation inferring that HATs are present at these genomic loci to acetylate histones. In order to further strengthen the cyclical histone acetylation model, it would be useful to perform quantitative histone acetylation assays after inhibiting HAT activity. A-485, a selective catalytic p300/CBP inhibitor (Lasko et al., 2017), can be used to treat embryos, followed by histone acetylation (pan-H3Kac) quantification. If the proposed model of cyclical histone acetylation is accurate, we then expect to observe a global histone hypoacetylation on the genome of embryos treated with the HAT inhibitor.

Hdac1 is broadly accepted as a transcriptional repressor because histone deacetylation is closely associated with gene repression. However, many studies, including ours, revealed that Hdac1 binds to many genomic loci with active transcription (Kurdistani et al., 2002; Wang et al., 2002; Wang et al., 2009; Kidder et al., 2012). This finding suggests a role for Hdac1 during active transcription. Interestingly, HDAC-containing complexes are recruited by co-transcriptional histone methylation in yeast (Carrozza et al., 2005; Keogh et al., 2005; Li et al., 2007). Furthermore, HDAC depletions in yeast, mammalian cells, and result in attenuated transcription fidelity characterized by the production of cryptic transcripts (Kim and Buratowski, 2009; Kim et al., 2012; Heo at al., 2021; Brocks et al., 2017; Milstone et al., 2020). Unfortunately, detection of cryptic transcripts in *Xenopus* is challenging because the current *Xenopus* gene annotation is not good enough to inform the precise transcriptional start sites. Instead, 5' RNA sequencing methods such as CAGE (cap analysis of gene expression)

sequencing (Adiconis et al., 2018) can be used to identify the transcription start sites and to map differential utilization of canonical versus cryptic transcription start sites in embryos treated with the HDAC inhibitor.

We propose that Hdac1 can function at both inactive and active CRMs, yet the mechanism by which Hdac1 engages these two distinct states of CRMs is unknown. In physiological contexts, Hdac1 can be incorporated into four different protein complexes (Milazzo et al., 2020). Sin3A/ B complex is the prototypical co-repressor complex mediating histone deacetylation facilitated gene repression. CoREST complex consists of both deacetylase and demethylase (LSD1) to modulate histone states during gene repression. NuRD complex carries ATP-remodeling activity besides histone deacetylation activity. MiDac complex, the last Hdac1-containing complex identified, is shown to modulate cell cycles. Considering that Hdac1 can be assembled into different complexes with different activities, it is tempting to speculate that one type of Hdac1-containing complex may play a role in inactive CRMs, while another kind of Hdac1-containing complex functions at active CRMs. If different Hdac1-containing complexes are deployed at distinct genomic loci, we then expect to find different Hdac1-containing complexes present at distinct genomic loci based on the activities of CRMs. For instance, NuRD complex can remodel the local chromatin states, reposition the nucleosome, and facilitate transcriptional machinery assembly. If HDAC activity of NuRD complex is responsible for the cyclical histone acetylation at active CRMs, we then expect to find specific subunits such as CHD3 or MTA1 in NuRD complex to be present at active but not inactive CRMs. A survey of ChIP-seq on several complex-specific subunits is needed to differentiate the genomic loci bound by one complex but not the other, which may infer the complex-specific role of Hdac1 at distinct CRMs.

It is puzzling that elevated histone acetylation on active CRMs leads to decreased gene expression levels (Figure 3.8G, 3.9E, 3.10G). How excessive histone acetylation impairs the expression level of nearby genes is unclear. Chromatin regions with overly acetylated histones may destabilize the chromatin leading to an unfavored platform for the assembly of transcriptional machinery. To test this possibility, the binding of RNA polymerase II should be assessed in embryos treated with and without the HDAC inhibitor through spike-in quantified ChIP-seq. Specifically, chromatin from *Drosophila* S2 cells can be mixed with *Xenopus* chromatin; an antibody against *Drosophila* histone variant H2Av is then introduced along with the antibody against the protein of interest. The relative signal density of the protein of interest can be obtained by normalizing to the signal density of H2Av (Egan et al., 2016). If excessive histone acetylation disrupts the assembly of general transcriptional machinery, we then expect to see a decreased signal density of RNA polymerase II at CRMs associated with downregulated genes. Histone acetylation reader proteins such as BRD proteins are essential to the formation of enhancer-promoter interaction, which in turn stimulates the activity of basal transcriptional machinery (Hassan et al., 2007; Filippakopoulos et al., 2012). It is also possible that a global histone hyperacetylation by HDAC inhibition alters the binding of BRD proteins, as reported (Slaughter et al., 2021). If global hyperacetylation reshapes the binding patterns of BRD proteins (e.g., BRD4), we would expect to find a reduced binding enrichment of BRD4 at genomic loci associated with downregulated genes through spike-in quantified ChIP-seq experiments. Histone acetylation also modulates the flexibility of chromatin, thereby determining the loop formation between chromatin regions (Li et al., 2006). Global histone hyperacetylation might result in aberrant chromatin loop formation (Gryder et al., 2019). If excessive histone acetylation at active CRMs affects loop formation, we would then expect to detect decreased chromatin contacts

between genomic loci associated with downregulated genes by chromatin conformation capture assays (Hagège et al., 2007).

Applications and limitations of STARR-seq in Xenopus embryos

The direct high-throughput functional measurement of CRM activities (STARR-seq) was conducted in various cultured cell lines (Arnold et al., 2013; Liu et al., 2017; Barakat et al., 2018; Schöne et al., 2018; Wang et al., 2018). However, it is unclear whether STARR-seq is applicable in developing embryos. We modified the published STARR-seq protocol (Neumayr et al., 2019) and applied it to *Xenopus* embryos. Our modified protocol showed effective capture of transcripts derived from input plasmids (Figure 4.2E), suggesting that the technical flow of STARR-seq is suitable *in vivo*. Yet, the activity of each CRM is stochastic, poorly correlating with the expression levels of associated genes. This indicates that the current modified protocol is suboptimal and requires further modifications.

The major challenges to applying STARR-seq in embryos are the range of library complexity and sample size. Unlike the cell culture system, where a (relatively) highly complex plasmid library can be assayed across tens or hundreds millions of cells, an early gastrula *Xenopus* embryo bears much fewer cells (~10,000 cells/ embryo), significantly limiting the testable complexity of the plasmid library. Therefore, the testable range of plasmid library complexity must be determined. Similar to the concept of spike-in strategy, a set of CRMs with known activities should be mixed in the plasmid libraries with different complexity. And the relative activities of these known CRM should be assessed. For example, the transcript levels of a set of ten CRMs can be examined individually so these ten plasmids can be served as a standard baseline. If the relative activities of these ten known CRMs were not affected after

mixing them with additional 90 different CRMs, we then can presume the activities of other 90 CRMs are accurate and conclude that an embryo can tolerate the STARR-seq library complexity of 100 different CRMs.

