Elucidating the role of chromatin organization during cell fate specification and vertebrate organogenesis

A Thesis

Submitted in partial fulfillment of the requirements

Of the degree of

Doctor of Philosophy

By

Saurabh Jagdish Pradhan
20173518

Under the guidance of

Prof. Sanjeev Galande and Prof. Carl-Philipp Heisenberg

Indian Institute of Science Education and Research Pune
February 2021

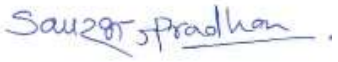
यशवंत सुत प्रार्थितो आई ...

Dedicated to four strong pillars of my life...

My Mom, Dad, sister
& almighty god

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have not been properly cited or from whom proper permission has not been taken when needed.

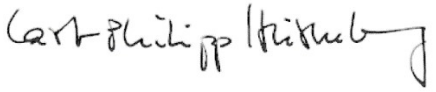


Date: 11.02.2020

Saurabh Jagdish Pradhan
ID: 20173518

CERTIFICATE

Certified that the work incorporated in the thesis entitled, "**Elucidating the role of chromatin organization during cell fate specification and vertebrate organogenesis**" submitted by **Saurabh Jagdish Pradhan** was carried out by the candidate under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.



Prof. Sanjeev Galande	Prof. Carl-Philipp Heisenberg
Supervisor	Co-Supervisor
IISER Pune	IST Austria

Date: 11th February 2021

Acknowledgements

On the completion of the most exciting and memorable journey of my life, I would like to express my love and sincere gratitude towards everyone who has been a big part of it. It's been a truly amazing one. I started taking baby steps towards a scientific journey in my undergraduate studies, but I feel extremely fortunate to join IISER-Pune initially as a research fellow and then as a PhD student. This work would not have taken its current shape without the support from everyone present on this beautiful IISER campus.

To begin with, I would like to express my sincere gratitude towards my supervisor, Prof. Sanjeev Galande. I still remember when I met him first on the stairs of the Sai Trinity campus and asked him to allow me to explore his area of research. He has been always encouraging and supportive to me since that day. He kept his faith in my work, and this allowed me to convert my thoughts into reality. I cannot thank him enough for his constant encouragement and guidance. During my initial days as a JRF, he always treated me at par with his PhD students. This was my biggest motivation and encouragement to continue in science and pursue doctoral studies. He allowed me to start exploring a new model system in his lab, zebrafish, to answer key questions in evo-devo. The freedom I got from him to design and execute my experiments and learn from failures allowed me to evolve as a researcher. He always supported the idea of discussing my research work with peers and establishing new collaborations. This element was a crucial part of his mentoring. Since this provided me with unique

opportunities to learn from experts in their respective fields. Thank you, Sanjeev, for this invaluable experience.

As mentioned earlier, I took baby steps during my undergraduate studies and learnt to walk holding Sanjeev's hand. But one man who inspired me to ask key questions in developmental biology and helped me to channelize my thoughts by stressing the importance of designing elegant experiments is my co-supervisor Prof. Carl-Philipp Heisenberg. I can not thank him enough to host me in his laboratory and expose me to a unique scientific environment. Discussion sessions with him were always enlightening. I still remember his briefings about project designing and scientific writing and will always be my companion in my future journey. He also provided me with absolute freedom to design and execute my experiments in his labs. He provided me with unique opportunities to interact with many greats during their visits to IST. He was truly there with me as a mentor when I needed it the most. Thank you, CP. There are many more things I am yet to learn from you. Allow me to bother.

Thank you both Sanjeev and Carl-Philipp for this excellent collaboration and I wish you both great success in the coming years.

I would not have enjoyed developmental biology if I had not met Dr Mahendra Sonawane. He has been most patient and at the same time encouraging, allowing me to learn this beautiful model system. His advice to be focused and patient during testing times of the project helped me a lot to eventually get over the line. He granted me access to his fish facility and other resources which helped me

tremendously in getting significant directions in my project. His scientific contributions to my research are immensely invaluable. Thank you, Mahendra for bringing out the developmental biologist in me.

I would like to thank my Research Advisory committee members, Prof. L.S. Shashidhara, Dr Chinmoy Patra and Dr Krishanpal Karmodiya for their valuable insights and engaging scientific discussions. Discussing concepts of developmental biology was always fun with you Shashi. Thank you for your valuable time and support. I would particularly like to extend my sincere regards to Krish for his constant guidance and also for allowing me to be a part of one of his research projects during the early days. The experience I gained from you in the field of transcriptional regulation and genomics is invaluable.

Many thanks to collaborators Dr Rakesh Mishra, Prof. Shubha Tole, Prof. Vidita Vaidya and Prof. Antonio Giraldez. Antonio has been truly inspirational and enthusiastic. His efforts in teaching me germline transplantations helped me immensely. Thank you, Shubha, and Vidita for allowing me to be a part of your research studies and at times lead from the front. These opportunities certainly exposed me to new areas of research. I would also like to thank Prof. Girish Deshpande for the brainstorming scientific discussions and all the guidance during manuscript preparation. His enthusiasm to discuss science from a broad perspective was truly inspirational. Thank you, Girish for your wonderful mentoring. Looking forward to learning more from you.

I am deeply grateful to Prof. K. N. Ganesh (Former Director, IISER-Pune) and Prof. L. S. Shashidhara (former head of the Dept. of Biology) for creating this wonderful scientific environment. Open science is the key and IISER-Pune has never failed to provide a suitable environment for it. I would also like to thank Prof. Jayant Udgaonkar (Director, IISER-Pune) and Prof. Thomas Pucadyil (Head of Dept. of Biology) for continuing to make IISER-Pune the home of Indian Science. Many thanks are due to Dr V. S. Rao (former officiating Registrar) and Col. (Retd) Raja Sekhar (Registrar) for all their support in academic and administrative processes. I also wish to thank the Dean academics for designing the curriculum and facilitating all the doctoral processes smoothly.

I am indebted to CSIR for the support extended through the direct SRF scheme towards my fellowship. Many thanks to DBT for support through multiple research fellow positions. I wish to thank the Company of Biologists for the travel award to visit the laboratory of Prof. Antonio Giraldez at Yale University. Many thanks to DST-FWF for supporting my exchange visits to the laboratory of Prof. Heisenberg. I would also like to thank DST-SERB and Infosys Foundation for supporting my travel to International conferences where I got an exciting platform to present my work.

I wish to express my gratitude towards all the members of the core facilities at IISER-Pune and IST Austria. Special thanks to the animal facility, microscopy facility, FACS facility and NGS facility at IISER-Pune and IST Austria. All the members and support staff at both places have been extremely accommodating

during my long hours or late-night experiments. Special mention to Shalini, who not only took great care of the fish facility but also assisted me in many of my experiments.

I would like to thank Dr Sridhar Sivasubbu for facilitating high-throughput sequencing for one of the projects at the sequencing facility of IGIB, New Delhi. I appreciate your timely help.

I wish to thank and express my gratitude towards all the past and present members of Sanjeev's lab. Thank you, Ranveer, Sunita, Kamal, Praveena, Rahul, Rafeeq, Mouli, Manjunath, Pavana, Rini, Aditi, Suyog, Satyajeet, Ameya, Smitha, Sneha Sagarkar, Sapana, Rutika, Ashwin, Indumathi, Manu, Ankitha, Akhila, Ankita, Meghana, Ayush, Keshav, Amit, Sneha, Suyash, Anirudh, Mrinmoy, Greg, Sudiksha, Harshada, Anjali, Ajay, Rahul, Tayyab and Yoshita for creating a wonderful environment. Special thanks to Rafeeq who taught me the basics of biochemistry and always mentored me throughout my early years at IISER Pune. Rahul who was always best in molecular biology techniques shared his special tricks with me to get the best results out of experiments. I would also like to convey my special thanks to Sunita, who was always accommodative and made sure I belong to this lab. I owe a lot to Suyog for not only helping me professionally but also for helping me to overcome daily frustration. Ankita deserves huge credit for helping me with the completion of this work as she was always available to assist and discuss experiments with me. Very aptly she describes herself as my 'scribble book' since I always used to rethink and reorder my experiments while discussing with her. Without your help

and support, I can only imagine achieving what I could. Thank you for everything. I would also like to convey my special regards to Mouli and Satyajeet for teaching the basics of bioinformatics. It helped me a long way during my PhD. Mouli also took the time to discuss my work and always provided new insights into my research questions. I am sure he will be a great supervisor and mentor. I am deeply grateful to Ankitha for extending as much help and support she could. Ankitha and Mukul are those I would always like to wish all the success for. I wish you both all the very best in your future endeavours. I would also like to thank Manu for always making sure the lab stays together and being friends with everyone. I am sure you will continue to spread such positivity wherever you go. Big thank you to Meghana, Greg and Mrinmoy for all the assistance they provided during manuscript preparation. Sincere thanks to Mr Vipradas for taking our care as he would for his children, for placing orders and making sure we remain strong in our testing times. Santosh has always made sure we never run out of glassware and plastic wares. He was always ready to assist whenever it was required. Thank you, Santosh.

Big thanks to all the members of the Heisenberg lab at IST. Many thanks to Philipp, Michael, Gabby, Vanessa, Matt, Keisuke, Jana, Daniel, Alex, Diana, Stefania, Nicoleta, Hitoshi, Shayan, Peng, Feyza, Silvia, Zoltan, Roland, Ehsan, Cony, Deborah, Koni and Karla for being wonderfully supportive during my stay. Each one of you has shared your valuable expertise allowing me to learn and establish this work. Special thanks to Michael and Keisuke for the assistance in designing and conducting experiments. Vanessa, Daniel and Alex were my go-to people whenever I was stuck with a technique or concept. They were also

supportive in my testing times making sure I get the best out of every visit to IST. I learnt how to think about scientific questions by interacting with each one of you. Many thanks and hugs to all of you.

I wish to thank the members of Krish lab for all their help and for creating a wonderful environment. Many thanks to Mukul (my little brother), Abhishek, Bhagyashree, Anjani, Ratnal, Deepak, Rashim, Bruhad and Bhagyashree. I would also like to thank members of Anjan lab, Ameya, Sneha, Bhoomi, Bhavani, Amit, Harpreet, Nilam, Shirsha for all your help. Many thanks are due to Soumen, Sanket, Devika, Maithilee, Trupti, Swati, Dyanesh for being there for all the chit-chats. I also like to thank the batch of 2017 for making the coursework period truly a fun time. I never enjoyed it as much after my college days. Thank you for bringing back all those memories.

I would like to convey my special regards towards all the staff of the biology department, administrative section, purchase section, IT section, academic and accounts section. Special mention to Piyush, Kalpesh, Mahesh, Mrinalini, Tushar, Sayalee, Anil Jadhav, Vrushali, Suresh, John, Sachin and Prabhas for all their help and support.

I wish to acknowledge all the faculty of the Biology department for keeping your doors always open for me. Discussing my work with you all was a true privilege. Special thanks to Girish, Kundan, Nagaraj, Sudha, Shashi, Aurnab, Thomas, Richa, Siddhesh and Krish.

Many thanks are due to Dr G. C. Kundu for providing me with the very first exposure to science. Your hard work and persistence truly inspired me. I learnt a lot about doing collaborative science from you which helped me throughout my PhD.

I would also like to thank the Head of Department, Modern College, Dr Rekha Gupta for always trying to push my limits to achieve greater success. My sincere thanks to all teachers who took extraordinary efforts for an ordinary student like me.

Acknowledgements will not be complete without mentioning my besties Pushkar, Ojas, Chaitanya. Nikhil and Aditya for being constantly with me throughout this journey. Ojas and Chaitanya were also involved experimentally and the support system provided by these were extremely vital for me to survive in many instances. I cannot thank them enough and wish them all the best for their future endeavours.

Last but not the least, I am indebted to my parents, Prasad, siblings, uncles and aunts for all the love they showered on me. Thanking my parents would be a crime for what they have gone through. They showed immense patience and a strong belief in me. My mom and dad always believed my education is the most precious thing they will possess. Many have heard about sacrifices made by Late Dr Abdul Kalam's sister for his education but I have personally experienced it. My Sister, Shivangi Tai made unimaginable sacrifices during the hard time we went through as a family. I am eternally grateful to her for this. Love you all for

all your blessings. Finally, I would like to thank almighty God for giving me valuable things even when I never deserved them.