The inconsistent applicability of STARR-seq between *Xenopus* embryos and cell culture might also be due to insufficient sample size. The input plasmid library is not freely diffusible but locally enriched around the site of microinjection (Jonas et al., 1989), and it is passively distributed into daughter cells after ~10 cycles of cell division at early gastrulation. Each daughter cell would receive different amounts of input plasmids; therefore, the output (transcript levels derived from the input plasmid library) would vary from one cell to another. This uneven distribution of plasmids across cells may skew the CRM activity because cells that receive fewer plasmids experience less competition between plasmids and endogenous transcription. Though this is the nature of embryo development, this issue can be potentially combated by increased sample size. The variation among cells within a single embryo can likely be decreased by repeated sampling at a greater amount. The exact sample size of embryos would need to be determined experimentally in conjunction with the complexity of the input plasmid library.

In summary, my dissertation work highlights the diverse and complex regulatory mechanisms governing ZGA. My work suggested a repressive role of maternal TF Foxh1 in the ectoderm, revealed a spatial and temporal regulation of gene expression by Hdac1 through histone acetylome refinement, and built a foundation to assess the functionality of numerous CRMs simultaneously in *Xenopus*. The proposed future directions discussed in this chapter will further advance our understanding of ZGA regulation at the molecular level.

REFERENCES

Adiconis X, Haber AL, Simmons SK, Levy Moonshine A, Ji Z, Busby MA, Shi X, Jacques J, Lancaster MA, Pan JQ, Regev A, Levin JZ. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat Methods*. 2018 Jul;15(7):505-511

Ahmad K, Henikoff S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell*. 2002 Jun;9(6):1191-200.

Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, Françoijs KJ, Stunnenberg HG, Veenstra GJ. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. *Dev Cell*. 2009 Sep;17(3):425-34.

Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol*. 2018 Apr;19(4):229-244.

Almouzni G, Khochbin S, Dimitrov S, Wolffe AP. Histone acetylation influences both gene expression and development of *Xenopus laevis*. *Dev Biol*. 1994 Oct;165(2):654-69.

Amin NM, Tandon P, Osborne Nishimura E, Conlon FL. RNA-seq in the tetraploid *Xenopus laevis* enables genome-wide insight in a classic developmental biology model organism. *Methods*. 2014 Apr 1;66(3):398-409.

Anderson JD, Lowary PT, Widom J. Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol*. 2001 Apr 6;307(4):977-85.

Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A. Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during *cis*-regulatory evolution. *Nat Genet*. 2014 Jul;46(7):685-92.

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. 2013 Mar 1;339(6123):1074-7.

Ayer S, Benyajati C. Conserved enhancer and silencer elements responsible for differential Adh transcription in *Drosophila* cell lines. *Mol Cell Biol*. 1990 Jul;10(7):3512-23.

Bailey TL, Machanick P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res*. 2012 Sep 1;40(17):e128.

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, Xie B, Daley GQ, Church GM. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*. 2009 Apr;27(4):361-8.

Baltus GA, Kowalski MP, Tutter AV, Kadam S. A positive regulatory role for the mSin3A-HDAC complex in pluripotency through Nanog and Sox2. *J Biol Chem*. 2009 Mar 13;284(11):6998-7006.

Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*. 1983 Jul;33(3):729-40.

Banerji J, Rusconi S, Schaffner W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 1981 Dec;27(2 Pt 1):299-308.

Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, Chambers I. Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell*. 2018 Aug 2;23(2):276-288.e8.

Beacon TH, Delcuve GP, López C, Nardocci G, Kovalchuk I, van Wijnen AJ, Davie JR. The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clin Epigenetics*. 2021 Jul 8;13(1):138.

Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell*. 2010 Nov 16;19(5):698-711.

Berger MF, Mardis ER. The emerging clinical relevance of genomics in cancer medicine. *Nat Rev Clin Oncol*. 2018 Jun;15(6):353-365.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006 Apr 21;125(2):315-26.

Bernstein BE, Tong JK, Schreiber SL. Genomewide studies of histone deacetylase function in yeast. *Proc Natl Acad Sci U S A*. 2000 Dec 5;97(25):13708-13.

Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. 2002 Jan 1;16(1):6-21.

Blitz IL, Cho KW. Anterior neurectoderm is progressively induced during gastrulation: the role of the *Xenopus* homeobox gene orthodenticle. *Development*. 1995 Apr;121(4):993-1004.

Blitz IL, Cho KWY. Control of zygotic genome activation in *Xenopus*. *Curr Top Dev Biol*. 2021;145:167-204.

Blitz IL, Paraiso KD, Patrushev I, Chiu WTY, Cho KWY, Gilchrist MJ. A catalog of *Xenopus tropicalis* transcription factors and their regional expression in the early gastrula stage embryo. *Dev Biol*. 2017 Jun 15;426(2):409-417.

Blum M, Gaunt SJ, Cho KW, Steinbeisser H, Blumberg B, Bittner D, De Robertis EM. Gastrulation in the mouse: the role of the homeobox gene *goosecoid*. *Cell*. 1992 Jun 26;69(7):1097-106.

Blythe SA, Cha SW, Tadjuidje E, Heasman J, Klein PS. beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Dev Cell*. 2010 Aug 17;19(2):220-31.

Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJ, Gómez-Skarmeta JL. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res*. 2012 Oct;22(10):2043-53.

Bogdanovic O, Long SW, van Heeringen SJ, Brinkman AB, Gómez-Skarmeta JL, Stunnenberg HG, Jones PL, Veenstra GJ. Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Res*. 2011 Aug;21(8):1313-27.

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EE. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet*. 2012 Jan 8;44(2):148-56.

Bright AR, van Genesen S, Li Q, et al. Combinatorial transcription factor activities on open chromatin induce embryonic heterogeneity in vertebrates. *EMBO J*. 2021;40(9):e104913.

Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, Li J, Zhang B, Hou Y, Laudato S, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet*. 2017 Jul;49(7):1052-1060.

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011 Feb 4;144(3):327-39.

Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, Raju A, Lieberman Aiden E, Li S, Manikandan L, Tergaonkar V, Tucker-Kellogg G, Fullwood MJ. H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat Commun*. 2021 Jan 29;12(1):719.

Calhoun VC, Stathopoulos A, Levine M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc Natl Acad Sci U S A*. 2002 Jul 9;99(14):9243-7.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*. 2005 Nov 18;123(4):581-92.

Cha SW, 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.* 2012;7(7):e41782.

Chafin, D. R., Guo, H. & Price, D. H. Action of alpha-amanitin during pyrophosphorolysis and elongation by RNA polymerase II. *J. Biol. Chem*. 270, 19114–19119 (1995).

Chammas, P., Mocavini, I. & Di Croce, L. Engaging chromatin: PRC2 structure meets function. *Br J Cancer* 122, 315–328 (2020).

Chan SH, Tang Y, Miao L, Darwich-Codore H, Vejnar CE, Beaudoin JD, Musaev D, Fernandez JP, Benitez MDJ, Bazzini AA, Moreno-Mateos MA, Giraldez AJ. Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. *Dev Cell*. 2019 Jun 17;49(6):867-881.e8.

Chang YL, Peng YH, Pan IC, Sun DS, King B, Huang DH. Essential role of *Drosophila* Hdac1 in homeotic gene silencing. *Proc Natl Acad Sci U S A*. 2001 Aug 14;98(17):9730-5.

Charney RM, Forouzmand E, Cho JS, Cheung J, Paraiso KD, Yasuoka Y, Takahashi S, Taira M, Blitz IL, Xie X, Cho KW. Foxh1 Occupies *cis*-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program. *Dev Cell*. 2017 Mar 27;40(6):595-607.e4.

Chen C, Shen MM. Two modes by which Lefty proteins inhibit nodal signaling. *Curr Biol*. 2004 Apr 6;14(7):618-24.

Chen K, Johnston J, Shao W, Meier S, Staber C, Zeitlinger J. A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *Elife*. 2013 Aug 13;2:e00861.

Chen T, Dent SY. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*. 2014 Feb;15(2):93-106.

Chen X, Ke Y, Wu K, Zhao H, Sun Y, Gao L, Liu Z, Zhang J, Tao W, Hou Z, Liu H, Liu J, Chen ZJ. Key role for CTCF in establishing chromatin structure in human embryos. *Nature*. 2019 Dec;576(7786):306-310.

Chen X, Weisberg E, Fridmacher V, Watanabe M, Naco G, Whitman M. Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature*. 1997 Sep 4;389(6646):85-9.

Chiu WT, Charney Le R, Blitz IL, Fish MB, Li Y, Biesinger J, Xie X, Cho KW. Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program. *Development*. 2014 Dec;141(23):4537-47.

Cho JS, Blitz IL, Cho KWY. DNase-seq to Study Chromatin Accessibility in Early *Xenopus tropicalis* Embryos. *Cold Spring Harb Protoc*. 2019 Apr 1;2019(4):pdb.prot098335.

Cho KW, Blumberg B, Steinbeisser H, De Robertis EM. Molecular nature of Spemann's organizer: the role of the *Xenopus* homeobox gene *goosecoid*. *Cell*. 1991 Dec 20;67(6):1111-20.

Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*. 2009 Aug 14;325(5942):834-40.

Conlon FL, Miteva Y, Kaltenbrun E, Waldron L, Greco TM, Cristea IM. Immunoisolation of protein complexes from *Xenopus*. *Methods Mol Biol*. 2012;917:369-90.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 2010 Dec 14;107(50):21931-6.

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell*. 2009 Jan 9;4(1):80-93.

Cunliffe VT. Histone deacetylase 1 is required to repress Notch target gene expression during zebrafish neurogenesis and to maintain the production of motoneurones in response to hedgehog signalling. *Development*. 2004 Jun;131(12):2983-95.

Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, Li G, Kuan S, Li B, Lee AY, Preissl S, Jermstad I, Haugen MH, Suganthan R, Bjørås M, Hansen K, Dalen KT, Fedorcsak P, Ren B, Klungland A. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*. 2016 Sep 22;537(7621):548-552.

Danilov V, Blum M, Schweickert A, Campione M, Steinbeisser H. Negative autoregulation of the organizer-specific homeobox gene *goosecoid*. *J Biol Chem*. 1998 Jan 2;273(1):627-35.

Dennis K, Fan T, Geiman T, Yan Q, Muegge K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev*. 2001 Nov 15;15(22):2940-4.

Desvoyes B, Sequeira-Mendes J, Vergara Z, Madeira S, Gutierrez C. Sequential ChIP Protocol for Profiling Bivalent Epigenetic Modifications (ReChIP). *Methods Mol Biol*. 2018;1675:83-97.

Dimitrov S, Almouzni G, Dasso M, Wolffe AP. Chromatin transitions during early *Xenopus* embryogenesis: changes in histone H4 acetylation and in linker histone type. *Dev Biol*. 1993 Nov;160(1):214-27.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013 Jan 1;29(1):15-21.

Dorighi KM, Swigut T, Henriques T, Bhanu NV, Scruggs BS, Nady N, Still CD 2nd, Garcia BA, Adelman K, Wysocka J. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell*. 2017 May 18;66(4):568-576.e4.

Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y, Ma J, Zhang X, Zhang K, Wang Y, Zhang MQ, Gao J, Dixon JR, Wang X, Zeng J, Xie W. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature*. 2017 Jul 12;547(7662):232-235.

Dunican DS, Ruzov A, Hackett JA, Meehan RR. xDnmt1 regulates transcriptional silencing in pre-MBT *Xenopus* embryos independently of its catalytic function. *Development*. 2008 Apr;135(7):1295-302.

Eckersley-Maslin MA, Alda-Catalinas C, Reik W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat Rev Mol Cell Biol*. 2018 Jul;19(7):436-450.

Egan B, Yuan CC, Craske ML, Labhart P, Guler GD, Arnott D, Maile TM, Busby J, Henry C, Kelly TK, et al. An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition. *PLoS One*. 2016 Nov 22;11(11):e0166438.

Elgin SC. The formation and function of DNase I hypersensitive sites in the process of gene activation. *J Biol Chem*. 1988 Dec 25;263(36):19259-62.

ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020 Jul;583(7818):699-710.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57-74.

Esmaeili M, Blythe SA, Tobias JW, Zhang K, Yang J, Klein PS. Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Dev Biol*. 2020 Jun 1;462(1):20-35.

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A*. 2010 May 11;107(19):8689-94.

Filippakopoulos P, Picaud S, Mangos M, Keates T, Lambert JP, Barsyte-Lovejoy D, Felletar I, Volkmer R, Müller S, Pawson T, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*. 2012 Mar 30;149(1):214-31

Fortriede JD, Pells TJ, Chu S, Chaturvedi P, Wang D, Fisher ME, James-Zorn C, Wang Y, Nenni MJ, Burns KA, et al. Xenbase: deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database. *Nucleic Acids Res*. 2020 Jan 8;48(D1):D776-D782.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*. 2016 May 31;15(9):2038-49.

Furey TS. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*. 2012 Dec;13(12):840-52.

Fyodorov DV, Zhou BR, Skoultchi AI, Bai Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol*. 2018 Mar;19(3):192-206.

Gao L, Zhu X, Chen G, Ma X, Zhang Y, Khand AA, Shi H, Gu F, Lin H, Chen Y, Zhang H, He L, Tao Q. A novel role for Ascl1 in the regulation of mesendoderm formation via HDAC-dependent antagonism of VegT. *Development*. 2016 Feb 1;143(3):492-503.

Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet*. 2006 Sep;7(9):703-13.

Gaydos LJ, Wang W, Strome S. Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science*. 2014 Sep 19;345(6203):1515-8.

Gentsch GE, Owens ND, Martin SR, Piccinelli P, Faial T, Trotter MW, Gilchrist MJ, Smith JC. *In vivo* T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency. *Cell Rep*. 2013 Sep 26;4(6):1185-96.