Sincerely,

Saurabh

Table of Contents

List of Abbreviations

CRISPR	Clustered regularly interspaced short palindromic repeats
TALENs	Transcription activator-like effector nucleases
ZGA	Zygotic genome activation
YSL	Yolk syncytial layer
TGF- β	Transforming growth factor-beta
Bmp	Bone morphogenetic proteins
Fgf	Fibroblast growth factors
RA	Retinoic acid
Oep	One-eye pinhead
TF	Transcription factor
Chrd	Chordin
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
scRNA-seq	Single-cell RNA sequencing
ChIP-seq	chromatin immunoprecipitation with massively parallel DNA sequencing
GEM	Gel Bead-In Emulsions
GEO	Gene Expression Omnibus
Gsc	Goosecoid

Flh	Floating head (noto)
qPCR	quantitative-Polymerase chain reaction
Sox	SRY-related HMG-box genes
hpf	Hours post-fertilization
dpf	Days post-fertilization
DPBS	Dulbecco's phosphate-buffered saline
RIPA	Radioimmunoprecipitation assay buffer
FDR	False discovery rate
TSS	Transcription start site
UMI	Unique molecular indices
UMAP	Uniform Manifold Approximation and Projection
PCA	Principal component analysis
LSD1	Lysine-specific histone demethylase 1
JMJD3	Jumonji domain-containing protein D3
MLL1	Mixed lineage leukemia protein-1
eGFP	enhanced green fluorescent protein

GRNs	Gene regulatory networks
SATB1	Special AT-rich sequence binding protein 1
SATB2	Special AT-rich sequence binding protein 2
TADs	Topologically associated domains
MOs	Morpholinos
ASOs	Antisense oligonucleotides
IGV	Integrative genomics viewer
BIO	6-bromoindirubin-3-oxime
GO	Gene ontology
PVDF	Polyvinylidene fluoride
CPM	Counts per million
DE	Differentially expressed
NCCs	Neural crest cells
PFA	Paraformaldehyde
MZ	maternal/zygotic
MBT	Mid blastula transition

MBD	methyl-CpG-binding domain
SHH	Sonic Hedgehog
5mC	5-Methylcytosine
HDAC1	Histone deacetylase 1
HRP	Horseradish peroxidase
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis

Abstract

The basic body plan is determined within hours to a few days of embryonic development. This window is characterized by zygotic genome activation, lineage specification and morphogenetic events during organogenesis. Early embryonic patterning and morphogenesis are regulated by the interplay between various maternally deposited as well as zygotically transcribed RNA and protein determinants. However, understanding of precise transcriptional mechanisms sculpting embryonic structures remains inadequate. In the quest to identify novel mechanisms, here, we generated the lineage-specific transcriptional and chromatin accessibility profiles of gastrulating embryos. Our study highlighted Nodal signalling mediated dynamic chromatin remodelling activities necessary for segregation of axial mesoderm and endoderm progenitors. Moreover, our analysis offered an opportunity to identify and characterize understudied chromatin organizers during germ layer segregation. We characterized the function of one such lineage-restricted protein Satb2. Studies using transient knockdown allowed us to identify a novel Wnt-dependent role of Satb2 during early cell fate specification events. Generation of tractable loss of function and gain of function models in zebrafish enabled us to dissect molecular mechanisms underlying pathological conditions associated with *satb2* mutation. Moreover, integrative analysis of the transcriptome, genome-wide occupancy and chromatin accessibility revealed molecular interplays by which Satb2 performs contrasting functions during major gene regulatory transitions throughout early embryogenesis. We found maternal Satb2 negatively regulate zygotic genes by influencing the interplay between the pluripotency factors whereas, zygotic Satb2 activates the same group of genes during neural crest development. The comparative analysis underscores how these antithetical activities are temporally coordinated and functionally implemented. Taken together, our study highlights the evolutionary implications of chromatin organization in the regulation of landmark developmental transitions.

Synopsis

Early embryonic development is a culmination of carefully calibrated gene expression patterns orchestrated in a spatiotemporally coordinated manner. Typically this outcome is achieved by establishing higher-order gene regulatory networks or GRNs. How the GRNs are initiated, maintained and executed has been a major focus of inquiry amongst developmental biologists. Consequently, we have learnt a great deal about the regulation of tissue-specific gene expression leading to the differentiation of individual cell types. In recent years the focus has gradually shifted to the elucidation of genome-wide molecular changes that regulate critical cellular or developmental landmarks such as zygotic genome activation (ZGA), germ layer segregation, lineage commitment and organogenesis. However, studies focusing on the establishment of higher-order GRNs responsible for these processes through modulation of chromatin landscape have been limited.

In this study, we have attempted to delineate the molecular mechanisms underlying gene expression changes primarily dictated by dynamic changes in chromatin landscape during early embryogenesis in zebrafish. The main objective of this work is to understand how a basic body plan is determined in the vertebrates and extrapolate these findings to the mammalian systems. In the process, we also aim to provide insights into evolutionary changes at the molecular level which presumably occurred during the transition from chordates to vertebrates and differentiate these species from each other in mechanisms underlying body patterning.

In **Chapter 1**, we focused on understanding how the three germ layers are specified and segregated in early vertebrates. Towards this, we compared the single-cell transcriptome data of gastrulating embryos of zebrafish (vertebrates) and ascidians (chordates). Interestingly, we observed a unique cluster of cells that co-express markers of multiple

lineage progenitors. Strikingly, such cluster was observed only in fish but not in ascidians suggesting transient gene expression state of cells which follow inductive mechanism(s) during early gastrulation. To further dissect this we generated lineage-specific transcriptome and chromatin accessibility maps during gastrulation. We utilized fluorescently labelled transgenic lines for characteristic lineage marker genes, isolated the population of interest and performed bulk mRNA-seq. Integrating the transcriptome and chromatin accessibility profiles provided novel insights into the mechanisms of lineage segregation. Firstly, as the embryos develop the endoderm progenitor cells were found to gain a very distinct transcriptome signature compared to all other lineages. Secondly, we show that the segregation between the axial mesoderm and endoderm progenitors requires massive chromatin remodelling driven by fate determinant transcription factors. Our study indicates a gain in chromatin accessibility when endoderm is specified whereas the accessibility is lost to generate prechordal plate lineage. We perturbed the chromatin states using small molecular weight inhibitors against histone modifiers as well as modulating nodal signalling using chemical inhibitors and optogenetically induced systems. Results of these experiments further strengthened the notion that endoderm specification requires activatory mechanisms while axial mesoderm is specified through engaging repressive mechanisms. We also predicted transcriptional regulators which could play a potential role in modulating the local chromatin landscape to bring about the required transcriptional output. Further characterization of the genome-wide activities of these transcriptional regulators in a lineage-specific manner would provide significant direction towards our understanding of early embryogenesis.

In the quest to identify chromatin organizers which function actively during lineage segregation, we analyzed our lineage-specific transcriptome dataset and identified Satb2 as one of the factors enriched in the prechordal plate cells. In **Chapter 2,** we present evidence for the early divergence of Satb2 from its family member Satb1, suggesting a role of Satb2 during body plan determination. Loss of function studies using transient knockdown methods led to severe developmental arrest as embryos fail to develop beyond stages of early gastrulation. Transcriptome analysis of control and *satb2* knockdown embryos allowed us to identify the differential activity of Satb2 during

gastrulation. We show that while Satb2 positively regulates the ectodermal and neuroectodermal lineages, it can repress the endodermal lineages. These observations are extremely interesting and provide insights into how precise spatiotemporal developmental patterning is established. We further Integrated the genome-wide occupancy of Satb2 and the known regulators of mesendoderm such as Eomesa, Nanog, Ta, Tbx16, Mixl1 and Mxtx2. This comparative analysis indicated that Satb2 participates in two distinct gene regulatory modules and its function is defined based on whether Satb2 occupies the genomic sites which are either Eomesa bound or Nanog bound. Importantly, Satb2 negatively regulates endodermal genes when it shares its genomic occupancy sites with Nanog. To understand the upstream signalling inputs, we modulated Wnt and Nodal signalling using chemical inhibitors. Interestingly, we observed a negative correlation between Satb2 and Wnt signalling as in the case with Nanog, indicating Satb2 activity is restricted to specific cell types through active Wnt signalling. Collectively, these results provide novel mechanisms employed by Satb2 to regulate cell fate specification.

Although the transient knockdown approach shed great insights into Satb2's function, severe developmental defects generated by these methods restricted our understanding during stages of organogenesis. Thus, to distinguish between early and late activities of Satb2 during vertebrate embryogenesis, in **Chapter 3**, we generated *satb2* mutants in zebrafish using the CRISPR-Cas9 genome editing system. We not only could reproduce the phenotypic defects associated with genetic mutations in mammals, but our comprehensive genome-wide studies also highlighted the molecular mechanisms underlying neural crest specification and associated craniofacial abnormalities. We further show that Satb2 regulates key determinant genes involved in early neurogenesis such as *sox9*, *sox10*, *zic1* by regulating chromatin accessibility and nucleosome positioning. Comparative analysis using the mouse model presented further suggested that Satb2 primarily functions in the cranial neural crest cells than the peripheral neural crest cells. Taken together, our study highlights the conserved molecular mechanisms by which Satb2 establishes higher-order GRNs across vertebrates during the early stages of neurogenesis and neural crest development.

Fascinated by the biphasic nature of *satb2* expression, in **Chapter 4**, we characterize the function of maternally deposited *satb2*. Maternally deposited *satb2* was observed to be reduced significantly at the onset of ZGA. To test the significance of the depletion of *satb2*, we overexpressed *satb2* during the early cleavage stages of embryogenesis. Similar to the knockdown approaches the overexpression of *satb2* also led to severe developmental defects. Transcriptome and genome-wide occupancy analysis revealed that maternally deposited *satb2* functions as a transcriptional repressor in a cell-autonomous manner to establish precocious zygotic genome activation. We further show that Satb2 negatively regulates the 'first wave' zygotic genes by altering the local histone modification marks. Thus, downregulation of *satb2* is essential for the successful onset of ZGA. Strikingly, complementing our genome-wide studies with genetic manipulations, we identified a negative feedback loop between *satb2* and the pluripotency factor *pou5f3* (mammalian homolog of OCT4). We further extended these observations using a human embryonic stem cell differentiation model. Expectedly, we observed a similar biphasic nature of *satb2* expression during differentiation to neuronal precursors. Taken together, this section of our study uncovers a novel biphasic and contrasting function for the global regulator Satb2 during ZGA and later during organogenesis. Our study also emphasizes how a single determinant regulates temporally distinct yet interlinked processes by employing differential regulatory mechanisms during the short period of embryonic development.

Chapter 1

Understanding the lineage-specific chromatin landscape at the onset of zebrafish gastrulation

1.1 Introduction

Cell lineage specification and segregation into distinct germ layers are the most crucial events occurring during gastrulation and are necessary for the formation of complex tissues and organs in development. These processes involve an interplay of various signalling molecules controlled in a Spatio-temporal manner, which results in an organized animal body plan (Carmany-Rampey and Schier, 2001; David and Rosa, 2001; Feldman et al., 2000; Gritsman et al., 1999; Keller et al., 2008; Schier and Talbot, 2005). Nearly about 85% of animal phyla constitute triploblastic organisms i.e. organisms with three distinct germ layers, namely ectoderm, mesoderm and endoderm.

Ectoderm is known to give rise to epidermal and neural tissues, mesoderm gives rise to blood and muscle tissues while endoderm is a major contributor to many of the vital organs such as heart, lung, liver and kidney. Hence, an accurate blueprint of specification of these early lineages during gastrulation holds significant value in patterning and the growth of an organism. Although these germ layers are considered to be distinct, they often contribute to each other's function. For example, neural crest cells are ectodermal in origin but many recent reports suggest their crucial role in cardiac function which is of mesoderm origin (Donoghue et al., 2008). Similarly, mesoderm gives rise to all muscle tissue types and participates in the function of all major internal organs of endodermal origin (Ferretti and Hadjantonakis, 2019; Tirosh-Finkel, 2006). Thus, communication between these three germ layers is critical during organogenesis.

Various theories have been proposed in the past centuries on how three germ layers obtain individual identities and segregate from each other (Alexanian et al., 2017; Ferretti and Hadjantonakis, 2019; Kimelman and Griffin, 2000; Rodaway and Patient, 2001; Slagle et al., 2011). To test these various hypotheses, the identification of new genetic regulators has been of interest for many decades. Earlier attempts towards this were mostly restricted to characterizing the function(s) of transcription factors or morphogen gradients regulating the specification process by cell-cell signalling. Importantly, gene regulatory cascades underlying these processes are regulated in a hierarchical manner and genetic mechanisms controlling major developmental signalling pathways

responsible for cell fate specification have been studied extensively. However, the epigenetic mechanisms including alteration in chromatin architecture and their impact on cell fate specification have come into the limelight more recently (Barrero et al., 2010, 2012; Ruthenburg et al., 2007; Szutorisz and Dillon, 2005; Szutorisz et al., 2005). Furthermore, how exactly the embryonic system modulates itself to bring about these changes remain elusive.