Gentsch GE, Spruce T, Owens NDL, Smith JC. Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nat Commun*. 2019 Sep 19;10(1):4269.

Germain S, Howell M, Esslemont GM, Hill CS. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. *Genes Dev*. 2000 Feb 15;14(4):435-51.

Gillies SD, Morrison SL, Oi VT, Tonegawa S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*. 1983 Jul;33(3):717-28.

Göttlicher M, Minucci S, Zhu P, Krämer OH, Schimpf A, Giavara S, Sleeman JP, Lo Coco F, Nervi C, Pelicci PG, Heinzel T. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J*. 2001 Dec 17;20(24):6969-78.

Gregoretti IV, Lee YM, Goodson HV. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J Mol Biol*. 2004 Apr 16;338(1):17-31.

Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem*. 1988;57:159-97.

Gryder BE, Wu L, Woldemichael GM, Pomella S, Quinn TR, Park PMC, Cleveland A, Stanton BZ, Song Y, Rota R, et al. Chemical genomics reveals histone deacetylases are required for core regulatory transcription. *Nat Commun*. 2019 Jul 8;10(1):3004.

Gupta R, Wills A, Ucar D, Baker J. Developmental enhancers are marked independently of zygotic Nodal signals in *Xenopus*. *Dev Biol*. 2014 Nov 1;395(1):38-49.

Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forné T. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*. 2007;2(7):1722-33.

Hake SB, Allis CD. Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc Natl Acad Sci U S A*. 2006 Apr 25;103(17):6428-35.

Hardison RC, Taylor J. Genomic approaches towards finding *cis*-regulatory modules in animals. *Nat Rev Genet*. 2012 Jun 18;13(7):469-83.

Hassan AH, Awad S, Al-Natour Z, Othman S, Mustafa F, Rizvi TA. Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochem J*. 2007 Feb 15;402(1):125-33.

Heasman J. Patterning the early Xenopus embryo. *Development*. 2006 Apr;133(7):1205-17.
Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, Zhang MQ, Snyder MP. Genome-wide map of regulatory interactions in the human genome. *Genome Res*. 2014 Dec;24(12):1905-17.

Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, Zhang MQ, Snyder MP. Genome-wide map of regulatory interactions in the human genome. *Genome Res*. 2014 Dec;24(12):1905-17.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 2009 May 7;459(7243):108-12.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007 Mar;39(3):311-8.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010 May 28;38(4):576-89.

Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, Marazzi I, García-Sastre A, Shaw ML, Benner C. Transcription Elongation Can Affect Genome 3D Structure. *Cell*. 2018 Sep 6;174(6):1522-1536.e22.

Henikoff S. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet*. 2008 Jan;9(1):15-26.

Heo DH, Kuś K, Grzechnik P, Tan-Wong SM, Birot A, Kecman T, Nielsen S, Zenkin N, Vasiljeva L. Transcription and chromatin-based surveillance mechanism controls suppression of cryptic antisense transcription. *Cell Rep*. 2021 Sep 7;36(10):109671.

Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes Dev*. 2012 Dec 1;26(23):2604-20.

Hill CS. TGF-beta signalling pathways in early *Xenopus* development. *Curr Opin Genet Dev*. 2001 Oct;11(5):533-40.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009 Jun 9;106(23):9362-7.

Hontelez S, van Kruijsbergen I, Georgiou G, van Heeringen SJ, Bogdanovic O, Lister R, Veenstra GJC. Embryonic transcription is controlled by maternally defined chromatin state. *Nat Commun*. 2015 Dec 18;6:10148.

Howell M, Inman GJ, Hill CS. A novel Xenopus Smad-interacting forkhead transcription factor (XFast-3) cooperates with XFast-1 in regulating gastrulation movements. *Development*. 2002 Jun;129(12):2823-34.

Hrit J, Goodrich L, Li C, Wang BA, Nie J, Cui X, Martin EA, Simental E, Fernandez J, Liu MY, Nery JR, Castanon R, Kohli RM, Tretyakova N, He C, Ecker JR, Goll M, Panning B. OGT binds a conserved C-terminal domain of TET1 to regulate TET1 activity and function in development. *Elife*. 2018 Oct 16;7:e34870.

Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*. 2017 Apr 6;169(2):216-228.e19.

Hughes MW, Jiang TX, Lin SJ, Leung Y, Kobielak K, Widelitz RB, Chuong CM. Disrupted ectodermal organ morphogenesis in mice with a conditional histone deacetylase 1, 2 deletion in the epidermis. *J Invest Dermatol*. 2014 Jan;134(1):24-32.

Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. *Exp Mol Med*. 2017 Apr 28;49(4):e324.

Inoue A, Fujimoto D. Enzymatic deacetylation of histone. *Biochem Biophys Res Commun*. 1969 Jul 7;36(1):146-50.

Inoue A, Jiang L, Lu F, Suzuki T, Zhang Y. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature*. 2017 Jul 27;547(7664):419-424.

Iratni R, Yan YT, Chen C, Ding J, Zhang Y, Price SM, Reinberg D, Shen MM. Inhibition of excess nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. *Science*. 2002 Dec 6;298(5600):1996-9.

Jansen C, Paraiso KD, Zhou JJ, Blitz IL, Fish MB, Charney RM, Cho JS, Yasuoka Y, Sudou N, Bright AR, Wlizla M, Veenstra GJC, Taira M, Zorn AM, Mortazavi A, Cho KWY. Uncovering the mesendoderm gene regulatory network through multi-omic data integration. *Cell Rep*. 2022 Feb 15;38(7):110364.

Ji X, Dadon DB, Abraham BJ, Lee TI, Jaenisch R, Bradner JE, Young RA. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc Natl Acad Sci U S A*. 2015 Mar 24;112(12):3841-6.

Jonas EA, Snape AM, Sargent TD. Transcriptional regulation of a *Xenopus* embryonic epidermal keratin gene. *Development*. 1989 Jun;106(2):399-405.

Jukam D, Shariati SAM, Skotheim JM. Zygotic Genome Activation in Vertebrates. *Dev Cell*. 2017 Aug 21;42(4):316-332.

Kaaij LJT, van der Weide RH, Ketting RF, de Wit E. Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. *Cell Rep*. 2018 Jul 3;24(1):1-10.e4.

Kakebeen AD, Chitsazan AD, Williams MC, Saunders LM, Wills AE. Chromatin accessibility dynamics and single cell RNA-Seq reveal new regulators of regeneration in neural progenitors. *Elife*. 2020 Apr 27;9:e52648.

Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, Yao X, Li F, Zhu W, Gao L, Chen H, Du Z, Xie W, Xu X, Huang X, Liu J. 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell*. 2017 Jul 13;170(2):367-381.e20.

Keller G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev*. 2005 May 15;19(10):1129-55.

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*. 2005 Nov 18;123(4):593-605.

Khochbin S. Histone H1 diversity: bridging regulatory signals to linker histone function. *Gene*. 2001 Jun 13;271(1):1-12.

Kidder BL, Palmer S. HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. *Nucleic Acids Res*. 2012 Apr;40(7):2925-39.

Kim T, Buratowski S. Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell*. 2009 Apr 17;137(2):259-72.

Kim T, Xu Z, Clauder-Münster S, Steinmetz LM, Buratowski S. Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. *Cell*. 2012 Sep 14;150(6):1158-69

Koenecke N, Johnston J, He Q, Meier S, Zeitlinger J. *Drosophila* poised enhancers are generated during tissue patterning with the help of repression. *Genome Res*. 2017 Jan;27(1):64-74.

Kofron M, Demel T, Xanthos J, Lohr J, Sun B, Sive H, Osada S, Wright C, Wylie C, Heasman J. Mesoderm induction in *Xenopus* is a zygotic event regulated by maternal VegT via TGFbeta growth factors. *Development*. 1999 Dec;126(24):5759-70.

Koide T, Hayata T, Cho KW. *Xenopus* as a model system to study transcriptional regulatory networks. *Proc Natl Acad Sci U S A*. 2005 Apr 5;102(14):4943-8.

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet*. 2009 Mar;41(3):376-81.

Kundu S, Ji F, Sunwoo H, Jain G, Lee JT, Sadreyev RI, Dekker J, Kingston RE. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell*. 2017 Feb 2;65(3):432-446.e5.

Kurdistani SK, Robyr D, Tavazoie S, Grunstein M. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat Genet*. 2002 Jul;31(3):248-54.

Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature*. 2014 Aug 7;512(7512):91-5.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012 Mar 4;9(4):357-9.

Lasko LM, Jakob CG, Edalji RP, Qiu W, Montgomery D, Digiammarino EL, Hansen TM, Risi RM, Frey R, Manaves V, Shaw B, Algire M, Hessler P, Lam LT, Uziel T, Faivre E, Ferguson D, Buchanan FG, Martin RL, Torrent M, Chiang GG, Karukurichi K, Langston JW, Weinert BT, Choudhary C, de Vries P, Van Drie JH, McElligott D, Kesicki E, Marmorstein R, Sun C, Cole PA, Rosenberg SH, Michaelides MR, Lai A, Bromberg KD. Discovery of a selective catalytic

p300/CBP inhibitor that targets lineage-specific tumours. *Nature*. 2017 Oct 5;550(7674):128-132.

Laurent MN, Blitz IL, Hashimoto C, Rothbächer U, Cho KW. The *Xenopus* homeobox gene twin mediates Wnt induction of *goosecoid* in establishment of Spemann's organizer. *Development*. 1997 Dec;124(23):4905-16.

Le TB, Laub MT. Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *EMBO J*. 2016 Jul 15;35(14):1582-95.

Lee BB, Choi A, Kim JH, Jun Y, Woo H, Ha SD, Yoon CY, Hwang JT, Steinmetz L, Buratowski S, Lee S, Kim HY, Kim T. Rpd3L HDAC links H3K4me3 to transcriptional repression memory. *Nucleic Acids Res*. 2018 Sep 19;46(16):8261-8274.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NA, Margaritis T, van de Pasch LA, van Heesch SA, Brok MO, Groot Koerkamp MJ, et al. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell*. 2011 May 20;42(4):536-49.

Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010 Sep 14;20(17):R754-63.

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011 Aug 4;12:323.

Li B, Gogol M, Carey M, Lee D, Seidel C, Workman JL. Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science*. 2007 May 18;316(5827):1050-4.

Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 1992 Jun 12;69(6):915-26.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9.

Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *Elife*. 2014 Oct 14;3:e03737.

Li Q, Barkess G, Qian H. Chromatin looping and the probability of transcription. *Trends Genet.* 2006 Apr;22(4):197-202.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat*. 5, 1752–1779.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M,

Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009 Oct 9;326(5950):289-93.

Lin X, Duan X, Liang YY, Su Y, Wrighton KH, Long J, Hu M, Davis CM, Wang J, Brunicardi FC, Shi Y, Chen YG, Meng A, Feng XH. PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. *Cell*. 2006 Jun 2;125(5):915-28.

Lindeman LC, Andersen IS, Reiner AH, Li N, Aanes H, Østrup O, Winata C, Mathavan S, Müller F, Aleström P, Collas P. Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev Cell*. 2011 Dec 13;21(6):993-1004.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009 Nov 19;462(7271):315-22.

Liu G, Wang W, Hu S, Wang X, Zhang Y. Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. *Genome Res*. 2018 Jul;28(7):998-1007.

Liu Y, Yu S, Dhiman VK, Brunetti T, Eckart H, White KP. Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol*. 2017 Nov 20;18(1):219.

Local A, Huang H, Albuquerque CP, Singh N, Lee AY, Wang W, Wang C, Hsia JE, Shiau AK, Ge K, Corbett KD, Wang D, Zhou H, Ren B. Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat Genet*. 2018 Jan;50(1):73-82.

Longo PA, Kavran JM, Kim MS, Leahy DJ. Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol*. 2013;529:227-40.

Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).

Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*. 2013 Apr 11;153(2):320-34.

Lu F, Liu Y, Inoue A, Suzuki T, Zhao K, Zhang Y. Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell*. 2016 Jun 2;165(6):1375-1388.

Lucio-Eterovic AK, Singh MM, Gardner JE, Veerappan CS, Rice JC, Carpenter PB. Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc Natl Acad Sci U S A*. 2010 Sep 28;107(39):16952-7.

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell*. 2008 Mar 21;132(6):958-70.

Lustig KD, Kroll KL, Sun EE, Kirschner MW. Expression cloning of a *Xenopus* T-related gene (Xombi) involved in mesodermal patterning and blastopore lip formation. *Development*. 1996 Dec;122(12):4001-12.

Lynch MD, Smith AJ, De Gobbi M, Flenley M, Hughes JR, Vernimmen D, Ayyub H, Sharpe JA, Sloane-Stanley JA, Sutherland L, Meek S, Burdon T, Gibbons RJ, Garrick D, Higgs DR. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J*. 2012 Jan 18;31(2):317-29.

Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, Etchin J, Lawton L, Sallan SE, Silverman LB, Loh ML, Hunger SP, Sanda T, Young RA, Look AT. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*. 2014 Dec 12;346(6215):1373-7.

Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011 Jan 20;469(7330):343-9.

Martin, B.J.E, Brind'Amour, J, Kuzmin, A, Jensen K.N, Liu Z.C, Lorincz M, Howe L.J. Transcription shapes genome-wide histone acetylation patterns. *Nat Commun* 12, 210 (2021).

Massagué, J. TGFβ signalling in context. *Nat Rev Mol Cell Biol* 13, 616–630 (2012).

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010 Jul 8;466(7303):253-7.