1.1.1 Cell fate specification during gastrulation

Cell fate specification is the earliest event in the process to form a particular cell and tissue type. During this process, regulators of a particular cell type take over the charge and drive the underlying gene network of lineage commitment. At this stage, the cell is committed to a specific fate. Davidson et al. classified three basic types of lineage commitment (Davidson, 1991). The first type is the autonomous specification, in which the donor cells transplanted into a host embryo will develop the same cell types as they would have in the donor embryo while the donor embryos will lack these particular cell types. The autonomous specification is generally a characteristic of invertebrate embryos like molluscs and tunicates (e.g. Ascidians). The second type of specification is very appropriately termed as conditional specification since the specification of the transplanted cells from donor embryos into host embryos depends upon the conditions in which the cells are transplanted. If cells are transplanted before they are fate committed, they acquire a new identity based on the surrounding environment where these cells are transplanted in. It is also intriguing to see how surrounding cells from donor embryos reorganize themselves to generate missing cell types. This mode of the specification is a characteristic of vertebrate embryos (e.g. Zebrafish, Xenopus and mammals). The third type is the syncytial specification. Almost all insects are syncytial during early development wherein many nuclei share the same cytoplasm of the large egg cell. Differential expression patterns across the anterior-posterior axis of the syncytial cell are necessary for cell fate specification. For example, in Drosophila, the anterior half is marked with the localization of bicoid while the posterior half is with Nanos and as the embryos develop further, other cells are directed towards particular cell fate based on the gradients formed by these proteins as a signalling cue (Rivera-Pomar and Jäckle, 1996).

Large numbers of such studies over many decades conducted in various organisms like a worm, sea urchin, fruit fly, African frog, chick and mouse has contributed to our current understanding of embryonic development (Bedzhov et al., 2014; Howard-Ashby et al., 2006; Maduro, 2010; Muñoz-Sanjuán and -Brivanlou, 2001; Streisinger et al., 1981). However, studying early events during vertebrate embryogenesis was still limited because of the difficulties associated with placental development, whole embryo imaging methods and longer generation time.

1.1.2 Zebrafish as a model to study early events of lineage commitment

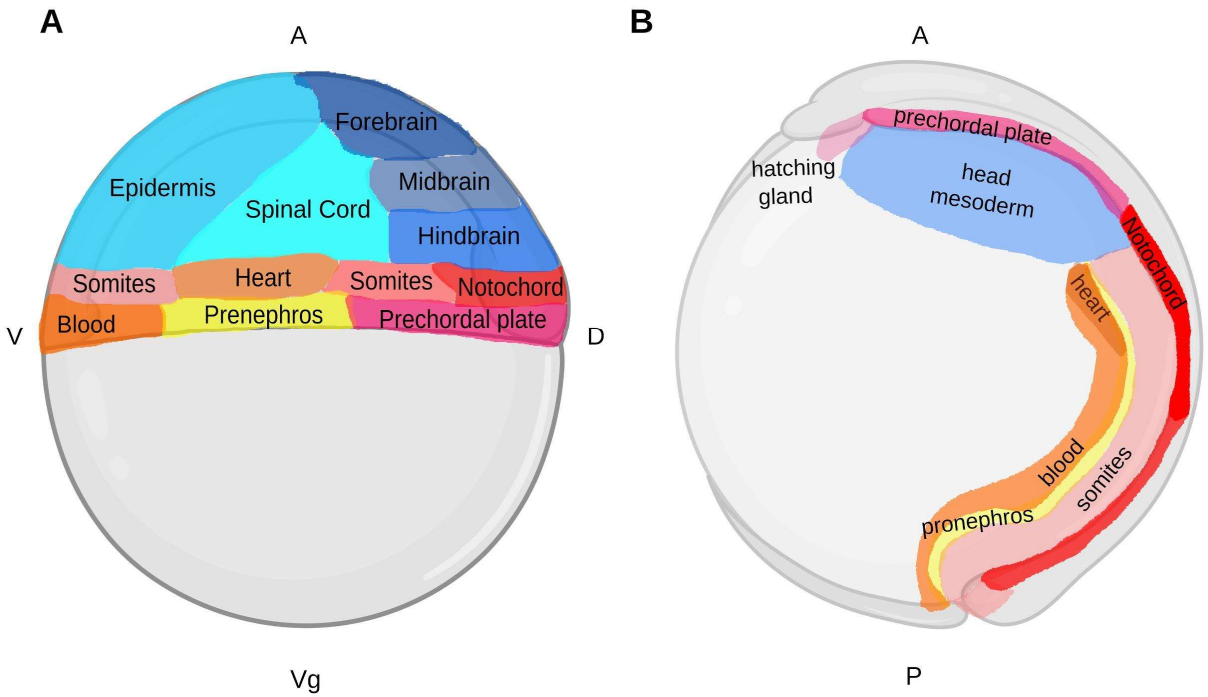
The introduction of zebrafish as a model system by Streisinger helped developmental biologists to overcome the limitations associated with other vertebrate systems to study cell fate specification during early vertebrate embryonic development (Streisinger et al., 1981). Large fecundity, external fertilization and faster development made zebrafish a popular choice to study vertebrate embryogenesis. Moreover, the optical transparency of embryos was advantageous in the development of time-lapse imaging and lineage tracing. All these features of zebrafish in combination with the development of genetic tools led to the first forward genetic screen for the vertebrate system. Genetic mutants isolated in the screens by Christiane Nüsslein-Volhard and Wolfgang Driever were extensively documented in a special issue of the journal '*Development*', volume 123, published in 1996 (Nüsslein-Volhard, 2012). These screens were instrumental in the identification of genes responsible for a variety of processes ranging from axis determination to organ development. The simultaneous rise in the development of key technologies such as whole-genome sequencing, large genomic libraries, reverse genetic tools such as morpholinos, TALENs, CRISPR-Cas system helped the community immensely to understand molecular mechanisms underlying various key processes during zebrafish development (Bedell et al., 2011; Hwang et al., 2014; Li et al., 2016)

1.1.3 Early embryonic development and cell fate specification in zebrafish

Rapid embryonic development is the key characteristic of zebrafish embryogenesis. Within 10 hours after fertilization, the basic body plan becomes visually evident (Figure 1.1) (Schier and Talbot, 2005). Following transcriptional competence established during zygotic genome activation (ZGA), At 6 hpf, tissue structure, equivalent to the amphibian dorsal organizer (Spemann-Mangold organizer) described as 'shield' in zebrafish becomes visually apparent. Internalization of Meso-endoderm cells initiates from the shield and further undergoes a process of convergence and extension. These series of events are classified as gastrulation and mark the time point of initiation of lineage segregation. Concurrently, the blastoderm continues to cover the yolk sack till 10 hpf or the bud stage which marks the end of gastrulation. The majority of cell types are committed and specified by this time point.

Embryonic transplantation experiments, lineage tracing and genetic mutation studies have contributed immensely to the understanding of the establishment of fate maps in early zebrafish embryos. The cell fate map can be traced back to as early as from the late blastula stage of development, 3.5 hpf (along the animal-vegetal axis) wherein, the ectoderm is present mostly at the animal pole region; mesoderm, near the margin and endoderm at the margin just above the yolk syncytial layer (YSL). YSL is known to play a crucial role in the induction of the mesendoderm population (Chen and Kimelman, 2000; Erter et al., 1998). Results from transplantation and live-cell imaging experiments also indicate the overall arrangement of sub-classified cell types such as axial mesoderm and notochord (mostly dorsal), blood and somites (mostly ventral), neural ectoderm cells contributing to forebrain and midbrain (dorsally towards the animal pole) and non-neural ectoderm, epidermal cells (ventrally) along the dorsal-ventral (D-V) axis.

Figure 1.1 Cell fate map during zebrafish gastrulation.



(A) Cell fate map at the onset of gastrulation (6hpf, shield stage). Three major germ layers ectoderm, mesoderm and endoderm are arranged in the Animal-vegetal axis (A-Vg). Various mesendoderm progenitors are arranged in the dorso-ventral axis (D-V). **(B)** Different mesoderm derived tissues are depicted in the anterior-posterior axis (A-P). (Schematic representations are an adaptation from Schier and Talbot 2005.)

Mckenna et al., and Spanjaard et al., in parallel studies, made use of the CRISPR-Cas system to induce scars in the most ubiquitous genomic sequences for lineage tracing (McKenna et al., 2016; Spanjaard et al., 2018). Newly developed single-cell transcriptomics methods (ScRNA-seq), TracerSeq by Wagner et al., and URD by Farrell et al., also validate this observation. Results from these studies reveal intermixing of cells from different lineages. Further, cell trajectories constructed using computer simulations display a highly dynamic nature throughout different stages, suggesting plasticity during

embryonic development (Farrell et al., 2018; Wagner et al., 2018). TracerSeq method is based on the use of transposon-mediated barcoding for lineage tracing while URD is based on the idea of sequencing multiple stages within a short interval of time and creating diffusion maps using transcriptional signatures. Overall, these studies further strengthen the notion of the dynamic nature of cell lineage trees and suggest a possible role for gene regulators or what can be called 'fate drivers' in shaping up defined cell boundaries leading to tissue organization.

1.1.4 Signalling during cell fate specification in zebrafish

Saturated mutation screens mentioned above, contributed largely towards the identification and characterization of 'fate driver' genes and signalling pathways, which could be responsible for fate determination. Post-ZGA, TGF- β superfamily transcription factors such as smad1/smad2/samd5 are highly expressed and establish themselves as a major determinant of Nodal, Bmp and FGF signalling. Nodal signalling is involved in the animal-vegetal axis whereas Bmps are shown to be involved in the specification of D-V patterning (Little and Mullins, 2006; Ramel and Hill, 2012; Schier, 2003; Shen, 2007). In zebrafish, nodal is activated by nuclear β-catenin by directly binding to cis-regulatory regions (Schneider et al., 1996). Studies with one eye pinhead (oep) mutant (mutation in acvr, receptors of nodal signalling) suggested that loss of nodal signalling leads to loss of mesendoderm specification (Schier et al., 1997). Upon activation of Nodal signalling, effector molecule Smad2 gets phosphorylated and interacts with co-factors to regulate transcription of the target genes such as *gsc*, *sox17*, *foxa2* and *sox32*. One of the crucial TF, Foxh1 is important for mesendoderm specification and craniofacial development has been shown to interact with smad2 (Nelson et al., 2014; Pei et al., 2007). Moreover, differential nodal signalling has been shown to induce different germ layers. A recent study from Sako et al. suggested both strength and duration are important for the induction of layers. High level and long exposure of Nodal signalling induce mesoderm whereas low levels induce posterior mesoderm markers required for the formation of the notochord (Sako et al., 2016). Further, Bmp signalling molecules have been shown to regulate the expression of *chd* and *noggin* (Kondo, 2007). Overexpression of these leads to severe dorsalized phenotypes. Many such signalling cross-talks are required for

morphogenesis during gastrulation and have been extensively reviewed (Pinheiro and Heisenberg, 2020; Schier and Talbot, 2005; Zinski et al., 2018).

1.1.5 Role of chromatin architecture and epigenetic mechanisms in cell lineage commitment during development

Suggestive involvement of chromatin architecture in cell fate specification can be traced way back to experiments conducted in 1969 by Nicole LeDourian with chick-quail chimaeras. Her experiments are the earliest examples describing the association of conditional specification with chromatin compaction in cell fate specification (Le Douarin et al., 1997). Nicole LeDourian made systematic observations using Feulgen-Rosenbeck's procedure that stains DNA. She observed that the embryonic and adult cells of quail contained enlarged nucleoli and large amounts of heterochromatic DNA, which helped her to distinguish quail cells from chicken cells. Kashif Ahmed and co-workers further made use of electron spectroscopic imaging to directly examine structural changes in chromatin architecture of early mouse embryos reflecting pluripotency and lineage commitment (Fussner et al., 2010). They observed that chromatin is less compact in pluripotent cells in 2-cell stage mouse embryos as compared to a fate determined cells from 8-cell stage embryos. This study supported and strengthened the observations made by Nicole LeDourian and highlighted the importance of chromatin structure analysis in the identification of the differentiation state of cells.

C. H. Waddington in 1942 coined the term ‘epigenetics’ to explain the involvement of genetic elements in driving cell differentiation. However, with constant contributions from various labs across the globe, today we briefly summarize epigenetics as the inheritance of phenotypic variations through gene regulation mechanisms that do not involve modulation in underlying DNA sequences (Bonasio et al., 2010). During the late 1990s, scientists were able to decode core chromatin structure (Luger et al., 1997). Depending upon the status of cytosine methylation at CpG islands, various histones tail modifications such as methylation, acetylation, ubiquitination, nucleosome positioning and histone variants; chromatin is either euchromatinised i.e. more accessible to binding of

transcription-associated factors resulting in increased transcriptional rate or undergo heterochromatinization, leading to decreased expression.

Recent advances in tools to analyze chromatin states by assessing chromatin accessibility (ATAC-seq), levels of histone modifications such as H3K4me1, H3K27Ac and occupancy of several transcription factors using ChIP-seq and long-range interactions through Hi-C maps have facilitated the identification of regulatory elements which can be successfully validated using live reporter systems (Andrey and Mundlos, 2017). It has been observed that multiple enhancer elements function together to regulate the expression of a subset of genes and are referred to as super-enhancers (Pott and Lieb, 2014). Moreover, large changes in interactions between genomic elements have been observed during the differentiation regime of human pluripotent stem cells regulating lineage commitment (Dixon et al., 2015). Gene expression programs are regulated at a higher-order level by *cis*-regulatory binding factors such as CTCF and cohesin complexes. Recently, CTCF and cohesin complexes have been implicated in the regulation of gene expression during gastrulation in zebrafish and medaka fish by generating chromatin loops (Meier *et al.,* 2018; Nakamura *et al*., 2018). Many histone modifier complexes including the polycomb group (PCG), MLL and SAGA are shown to play a crucial role during embryonic development by mediating long-range interactions (reviewed in Andrey and Mundlos, 2017). Hi-C studies revealed long-range interacting regions of transcriptionally active and inactive domains which are called A/B compartments (Lieberman-Aiden *et al*., 2009) and identified topologically associating domains (TADs) which are shown to facilitate enhancer-promoter interactions (Gomez-Diaz and Corces, 2014). Kaaij *et al.,* in their study on characterization of enhancers in zebrafish, made an interesting observation that in contrast to studies in other mammalian species, active enhancers are hyper-methylated, while hypo-methylated enhancer regions are mostly associated with transcription factors which are functional at later stages of embryonic development (Kaaij *et al*., 2016). 4C experiments suggest constant interaction of these enhancer elements with respective promoters throughout embryonic development but their activity is dependent on priming by H3K4me1. These findings make the overall regulation even more complex to understand but interesting to decode further.

These observations were further extended by generating systematic Hi-C contact maps throughout zebrafish early embryogenesis. As seen in mammals, the zebrafish genome also exhibits the dynamic nature of chromatin architecture by actively dissolving TADs during ZGA and then re-establishing during the stages of organogenesis (Kaji *et al*., 2018). These observations do not go completely per what is observed in studies using the fruit fly wherein the chromatin architecture is more structured upon onset of major ZGA suggesting differential higher-order regulation between invertebrates and vertebrate species (Hug *et al*., 2017). However, upon the onset of minor ZGA, the genome appears to be more unstructured. This differential regulation could be because of the absence of Zelda like factors in zebrafish and it would be interesting to test this hypothesis further. Lost chromatin architecture during ZGA re-establishes as embryos develop and a completely structured genome can be seen again at 24 hpf highlighting the importance of higher-order structural changes required during the period of gastrulation. Thus, making it interesting to characterize the lineage-specific transcriptome and associated higher-order structural changes in chromatin to delineate the mechanisms underlying lineage segregation during embryonic development.

1.2 Methods

1.2.1 Experimental models

Zebrafish (*Danio rerio*) were maintained as described previously (Westerfield, 2000). All the experimental procedures were carried out following the guidelines from the institutional animal ethics committee at IISER Pune and IST Austria. Embryos were raised at 28-31 ^{o}C in E3 medium and staged as described previously (Kimmel et al., 1995).

1.2.2 Single-cell RNAseq analysis

To generate single-cell gene expression datasets, Wild type TU/TL embryos were raised to 6 hpf at 28.5 ^{o}C, hand dechorionated and dissociated further in 1x DMEM-F12 + 125 mM EGTA by hand tapping. Cells were collected by centrifugation at 300x g for 2 min and resuspended in 1X DPBS + 0.1% BSA. The cell suspension was passed through a 40 µm Flowmi cell strainer (Sigma). Cell viability was estimated using Countess-II automated cell counters and only samples with more than 95% viability were processed further with 10x Genomics Chromium. Approximately 8000 cells were captured in GEMs. scRNA-seq libraries were prepared as described in 10X Genomics manuals (Single Cell 3′ Reagent Kits V3, User Guide PN-1000092). Immediately following GEM generation, reverse transcription reaction was carried out using Eppendorf master cycler pro by incubating at 53 ^{o}C for 45 min followed by denaturation at 85 ^{o}C for 5 min. Single-stranded cDNA was purified using DynaBeads MyOne Silane Beads (Thermo). cDNA amplification was performed for 11 cycles with an initial denaturation at 98^{o}C 3 min followed by repeated cycles of 98 ^{o}C for 15 seconds, 63 ^{o}C for 20 seconds and 72 ^{o}C for 1 min. The final extension was performed at 72 ^{o}C for 1 min. cDNA quality was determined using Bioanalyzer 2100 high sensitivity assay. cDNA was further fragmented, end-repaired and A-tailed according to the manufacturer's instructions. Before proceeding with adapter ligation, samples were purified using a double-side cleanup protocol with SPRI beads (Beckman Coulter). The adapter ligated sample was subjected to amplification for 11 cycles using indexing primer from

Chromium i7 Multiplex Kit (PN-120262). Amplified libraries were again subjected to double-side purification using SPRI and quantified using Qubit fluorometer and library size was estimated using Bioanalyzer 2100 (Agilent).

1.5 pM of the denatured libraries were used as an input to obtain sequencing reads using 28 cycles for read 1, 8 cycles for indexes and 101 cycles for read 2 on Nextseq 550 (Illumina) at IISER Pune.

Sequencing data were further processed with default parameters using cell Ranger 3.0.2 (10X Genomics). Reads were aligned to the danRer10 genome using STAR aligner. Feature matrices generated using the CellRanger were utilized for further analysis using Seurat 3.0.

1.2.3 Analysis of Single-cell RNAseq data from *Ciona intestinalis*.

Feature matrices generated using the CellRanger pipeline for mid gastrulation stages were downloaded from GEO (GSE131155) and further analysed using Seurat 3.0.

1.2.4 Isolation of specific lineages using Fluorescence-Activated Cell Sorting (FACS).

To isolate specific lineages for downstream analysis, males from each transgenic line were out crossed to wild-type females. This was done particularly to avoid maternal contribution hindering specific labelling of cell types. To isolate ectoderm, we crossed *krt4*: egfp line with *sebox*: GFP line and isolated negative population by sorting on ARIA III cell sorter system (BD, USA). Axial mesoderm progenitors were isolated using *gsc*:CAAX:egfp (Barone et al., 2017). To avoid any non-specific readout, only cells with higher expression levels were isolated. Notochord progenitors were isolated using the *flh*:egfp line. To isolate pan mesendoderm cells *sebox*:eGFP transgenic fish were used. Endoderm cells were isolated using *sox17*: GFP line. To avoid contaminating signals from Kuffler's vesicles, high *sox17* expressing cells were excluded from sorting and downstream analysis specifically at the bud stage. All the cell populations were directly

collected into Trizol. Purity and enrichment of specific cell populations were analyzed using qPCR and further subjected to bulk-mRNA seq analysis.

1.2.5 mRNA, Morpholino injections and inhibitor treatment.

In-vitro mRNA transcription was performed using SP6 mMessage mMachine Kit as per the instructions from the manufacturer (Ambion). Synthesized mRNA was treated with Turbo-DNAse and precipitated using LiCl overnight at -20 ^{o}C. Synthesized mRNA was checked for quality and quantity by agarose gel electrophoresis and Nanodrop2000 respectively. Small single-use aliquots of mRNAs were kept frozen at -80 ^{o}C till further use. Glass capillaries (WPI) were pulled using a needle puller (P-97, Sutter Instruments) and mounted on a microinjection system (PV820, World Precision Instruments).

For ubiquitous overexpression and morpholino mediated knockdown studies, embryos were arranged in agarose moulds and injections were performed at the 1 cell stage as described previously (Westerfield, 2007). For inducing mesoderm, a cocktail of 100 pg of pCS2-*ndr1* and 2 ng of morpholino against *sox32* were injected at the 1 cell stage. To induce endoderm progenitor cells, 50 pg of *sox32* mRNA were injected at the 1 cell stage.

To perturb nodal signalling, embryos were incubated in an E3 medium containing 30 µM of SB505124. The status of histone modifiers was modulated using previously characterized inhibitors. Inhibitors were added to the medium at the 1K cell stage. Embryos were harvested at 10 hpf and imaged using stereoscopes.

Small molecule inhibitor	Final concentration	Molecular target	Effect on chromatin state after drug treatment	Company and Cat number
GSK343	50 µM	EZH2	Active	SIGMA (SML0766)
GSK-LSD1	150 µM	LSD1	Repressive	SIGMA (SML1072)
MM102	50 µM	MLL1	Repressive	Calbiochem (5006490001)
BIX-01294	10 µM	G9a	Active	SIGMA (B9311)
GSK-J4	7.5 µM	JMJD3	Repressive	SIGMA (SML0701)
SB505124	25 µM	Nodal	NA	SIGMA (S4696)

1.2.6 LED Light Activation of Opto-*acvr1b* and Opto-*acvr2b* Receptors.

Embryos from the inter cross of oep$^{-/-}$ /gsc:caax:eGFP$^{+/+}$ were injected with Opto-Acvr1a and Opto-Acvr2b (Sako et al., 2016). Injected embryos were exposed to blue LED light using an incubator (Herp Nursery II, 69802, Lucky Reptile) equipped with 300 light-emitting diodes (SMD5050) (Grusch et al., 2014) and set to 29.5 ºC to activate the Nodal signalling. Light intensity was controlled with an analogue dimmer and measured with a digital power meter (PM100D; Thorlabs). Embryos were exposed for the desired time interval with the maximum intensity of blue light (5.12 mW/mm2) unless mentioned otherwise. Embryos kept in a dark box covered with aluminium foil and raised in the same incubator as the light-exposed embryos were used as a control. After the desired stage is reached, embryos were visualized under a fluorescent stereomicroscope to confirm the activation of Nodal signalling and immediately harvested in Trizol for RNA extraction followed by bulk-mRNA seq as described above

1.2.7 Bulk mRNA seq

For lineage-specific transcriptome analysis, 100-200 ng of total RNA with RIN > 8 from each sorted sample was used as a starting material to prepare mRNA sequencing libraries using Illumina Truseq mRNA seq kit. All cDNA samples were amplified for 10-15 cycles depending on cycle number estimation by qPCR (Kapa library quantitation kit, KAPA Biosystems). Amplified libraries were purified using two rounds of 0.8x volume of Ampure XP beads (Beckman Coulter). The library concentrations were estimated using the Qubit DNA HS system. Finally, all the libraries were pooled and subjected to high throughput sequencing using Hiseq 2500 (IGIB, New Delhi).

For transcriptome analysis after nodal inhibition and optogenetic modulation, embryos were harvested at the desired developmental stage in Trizol. 500 ng of total RNA was used as a starting material to prepare libraries using NEB ultra II RNA kit (NEB, USA). Purified and pooled libraries were sequenced using 76 bp PE chemistry on Nextseq 550 at IISER Pune.

Sequencing reads were trimmed for quality using Trimmomatic and aligned to daRer10 genome assembly using HISAT2 aligner. Counts for each gene feature was estimated using the FeatureCounts package from Rsubread. Differential expression analysis was performed for replicates using EdgeR. Gene ontology analysis for upregulated and downregulated genes were performed using a web-based tool, gProfiler and WebGestalt (Reimand et al., 2007-7).

1.2.8 ATAC seq

To determine chromatin accessibility of Gsc^{+} and $Sox17^{+}$ progenitors, 10,000 isolated cells from the out cross of these transgenic fish were harvested at 6 hpf and processed for Omni-ATAC seq as described previously (Corces et al., 2017) with inputs from Dr Amanda Ackermann. Briefly, cells were washed with 1x DPBS and resuspended in cell lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM KCl, 0.1 % NP 40, 0.1% Tween20 and 0.01% Digitonin) and incubated on ice for 3 min. Further, cells were washed with a wash buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM KCl and 0.1% Tween20) by

centrifugation at 500 g for 10 min 4°C. The supernatant was discarded and the pellet was resuspended in 25 µl 2x Tagmentation buffer (Illumina, catalogue # 15027866), 16.5 µl DPBS, 0.5 µl 10% Tween 20, 0.5 µl 1% Digitonin, 5 µl nuclease and 2.5 µl Tn5 transposase enzyme (TDE1, Illumina, catalogue # 15027865) and incubated for 28 min at 37°C. After the tagmentation reaction, DNA was isolated using the Qiagen MinElute Reaction Cleanup Kit (Qiagen). Purified DNA was used as an input to generate a library by amplifying with 2x Q5 DNA polymerase mix (NEB) and indexing primers. Optimal cycles were determined using qPCR analysis. Amplified libraries were purified using Agencourt ampure XP beads to remove adapters and larger fragments.