Meganathan K, Jagtap S, Srinivasan SP, Wagh V, Hescheler J, Hengstler J, Leist M, Sachinidis A. Neuronal developmental gene and miRNA signatures induced by histone deacetylase inhibitors in human embryonic stem cells. *Cell Death Dis*. 2015 May 7;6(5):e1756.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG Jr, Kinney JB, Kellis M, Lander ES, Mikkelsen TS. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol*. 2012 Feb 26;30(3):271-7.

Methot SP, Padeken J, Brancati G, Zeller P, Delaney CE, Gaidatzis D, Kohler H, van Oudenaarden A, Großhans H, Gasser SM. H3K9me selectively blocks transcription factor activity and ensures differentiated tissue integrity. *Nat Cell Biol*. 2021 Nov;23(11):1163-1175.

Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, Mushayamaha T, Thomas PD. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D394-D403.

Milazzo G, Mercatelli D, Di Muzio G, Triboli L, De Rosa P, Perini G, Giorgi FM. Histone Deacetylases (HDACs): Evolution, Specificity, Role in Transcriptional Complexes, and Pharmacological Actionability. *Genes (Basel).* 2020 May 15;11(5):556.

Milstone ZJ, Saheera S, Bourke LM, Shpilka T, Haynes CM, Trivedi CM. Histone deacetylases 1 and 2 silence cryptic transcription to promote mitochondrial function during cardiogenesis. *Sci Adv*. 2020 Apr 10;6(15):eaax5150.

Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology*. 2013 Jan;38(1):23-38.

Moorthy SD, Davidson S, Shchuka VM, Singh G, Malek-Gilani N, Langroudi L, Martchenko A, So V, Macpherson NN, Mitchell JA. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res*. 2017 Feb;27(2):246-258.

Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res*. 1981 Nov 25;9(22):6047-68.

Morgunova E, Taipale J. Structural perspective of cooperative transcription factor binding. *Curr Opin Struct Biol.* 2017 Dec;47:1-8.

Moshe A, Kaplan T. Genome-wide search for Zelda-like chromatin signatures identifies GAF as a pioneer factor in early fly development. *Epigenetics Chromatin*. 2017 Jul 4;10(1):33.

Neumayr C, Pagani M, Stark A, Arnold CD. STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. *Curr Protoc Mol Biol*. 2019 Sep;128(1):e105.

Ng RK, Gurdon JB. Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nat Cell Biol*. 2008 Jan;10(1):102-9.

Nieuwkoop, P.D., and Faber, J. (1994). Normal Table of *Xenopus laevis* (Daudin) (*Garland Publishing Inc*), ISBN 0-8153-1896-0.

Niu L, Shen W, Shi Z, Tan Y, He N, Wan J, Sun J, Zhang Y, Huang Y, Wang W, Fang C, Li J, Zheng P, Cheung E, Chen Y, Li L, Hou C. Three-dimensional folding dynamics of the *Xenopus tropicalis* genome. *Nat Genet*. 2021 Jul;53(7):1075-1087.

Ogino H, McConnell WB, Grainger RM. Highly efficient transgenesis in *Xenopus tropicalis* using I-SceI meganuclease. *Mech Dev*. 2006 Feb;123(2):103-13.

Ogiyama Y, Schuettengruber B, Papadopoulos GL, Chang JM, Cavalli G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Mol Cell*. 2018 Jul 5;71(1):73-88.e5.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell*. 1999 Oct 29;99(3):247-57.

Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014 Apr;15(4):234-46.

Onichtchouk D, Chen YG, Dosch R, Gawantka V, Delius H, Massagué J, Niehrs C. Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. *Nature*. 1999 Sep 30;401(6752):480-5.

Oppikofer M, Bai T, Gan Y, Haley B, Liu P, Sandoval W, Ciferri C, Cochran AG. Expansion of the ISWI chromatin remodeler family with new active complexes. *EMBO Rep*. 2017 Oct;18(10):1697-1706.

Osada SI, Saijoh Y, Frisch A, Yeo CY, Adachi H, Watanabe M, Whitman M, Hamada H, Wright CV. Activin/nodal responsiveness and asymmetric expression of a *Xenopus* nodal-related gene converge on a FAST-regulated module in intron 1. *Development*. 2000 Jun;127(11):2503-14.

Owens NDL, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KWY, Khokha MK. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep*. 2016 Jan 26;14(3):632-647.

Pan MR, Hsu MC, Chen LT, Hung WC. Orchestration of H3K27 methylation: mechanisms and therapeutic implication. *Cell Mol Life Sci*. 2018 Jan;75(2):209-223.

Paraiso KD, Blitz IL, Cho KWY. Maternal and zygotic factors sequentially shape the tissue regionalization of chromatin landscapes in early vertebrate embryos. *BioRxiv* 2021.04.14.439777; doi: https://doi.org/10.1101/2021.04.14.439777

Paraiso KD, Blitz IL, Coley M, Cheung J, Sudou N, Taira M, Cho KWY. Endodermal Maternal Transcription Factors Establish Super-Enhancers during Zygotic Genome Activation. *Cell Rep*. 2019 Jun 4;27(10):2962-2977.e5.

Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol*. 2009 Dec;27(12):1173-5.

Penzel R, Oschwald R, Chen Y, Tacke L, Grunz H. Characterization and early embryonic expression of a neural specific transcription factor xSOX3 in *Xenopus laevis*. *Int J Dev Biol*. 1997 Oct;41(5):667-77.

Petrykowska HM, Vockley CM, Elnitski L. Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. *Genome Res*. 2008 Aug;18(8):1238-46.

Pombo A, Dillon N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol*. 2015 Apr;16(4):245-57.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*. 2010 Feb 25;463(7284):1101-5.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010 Mar 15;26(6):841-2.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011 Feb 10;470(7333):279-83.

Radman-Livaja M, Rando OJ. Nucleosome positioning: how is it established, and why does it matter? *Dev Biol*. 2010 Mar 15;339(2):258-66.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014 Jul;42(Web Server issue):W187-91.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A*. 2000 May 9;97(10):5237-42.

Rao A, LaBonne C. Histone deacetylase activity has an essential role in establishing and maintaining the vertebrate neural crest. *Development*. 2018 Aug 8;145(15):dev163386.

Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*. 2010 Mar 5;140(5):744-52.

Reid CD, Steiner AB, Yaklichkin S, Lu Q, Wang S, Hennessy M, Kessler DS. FoxH1 mediates a Grg4 and Smad2 dependent transcriptional switch in Nodal signaling during *Xenopus* mesoderm development. *Dev Biol*. 2016 Jun 1;414(1):34-44.

Reid G, Métivier R, Lin CY, Denger S, Ibberson D, Ivacevic T, Brand H, Benes V, Liu ET, Gannon F. Multiple mechanisms induce transcriptional silencing of a subset of genes, including oestrogen receptor alpha, in response to deacetylase inhibition by valproic acid and trichostatin A. *Oncogene*. 2005 Jul 21;24(31):4894-907.

Reiter F, Wienerroither S, Stark A. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev*. 2017 Apr;43:73-81.