Sequencing reads (76 bp PE) were obtained on NextSeq 550 (Illumina) at IISER Pune and trimmed for Nextera adapters using default parameters of Trimmomatic PE. Trimmed reads were aligned to danRer10 using default parameters of Bowtie2 (Langmead and Salzberg, 2012). Briefly, BAM files were subsampled to 55 million reads in each sample using bbmap and sorted by name. paired-end bed files were obtained using bedtools bamtobed. Reads were displaced by +4 bp and -5 bp. BigWig files were generated using bamCoverage (deepTools). Peaks were annotated to the nearest genes using Homer and classified into promoter (+/- 2 Kb) and non-promoter regions. K-means clustering was performed around +/- 2 Kb of TSS of danRer10 genome using deepTools.

To assess chromatin accessibility of nodal inhibitor-treated and induced mesendoderm progenitors, 10 embryos at 4.5 hpf were harvested and processed as described above for ATAC-seq. Pooled libraries were sequenced using 41 bp PE chemistry on NextSeq 550. Sequencing reads were further processed as described above.

Motif analysis was performed to identify consensus binding sites present in open chromatin regions using findMotifsGenome.pl from Homer.

1.2.9 Differential chromatin accessibility analysis (DiffBind)

The differential chromatin accessibility analysis was performed using the DiffBind (Stark et al., 2011). Significantly differentially accessible peaks were identified using the Deseq2 package and only sites with FDR < 0.05 and fold change of > Log2 (+/- 1.5) were used

for further analysis. Differential accessible sites were annotated to the nearest gene using Homer. Core promoter was defined as +/- 2 Kb from the TSS.

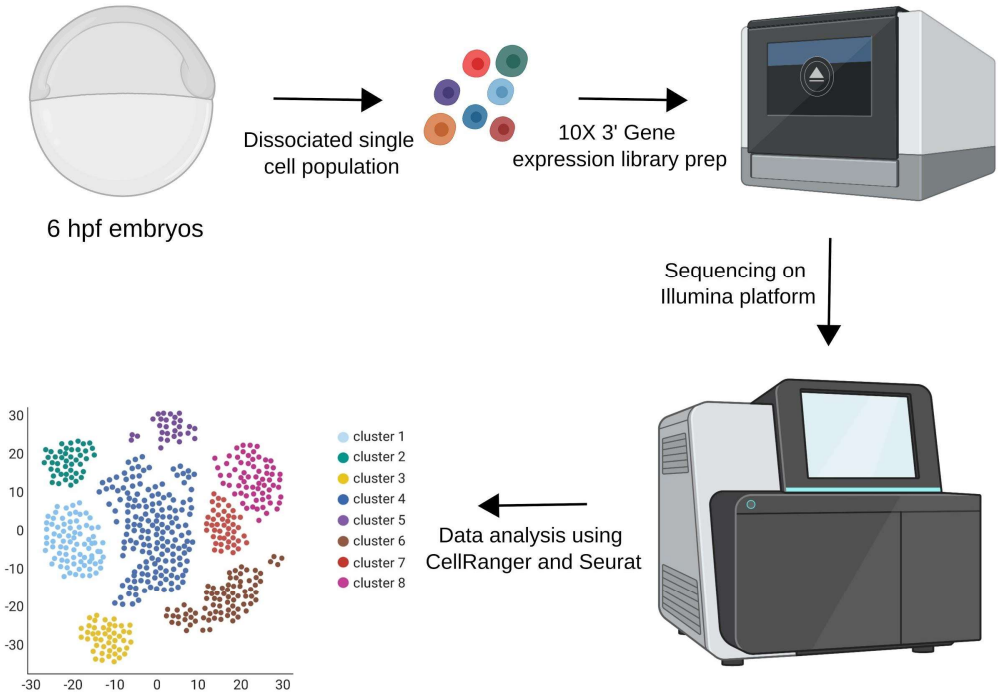
1.3 Results

1.3.1 Identification of gene expression modules at the onset of gastrulation

In an attempt to characterize the transcriptional states of three distinct germ layers during the early stages of gastrulation in vertebrates, we utilized the recently popularised method of single-cell 3' mRNA sequencing analysis using the 10X genomics platform. This allowed us to capture gene expression profiles of multiple cell types in one frame. We obtained a single cell transcriptomic map of zebrafish embryos at 6 hpf (onset of gastrulation with the clear formation of dorsal organizer equivalent, shield) and processed it to identify gene expression modules using Seurat 3.0 (Figure 1.2).

We generated high depth (~100,000 reads/cell) and high-quality scRNA seq data for nearly 3500 cells with detection of 7,902 unique molecular indices (UMIs) and an average of 2462 genes per cell (Figure 1.3A). Feature matrix generated using CellRanger was imported into the Seurat pipeline using R 3.4 and the quality of the dataset was assessed using several parameters. First, violin plots for RNA features and RNA count again highlighted the high quality of distribution across several cells. Reads mapping to mitochondrial gene sets were filtered during the alignment step and the dataset was free of mitochondrial contamination as evident in a quality check analysis (Figure 1.3B). Second, Jackstraw plots and elbow plots were used to estimate the number of principal components (PCs) required for successful data integration. The number of PCs were estimated to be 15 and used to generate UMAP.

Figure 1.2 Schematic for the workflow of single-cell transcriptome analysis.



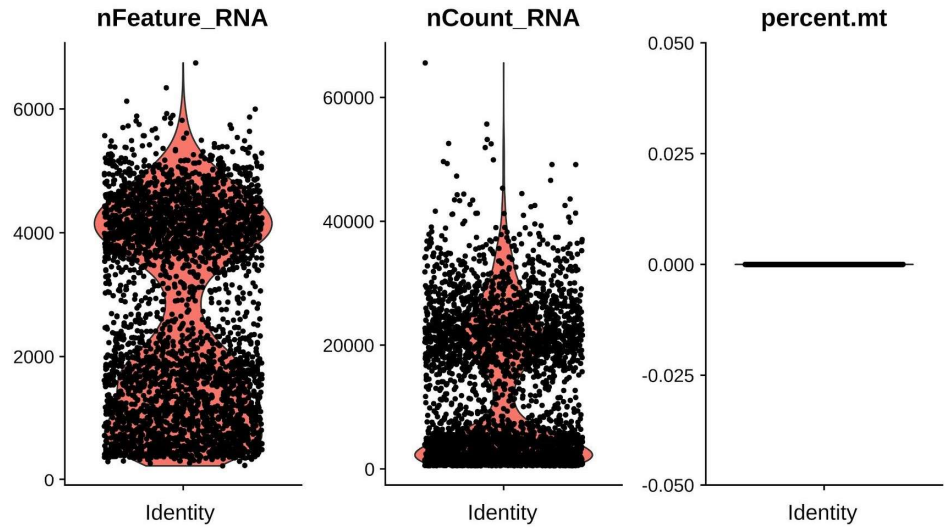
Gastrulating zebrafish embryos were harvested at 6 hpf and dissociated into single cells. In general, single cells are encapsulated in gel beads providing a unique identity to each cell using a barcoding approach. A pool of uniquely barcoded single cells was sequenced using Nextseq 550 and Seurat 3.0 was used for annotating clusters and studying gene expression patterns.

Figure 1.3 Run statistics and quality control analysis of single-cell data.

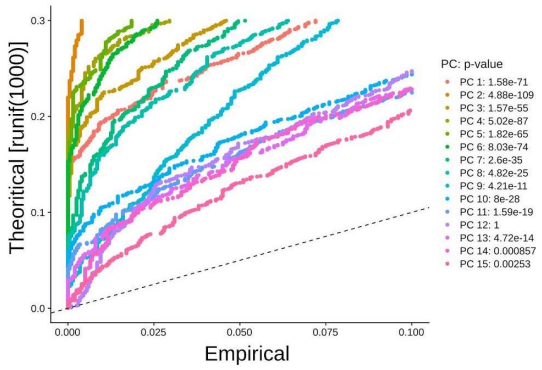
A

% Mapped Reads	Number of cells	Mean Reads	Genes per cell	UMI per cell
85.6%	3424	118,024	2,462	7,902

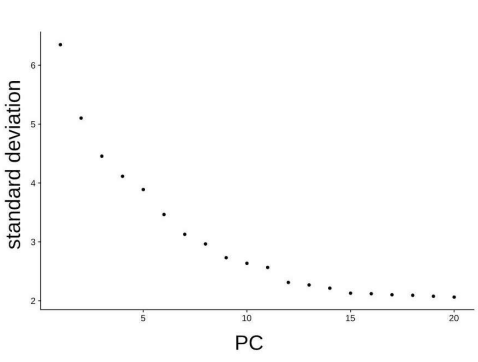
B



C



D



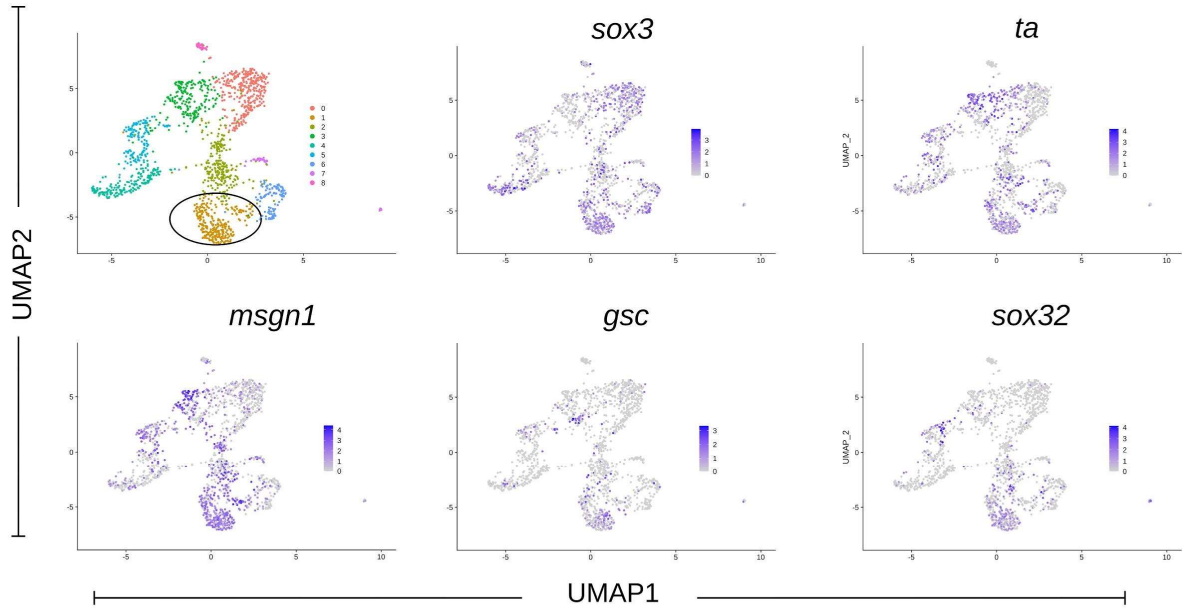
(A) Sequencing run statistics summarizing the number of cells sequenced, percentage of reads mapped to the genome, mean reads per cell, an average number of genes annotated for each cell and average number of UMIs per cell. **(B)** Violin plots for a number

of genes (nFeature), number of RNA molecules (nCount) and percentage of mitochondrial genes across each individual cell in the scRNA-seq dataset. **(C)** Jackstraw plot representing dimensionality of principal components (n=15) with high enrichment and significant p-value. **(D)** Elbow plot depicting the ranking of principal components based on a percentage of variance. Variance saturation was observed after PC 15, hence only 15 PCs were used for downstream analysis.

UMAP clustering allowed us to identify 9 distinct clusters based on gene expression profiles of single cells. We were particularly Intrigued by the behaviour of cluster number 1 (yellow, marked with an elliptical circle) since cells belonging to this cluster expressed markers of multiple cell lineages such as *sox3* (ectoderm), *ta* (mesoderm), *msgn1* (paraxial mesoderm), *gsc* (axial mesoderm) and *sox32* (endoderm) (Figure 1.4). Whereas, cluster number 0 expresses markers of ectoderm exclusively. We could also identify a cluster that potentially gives rise to rare and specialized tissue type enveloping layers (EVL).