Richards EJ, Elgin SC. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell*. 2002 Feb 22;108(4):489-500.

Rickels R, Herz HM, Sze CC, Cao K, Morgan MA, Collings CK, Gause M, Takahashi YH, Wang L, Rendleman EJ, Marshall SA, Krueger A, Bartom ET, Piunti A, Smith ER, Abshiru NA, Kelleher NL, Dorsett D, Shilatifard A. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet*. 2017 Nov;49(11):1647-1653.

Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D, et al. Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res*. 2011 Feb;21(2):147-63.

Ring C, Ogata S, Meek L, Song J, Ohta T, Miyazono K, Cho KW. The role of a Williams-Beuren syndrome-associated helix-loop-helix domain-containing transcription factor in activin/nodal signaling. *Genes Dev*. 2002 Apr 1;16(7):820-35.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011 Jan;29(1):24-6.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010 Jan 1;26(1):139-40.

Roh TY, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev*. 2005 Mar 1;19(5):542-52.

Schier AF. Nodal signaling in vertebrate development. *Annu Rev Cell Dev Biol*. 2003;19:589-621.

Schöne S, Bothe M, Einfeldt E, Borschiwer M, Benner P, Vingron M, Thomas-Chollier M, Meijsing SH. Synthetic STARR-seq reveals how DNA shape and sequence modulate transcriptional output and noise. *PLoS Genet*. 2018 Nov 14;14(11):e1007793.

Schübeler D. Function and information content of DNA methylation. *Nature*. 2015 Jan 15;517(7534):321-6.

Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. *Nat Rev Genet*. 2019 Apr;20(4):221-234.

Seto E, Yoshida M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol*. 2014 Apr 1;6(4):a018713.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*. 2012 Feb 3;148(3):458-72.

Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. *Nat Protoc*. 2007;2(6):1445-57.

Shechter D, Nicklay JJ, Chitta RK, Shabanowitz J, Hunt DF, Allis CD. Analysis of histones in *Xenopus laevis*. I. A distinct index of enriched variants and modifications exists in each cell type and is remodeled during developmental transitions. *J Biol Chem*. 2009 Jan 9;284(2):1064-74.

Shen MM. Nodal signaling: developmental roles and regulation. *Development*. 2007 Mar;134(6):1023-34.

Simone C, Stiegler P, Forcales SV, Bagella L, De Luca A, Sartorelli V, Giordano A, Puri PL. Deacetylase recruitment by the C/H3 domain of the acetyltransferase p300. *Oncogene*. 2004 Mar 18;23(12):2177-87.

Sitbon D, Boyarchuk E, Dingli F, Loew D, Almouzni G. Histone variant H3.3 residue S31 is essential for *Xenopus* gastrulation regardless of the deposition pathway. *Nat Commun*. 2020 Mar 9;11(1):1256.

Skvortsova K, Iovino N, Bogdanović O. Functions and mechanisms of epigenetic inheritance in animals. *Nat Rev Mol Cell Biol*. 2018 Dec;19(12):774-790.

Slaughter MJ, Shanle EK, Khan A, Chua KF, Hong T, Boxer LD, Allis CD, Josefowicz SZ, Garcia BA, Rothbart SB, et al. HDAC inhibition results in widespread alteration of the histone acetylation landscape and BRD4 targeting to gene bodies. *Cell Rep*. 2021 Jan 19;34(3):108638.

Smith RC, Dworkin-Rastl E, Dworkin MB. Expression of a histone H1-like protein is restricted to early *Xenopus* development. *Genes Dev*. 1988 Oct;2(10):1284-95.

Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012 Sep;13(9):613-26.

Stancheva I, Meehan RR. Transient depletion of xDnmt1 leads to premature gene activation in *Xenopus* embryos. *Genes Dev*. 2000 Feb 1;14(3):313-27.

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

Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000 Jan 6;403(6765):41-5.

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*. 2012 May;139(9):1651-61.

Szenker E, Ray-Gallet D, Almouzni G. The double face of the histone variant H3.3. *Cell Res*. 2011 Mar;21(3):421-34.

Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. *Development*. 2009 Sep;136(18):3033-42.

Taira M, Jamrich M, Good PJ, Dawid IB. The LIM domain-containing homeo box gene Xlim-1 is expressed specifically in the organizer region of *Xenopus* gastrula embryos. *Genes Dev*. 1992 Mar;6(3):356-66.

Tavares AT, Andrade S, Silva AC, Belo JA. Cerberus is a feedback inhibitor of Nodal asymmetric signaling in the chick embryo. *Development*. 2007 Jun;134(11):2051-60.

Tettey TT, Gao X, Shao W, Li H, Story BA, Chitsazan AD, Glaser RL, Goode ZH, Seidel CW, Conaway RC, et al. A Role for FACT in RNA Polymerase II Promoter-Proximal Pausing. *Cell Rep*. 2019 Jun 25;27(13):3770-3779.e7.

Trojer P, Reinberg D. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell*. 2007 Oct 12;28(1):1-13.

Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*. 2014 Nov 20;7(1):33.

Ura K, Nightingale K, Wolffe AP. Differential association of HMG1 and linker histones B4 and H1 with dinucleosomal DNA: structural transitions and transcriptional repression. *EMBO J*. 1996 Sep 16;15(18):4959-69.

Vallot A, Tachibana K. The emergence of genome architecture and zygotic genome activation. *Curr Opin Cell Biol*. 2020 Jun;64:50-57.

van Heeringen SJ, Akkers RC, van Kruijsbergen I, Arif MA, Hanssen LL, Sharifi N, Veenstra GJ. Principles of nucleation of H3K27 methylation during embryonic development. *Genome Res*. 2014 Mar;24(3):401-10.

van Kruijsbergen I, Hontelez S, Elurbe DM, van Heeringen SJ, Huynen MA, Veenstra GJC. Heterochromatic histone modifications at transposons in *Xenopus tropicalis* embryos. *Dev Biol*. 2017 Jun 15;426(2):460-471.

Vardimon L, Kressmann A, Cedar H, Maechler M, Doerfler W. Expression of a cloned adenovirus gene is inhibited by *in vitro* methylation. *Proc Natl Acad Sci U S A*. 1982 Feb;79(4):1073-7.

Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. *Development*. 2019 Jun 12;146(11):dev161471.

Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, Rinn J, Schier AF. Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*. 2010 Apr 8;464(7290):922-6.

Veenstra GJ, Wolffe AP. Constitutive genomic methylation during embryonic development of *Xenopus*. *Biochim Biophys Acta*. 2001 Oct 31;1521(1-3):39-44.

Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol*. 2015 Mar;16(3):178-89.

Vidal M, Gaber RF. RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1991 Dec;11(12):6317-27.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*. 2015 Mar 3;10(8):1297-309.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009 Feb 12;457(7231):854-8.

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodeling and histone deacetylation. *Nat Genet*. 1999 Sep;23(1):62-6.

Wang A, Kurdistani SK, Grunstein M. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science*. 2002 Nov 15;298(5597):1412-4.