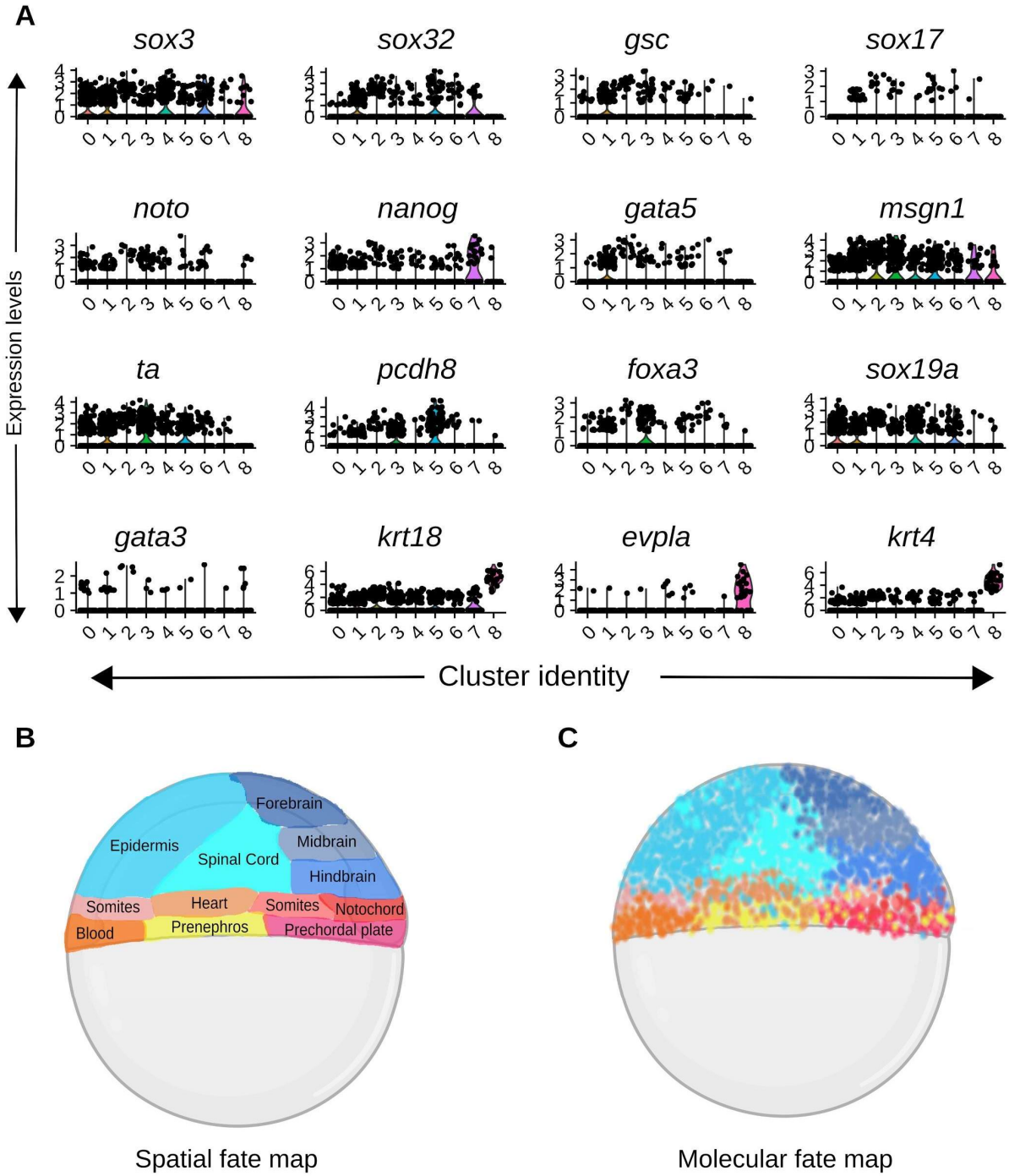
Cluster specific expression profiles for multiple lineage markers, *sox3* (ectoderm), *sox32* (endoderm), *gsc* (axial mesoderm), *sox17* (endoderm), *noto* (notochord, mesoderm), *nanog* (endoderm), *gata5* (endoderm), *msgn1* (paraxial mesoderm), *ta* (mesoderm), *pcdh8* (paraxial mesoderm), *foxa3* (ectoderm), *sox19a* (ectoderm), *gata3* (neural ectoderm), *krt18*, *evpla* and *krt4* (endoderm) confirmed these co-expression gene modules in various cell clusters (Figure 1.5A). These results prompted us to revisit the classical model of cell fate specification which suggests spatial segregation of lineage precursors at shield stage (6 hpf) (Schier and Talbot, 2005) (Figure 1.5B). The results from the scRNA-seq analysis, however, hint at the dynamic and transient transcriptional state of cells during gastrulation prior to lineage commitment instead of defined spatial compartments (Figure 1.5C). These results also support previous observations that at least in part mesoderm and endoderm is derived from common progenitors (Kimelman and Griffin, 2000).

Figure 1.4 Embryonic cells at the onset of gastrulation exhibit co-expression of lineage markers.



(A) Uniform Manifold Approximation and Projection (UMAP) representation of single-cell sequencing data depicting various clusters annotated based on co-expression gene modules. Cluster 1 (yellow) marked with an elliptical ring denotes a group of cells that co-express markers of all three major lineages. Feature Plots for ectoderm (*sox3)*, mesoderm (*ta)*, paraxial mesoderm (*gsc*) and endoderm (*sox32*).

Figure 1.5 Cluster specific expression patterns of lineage markers.



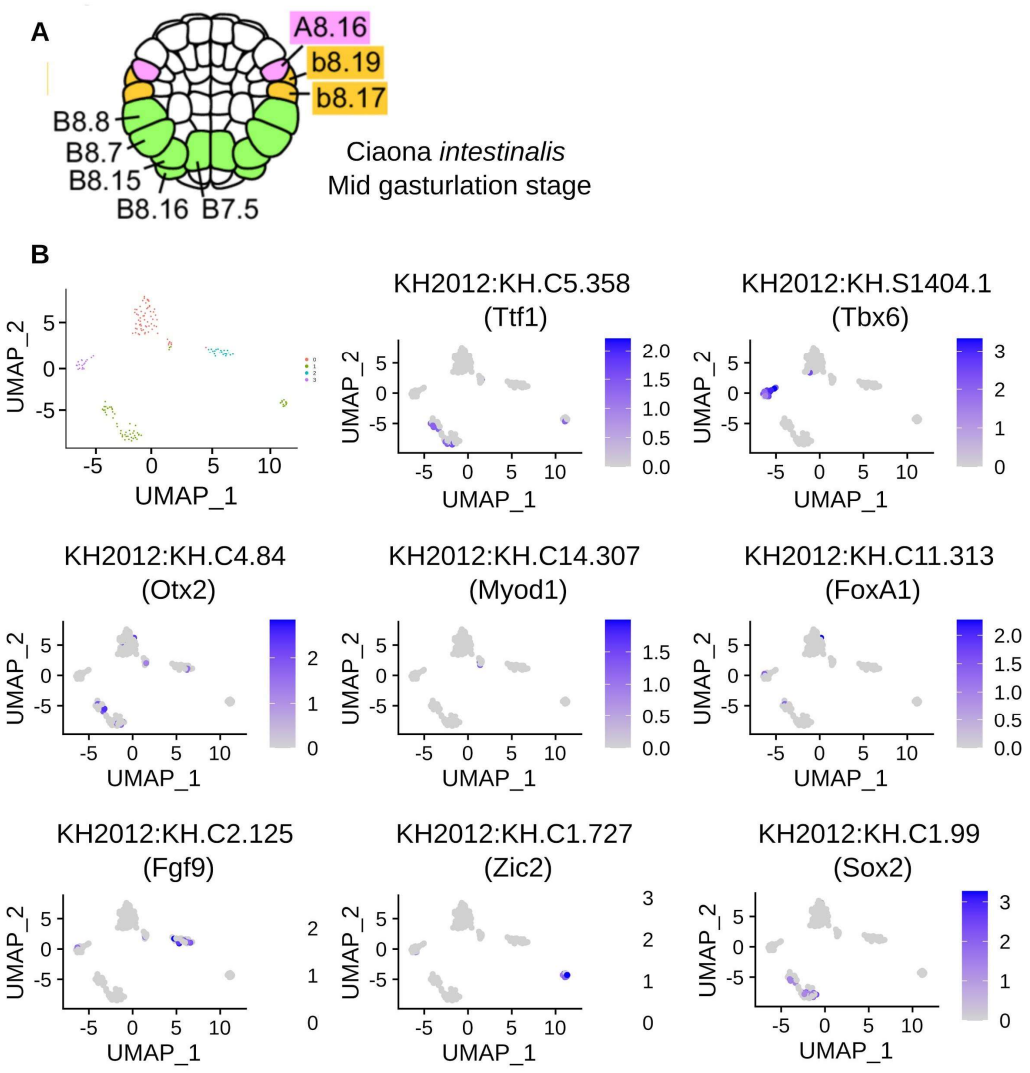
(A) Violin plots representing cluster-specific expression (based on UMAP clustering) patterns for lineages, Ectoderm (*sox3*, *sox19a*, *gata3*), Mesoderm (*noto*, *ta*), Axial mesoderm (*gsc*, *foxa3*), Paraxial mesoderm (*pcdh8*, *msgn1)*, Endoderm (*sox17*, *gata5*,

sox32) and Enveloping layer (*krt18*, *krt4*, *evpla*) highlighting co-expression modules in cells belonging to cluster 1. **(B)** Cell fate map at the onset of gastrulation (6hpf, shield stage). Three major germ layers ectoderm, mesoderm and endoderm are arranged in the Animal-vegetal axis (A-Vg). Various mesendoderm progenitors are arranged in the dorso-ventral axis (D-V) (based on Schier and Talbot 2005). **(C)** Cell fate map based on lineage state based on gene expression patterns during development.

To gain further insights into the transient nature of cell fate in the context of evolution, we reanalyzed previously published scRNA-seq from mid gastrulating (110 cells) stages of ascidians (*Ciona intestinalis*) (Cao et al., 2019). Ascidians are tunicates belonging to the phylum Chordata. In ascidians, cell lineages are largely determined in an autonomous manner as opposed to inductive (conducive environment as a prerequisite) nature of fate specification in vertebrates and mammalian systems (Nishida and Stach, 2014). Recently, geometrical constraints along with morphogen gradients have been implicated in body plan determination and reproducibility in ascidians (Guignard et al., 2020). However, lineage fate is determined as early as 16 cell stages during development (prior to gastrulation) as opposed to vertebrate embryos in which cells commit later during gastrulation.

UMAP clustering and feature map extraction suggested, as, in the case of zebrafish, there are no clusters of cells that co-expresses multiple markers of different lineages. This data suggests the possibility of evolutionary mechanisms of cell lineage segregation during vertebrate gastrulation and made a stronger case for detailed lineage-specific molecular characterization.

Figure 1.6 Determinative (autonomous) mode of cell fate specification does not exhibit co-expression modules during gastrulation.

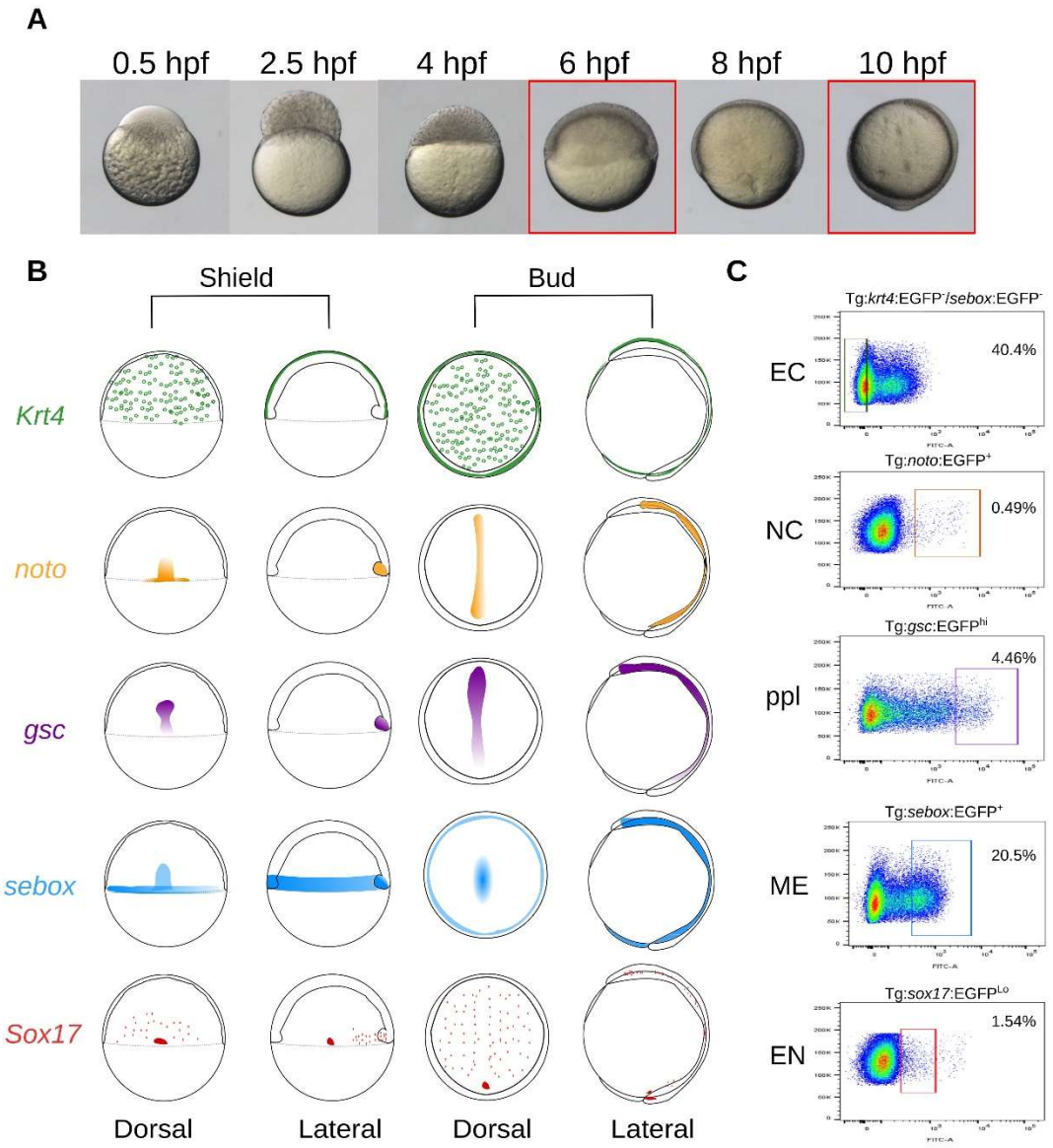


(A) Schematic representation of cell fate in an ascidian embryo at 110 cell stage, mid-gastrulation (reproduced from Tokuoka et al. 2007). **(B)** UMAP analysis highlighting different clusters and Feature plots for various lineage markers Ttf1, Tbx6, Otx2, Myod1, Foxa1, Fgf9, Zic2 and Sox2 highlighting exclusive expression of lineage markers in different cell clusters.