Wang C (a), Liu X, Gao Y, Yang L, Li C, Liu W, Chen C, Kou X, Zhao Y, Chen J, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol*. 2018 May;20(5):620-631.

Wang CI, Alekseyenko AA, LeRoy G, Elia AE, Gorchakov AA, Britton LM, Elledge SJ, Kharchenko PV, Garcia BA, Kuroda MI. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat Struct Mol Biol*. 2013 Feb;20(2):202-9.

Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 2016 Aug 5;353(6299):598-602.

Wang W, Li X, Huang J, Feng L, Dolinta KG, Chen J. Defining the protein-protein interaction network of the human hippo pathway. *Mol Cell Proteomics*. 2014 Jan;13(1):119-31.

Wang X (b), He L, Goggin SM, Saadat A, Wang L, Sinnott-Armstrong N, Claussnitzer M, Kellis M. High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. *Nat Commun*. 2018 Dec 19;9(1):5380.

Wang X, Moore SC, Laszckzak M, Ausió J. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *J Biol Chem*. 2000 Nov 10;275(45):35013-20.

Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*. 2009 Sep 4;138(5):1019-31.

Watabe T, Kim S, Candia A, Rothbächer U, Hashimoto C, Inoue K, Cho KW. Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between Xenopus and mouse. *Genes Dev*. 1995 Dec 15;9(24):3038-50.

Watanabe M, Whitman M. FAST-1 is a key maternal effector of mesoderm inducers in the early Xenopus embryo. *Development*. 1999 Dec;126(24):5621-34.

Weber CM, Ramachandran S, Henikoff S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell*. 2014 Mar 6;53(5):819-30.

Whitman M. Nodal signaling in early vertebrate embryos: themes and variations. *Dev Cell*. 2001 Nov;1(5):605-17.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013 Apr 11;153(2):307-19.

Winklbauer R. Mesoderm and endoderm internalization in the *Xenopus* gastrula. *Curr Top Dev Biol*. 2020;136:243-270.

Wolffe AP, Hayes JJ. Chromatin disruption and modification. *Nucleic Acids Res*. 1999 Feb 1;27(3):711-20.

Wu E, Vastenhouw NL. From mother to embryo: A molecular perspective on zygotic genome activation. *Curr Top Dev Biol*. 2020;140:209-254.

Wu MY, Hill CS. Tgf-beta superfamily signaling in embryonic development and homeostasis. *Dev Cell*. 2009 Mar;16(3):329-43.

Xanthos JB, Kofron M, Wylie C, Heasman J. Maternal VegT is the initiator of a molecular network specifying endoderm in *Xenopus laevis*. *Development*. 2001 Jan;128(2):167-80.

Xia W, Xu J, Yu G, Yao G, Xu K, Ma X, Zhang N, Liu B, Li T, Lin Z, Chen X, Li L, Wang Q, Shi D, Shi S, Zhang Y, Song W, Jin H, Hu L, Bu Z, Wang Y, Na J, Xie W, Sun YP. Resetting

histone modifications during human parental-to-zygotic transition. *Science*. 2019 Jul 26;365(6451):353-360.

Yamaguchi M, Tonou-Fujimori N, Komori A, Maeda R, Nojima Y, Li H, Okamoto H, Masai I. Histone deacetylase 1 regulates retinal neurogenesis in zebrafish by suppressing Wnt and Notch signaling pathways. *Development*. 2005 Jul;132(13):3027-43.

Yartseva V, Giraldez AJ. The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. *Curr Top Dev Biol*. 2015;113:191-232.

Yasuoka Y, Suzuki Y, Takahashi S, Someya H, Sudou N, Haramoto Y, Cho KW, Asashima M, Sugano S, Taira M. Occupancy of tissue-specific *cis*-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification. *Nat Commun*. 2014 Jul 9;5:4322.

Yeo CY, Chen X, Whitman M. The role of FAST-1 and Smads in transcriptional regulation by activin during early *Xenopus* embryogenesis. *J Biol Chem*. 1999 Sep 10;274(37):26584-90.

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schübeler D, Vinson C, Taipale J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017 May 5;356(6337):eaaj2239.

Yoshida M, Kijima M, Akita M, Beppu T. Potent and specific inhibition of mammalian histone deacetylase both *in vivo* and *in vitro* by trichostatin A. *J Biol Chem*. 1990 Oct 5;265(28):17174-9.

Zaitseva I, Zaitsev S, Alenina N, Bader M, Krivokharchenko A. Dynamics of DNA-demethylation in early mouse and rat embryos developed *in vivo* and *in vitro*. *Mol Reprod Dev*. 2007 Oct;74(10):1255-61.

Zaret KS. Pioneer Transcription Factors Initiating Gene Network Changes. *Annu Rev Genet*. 2020 Nov 23;54:367-385.

Zenk F, Loeser E, Schiavo R, Kilpert F, Bogdanović O, Iovino N. Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science*. 2017 Jul 14;357(6347):212-216.

Zhang B, Zheng H, Huang B, Li W, Xiang Y, Peng X, Ming J, Wu X, Zhang Y, Xu Q, Liu W, Kou X, Zhao Y, He W, Li C, Chen B, Li Y, Wang Q, Ma J, Yin Q, Kee K, Meng A, Gao S, Xu F, Na J, Xie W. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature*. 2016 Sep 22;537(7621):553-557.

Zhang J, Houston DW, King ML, Payne C, Wylie C, Heasman J. The role of maternal VegT in establishing the primary germ layers in *Xenopus* embryos. *Cell*. 1998 Aug 21;94(4):515-24.

Zhang J, King ML. Xenopus VegT RNA is localized to the vegetal cortex during oogenesis and encodes a novel T-box transcription factor involved in mesodermal patterning. *Development*. 1996 Dec;122(12):4119-29.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.

Zhang Y, Vastenhouw NL, Feng J, Fu K, Wang C, Ge Y, Pauli A, van Hummelen P, Schier AF, Liu XS. Canonical nucleosome organization at promoters forms during genome activation. *Genome Res*. 2014 Feb;24(2):260-6.

Zheng H, Huang B, Zhang B, Xiang Y, Du Z, Xu Q, Li Y, Wang Q, Ma J, Peng X, Xu F, Xie W. Resetting Epigenetic Memory by Reprogramming of Histone Modifications in Mammals. *Mol Cell*. 2016 Sep 15;63(6):1066-79.

Zhou S, Zawel L, Lengauer C, Kinzler KW, Vogelstein B. Characterization of human FAST-1, a TGF beta and activin signal transducer. *Mol Cell*. 1998 Jul;2(1):121-7.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019 Apr 3;10(1):1523.

Zupkovitz G, Tischler J, Posch M, Sadzak I, Ramsauer K, Egger G, Grausenburger R, Schweifer N, Chiocca S, Decker T, Seiser C. Negative and positive regulation of gene expression by mouse histone deacetylase 1. *Mol Cell Biol*. 2006 Nov;26(21):7913-28.