1.3.2 Characterization of lineage-specific transcriptome

scRNA seq studies revealed co-expressing cell clusters at the onset of gastrulation. However, as embryos develop, cells segregate into genetically distinct lineages and are defined spatially. Though current scRNAseq methods provide great insights into cellular heterogeneity, they have their limitations due to low depth, low genome coverage and high background noise making it difficult to precisely characterize transcriptomes. Moreover, low abundant factors which are crucial to modulate mRNA abundance are often unidentified through scRNAseq. To delineate this, we adapted an alternative strategy to generate lineage-specific transcriptome at two stages, 6 hpf (shield) and 10 hpf (bud) (Figure 1.7A). The shield stage marks the onset of gastrulation, at which mesendoderm progenitor cells start ingressing through dorsal organizer equivalent in fish. Bud stage denotes the completion of epiboly and gastrulation events. Primitive cell lineages at the bud stage are organized in an anterior-posterior axis (Figure 1.1). We utilized previously characterized transgenic zebrafish for markers of multiple lineages such as Tg:*Krt4*:eGFP (EVL), Tg:*noto*:eGFP (notochord), Tg:*gsc*:CAAX:eGFP (axial mesoderm), Tg:*sebox*:eGFP (mesendoderm) and Tg:*sox17*:eGFP (endoderm) (Figure 1.7B). Transgene labelled specific cells from these embryos were isolated using FACS and subjected for bulk mRNA seq to obtain mRNA abundance of these lineages (Figure 1.7C).

Figure 1.7 Strategy for lineage specific transcriptome.



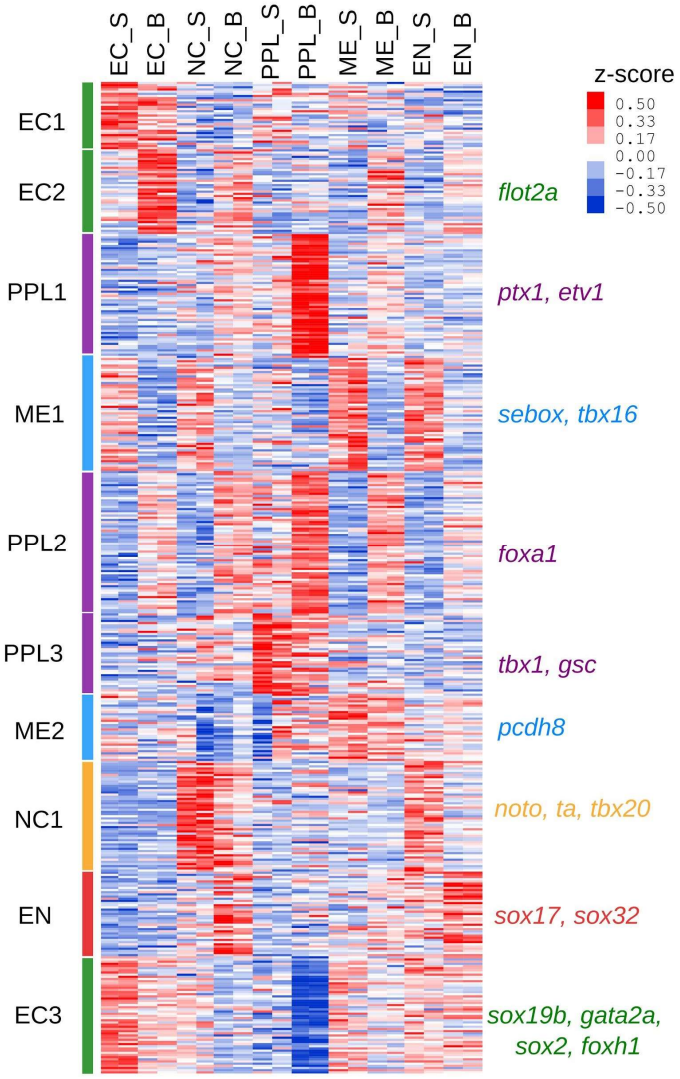
(A) Bright-field images of developmental series of zebrafish embryos. Bulk mRNA-seq analysis was performed at the onset of gastrulation (6 hpf, shield) and the end of gastrulation (10 hpf, bud). **(B)** Schematic representing dorsal and lateral view for expression patterns for transgenes *krt4:egfp*, *noto: GFP*, *gsc:caax: GFP*, *sebox: GFP* and

sox17: GFP at 6 hpf and 10 hpf. **(C)** Example scatter plots depicting sorting strategy to isolate various lineages such as ectoderm, mesoderm, axial mesoderm, paraxial mesoderm and endoderm for bulk mRNA-seq using transgenic lines represented earlier.

We obtained very high quality and consistent datasets in two biological replicates. K-means clustering of these datasets across two different developmental stages allowed us to identify various lineage-specific gene expression modules (Figure 1.8). These modules were classified and annotated to specific cell lineages based on the expression of well-characterized marker genes. For example, 3 different gene modules were assigned to ectoderm based on the enrichment. EC1 cluster represents genes enriched in ectoderm lineage at both shield and bud stage, whereas EC2 cluster represents genes expressed at higher levels in ectoderm lineage only at bud stage. We could also identify specific clusters of genes, EC3 belonging to the neural ectoderm represented by classical markers like *sox19b*, *gata2a*, *sox2* and *foxh1* which behaves similarly during developmental programs. Moreover, this analysis allowed us to identify clusters belonging to axial mesoderm (PPL #1-3, marked by *gsc*, *pitx1*, *foxa1*), notochord (NC, marked by *noto*, *ta*, *tbx20*), paraxial mesendoderm (ME, marked by *pcdh8*, *sebox* and *tbx16*) and endoderm lineage (EN, marked by *sox17* and *sox32*). We also observed engaging dynamics between the expression of lineage-specific markers *msgn1* and *lbx2.* Both these proteins are known to be expressed in muscle progenitors and lateral mesoderm. Our Spatio-temporal analysis revealed that *msgn1* is expressed at higher levels during early stages whereas, expression of *lbx2* is increased during late stages of gastrulation. These results together highlight the necessity of temporal activity of different lineage regulators to achieve specific functions.

Altogether, the generation of such a comprehensive transcriptome for more than 20,000 genes in a lineage-specific manner allowed us to identify expression dynamics of various transcription factors, cell adhesion proteins and could significantly help researchers to characterize functions of many understudied factors during the process of cell lineage commitment and morphogenesis.

Figure 1.8 Identification of lineage enriched factors.



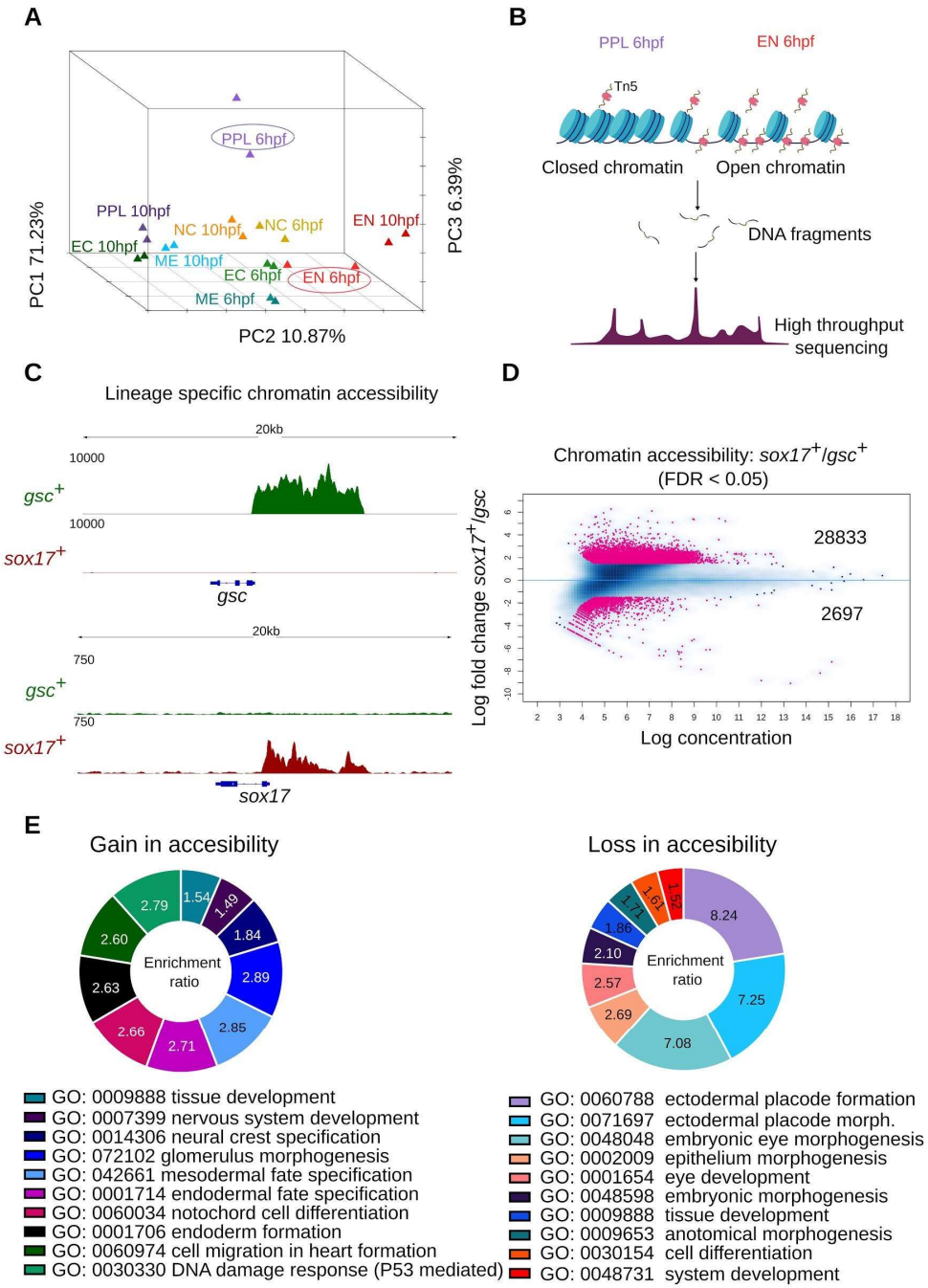
Heatmap representation of k-means clustering of lineage-specific transcriptome data at shield (S) and bud (B) stages of development to identify lineage enriched genes. Enriched gene clusters are colour coded as green (ectoderm), purple (prechordal plate), blue (paraxial mesoderm), yellow (notochord) and red (endoderm). Representative gene markers are highlighted for each cluster. Scale bar represents a normalized score based on counts per million values from bulk mRNA-seq experiments. blue (low) to red (high). Two independent experiments are clustered next to each other highlighting consistency between the replicates (N=2).

1.3.3 Lineage-specific chromatin accessibility analysis revealed endoderm progenitors exhibit higher chromatin accessibility

Three-dimensional principal component analysis of lineage-specific transcriptome suggested massive reprogramming during the stages of gastrulation. At 6 hpf, the gene expression profile of all the lineages clustered along one axis of PCA. However, as embryos develop, gene expression programs show clear segregation. Interestingly at 10 hpf, all the lineages except endoderm are seen to be clustered together, whereas endoderm showed the most distinct behaviour attaining special identity to gene expression programs (Figure 1.9A).

To understand the molecular mechanisms underlying these dynamic changes in the gene expression program, we analysed chromatin accessibility of closely spaced Gsc^{+} and $Sox17^{+}$ cells by utilizing Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) at 6 hpf (Figure 1.9B). We validated enrichment of specific lineages and confirmed the status of chromatin accessibility on the TSS of *gsc* and *sox17* locus in respective conditions (Figure 1.9C). Next, we performed differential chromatin accessibility analysis using DiffBind. Strikingly, chromatin accessibility in $Sox17^{+}$ cells is significantly higher at 28833 regions across the genome. Whereas, Gsc^{+} cells showed higher accessibility at only 2697 sites (Figure 1.9D). This observation was remarkable as it hints toward the active role of chromatin remodelling in the process of lineage commitment. Further, we annotated these differentially accessible sites to the nearest genes and subjected them to gene ontology programs. Gene ontology analysis suggested that differentially enriched regions (gain inaccessibility) were mostly associated with the genes involved in mesendoderm fate specification. Whereas, regions that showed a loss in chromatin accessibility were associated largely with ectodermal fate specification such as ectodermal placode and eye morphogenesis (Figure 1.9E). Collectively, data from our ATAC-seq studies indicated that changes in chromatin accessibility during gastrulation could play an important role in regulating gene expression programs necessary for lineage identity and segregation.

Figure 1.9 Endoderm progenitors exhibit distinct gene expression modules with higher chromatin accessibility.



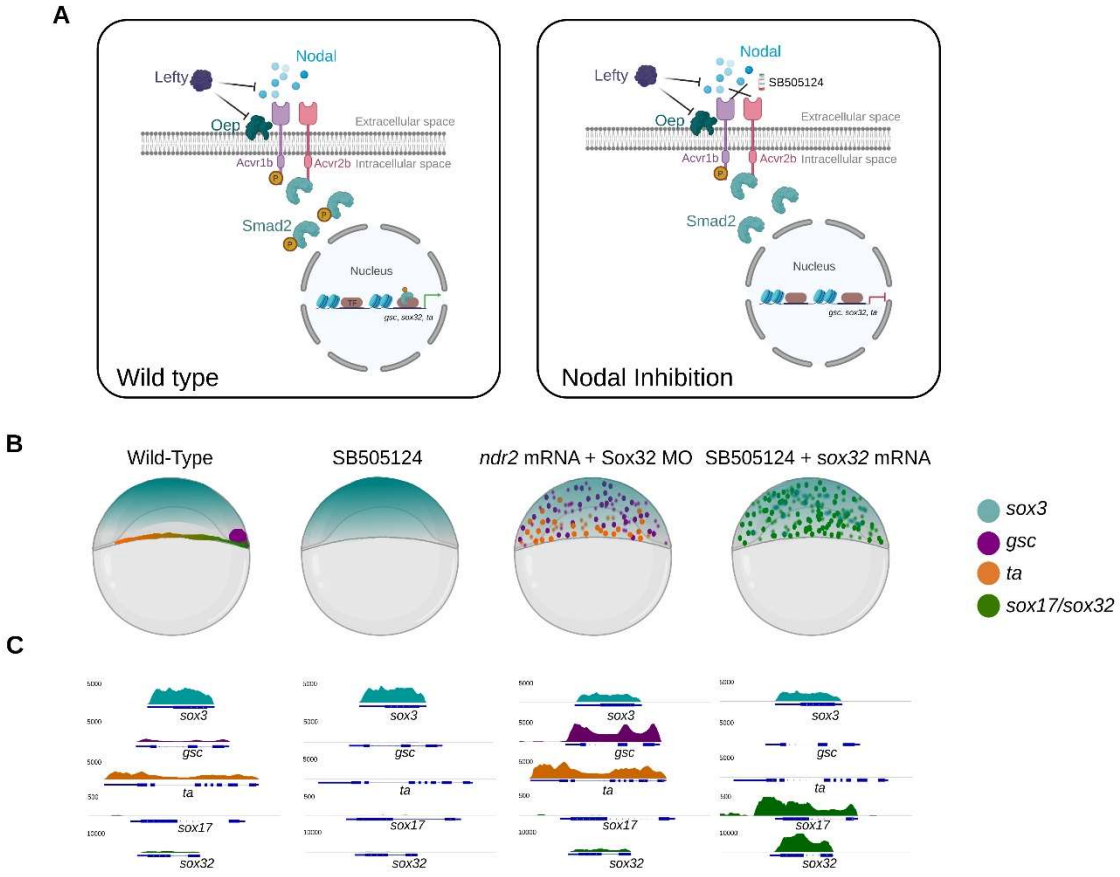
(A) Three-dimensional principal component analysis of lineage-specific transcriptome suggesting endoderm progenitors are clustered distantly from other lineages during embryonic development. **(B)** Schematic representation for chromatin accessibility

analysis using axial mesoderm and endoderm progenitors at shield stage. **(C)** Integrative genomic viewer snapshot to visualize sequencing reads from chromatin accessibility analysis on gene locus of *gsc* and *sox17* promoters validating the purity of progenitor cells. (Values represented on Y-axis are the average representation of two independent experiments, N=2). **(D)** Smear plot depicting differential chromatin accessibility sites. 28833 endoderm enriched and 2697 axial mesoderm enriched sites were identified using DiffBind analysis. **(E)** Gene ontology analysis for endoderm enriched (gain in chromatin accessibility) and axial mesoderm enriched (loss in chromatin accessibility).

1.3.4 Nodal targets drive changes in chromatin accessibility to drive cell fate decisions

Our lineage-specific transcriptome and chromatin accessibility studies highlighted the importance of active regulation of gene expression programs for lineage commitment. Moreover, this data also suggested that endoderm lineage has more accessible chromatin than cells of the axial mesoderm origin. However, we were stumbled with the question of which factors could potentially mediate this chromatin remodelling? To answer this in parts, we decided to perturb major signalling pathways and examine the effect on fate specification and chromatin accessibility. Signalling pathways such as Bmp, FGF, RA, WNT and Nodal have been widely implicated during vertebrate gastrulation (von Bubnoff and Cho, 2001; Muñoz-Sanjuán and -Brivanlou, 2001; Rankin et al., 2018; Schier, 2003; Schier and Talbot, 2005). Amongst these, Nodal signalling has been shown to regulate key lineage markers of mesendoderm cells such as *gsc*, *ta*, *sox17* and *sox32 (*Dougan et al., 2003; Sako et al., 2016; Schier, 2003*)*.

Figure 1.10 Schematic representation for a strategy to induce cell fate.



(A) Signalling pathway induced upon the interaction between nodal ligands and activin receptors leads to phosphorylation of cytoplasmic Smad2. Phosphorylated Smad2 shuttles in the nucleus and co-occupies gene locus with transcription factors such as Foxh1 to activate downstream mesendoderm specification genes (Left). Nodal signalling inhibition using small molecular inhibitors (SB505124) disrupting the interaction between nodal and activin receptors. Smad2 no longer can shuttle into the nucleus leading to failure in mesendoderm specification (right). **(B)** Resulting cell fate representation in our experimental setup: Induced ectoderm (SB505124), induced mesoderm (*ndr2* mRNA + *sox32* MO) and induced endoderm (SB505124 + *sox32* mRNA) denoted by expression domains of *sox3* (blue), *gsc* (purple), *ta* (yellow) and *sox17*/*sox32* (green)**. (C)** Integrative

genomic viewer snapshot for bulk mRNA-seq reads upon induction of respective cell fate (Values on Y-axis are average of three independent experiments, N=3).

Therefore, we next decided to perturb Nodal signalling using a well-characterized small molecular inhibitor SB505124. Exposure to SB505124 blocks interaction between nodal ligands and transmembrane activin receptors resulting in no phosphorylation of cytoplasmic smad2 preventing its nuclear localization to activate target genes (Figure 1.10A). We treated zebrafish embryos with 25 µM SB505124 from 2-4 cell stage and performed bulk transcriptome analysis at 4.5 hpf (Dome) stage. We deliberately decided to take a readout at 4.5 hpf to keep contributions from endogenous signalling at a minimum. Further, we decided to induce mesoderm cells by injecting nodal receptor 2 (*ndr2*) along with morpholino against endoderm determinant *sox32*. To test, if endoderm determinant *sox32* alone is sufficient to increase chromatin accessibility independent of other upstream molecular cues, we overexpressed *sox32* in the background of nodal inhibition (Figure 1.10B).

As documented in the literature, nodal inhibition with SB505124 resulted in a significant reduction in mRNA abundance of target genes *gsc*, *ta*, *sox17* and *sox32*. However, expression levels of ectoderm marker *sox3* remained unchanged (Figure 1.10C). We could also successfully induce the expression of mesodermal genes such as *gsc* and *ta* as well as endodermal markers, *sox17* and *sox32* in respective embryos.

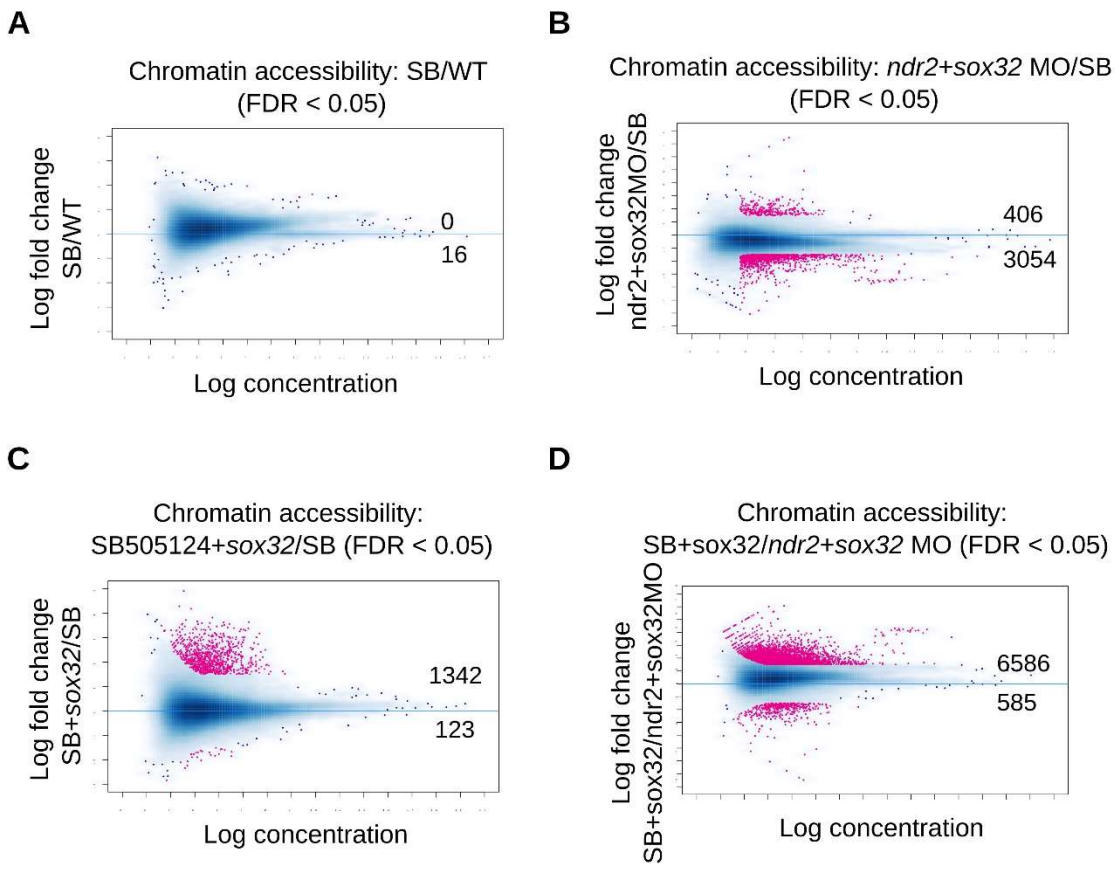
Next, we performed chromatin accessibility analysis using ATAC-seq and performed differential accessibility analysis. DiffBind analysis between wild-type embryos and embryos treated with nodal inhibitors did not show any significant changes in chromatin accessibility (Figure 1.11A). This validated our strategy of analysing accessibility at 4.5 hpf to avoid signals from endogenously expressed genes. However, similar to the observations made in ATAC-seq data with sorted cell populations, induced mesoderm using overexpression of *ndr2* + *sox32* MO showed reduced chromatin accessibility when compared to embryos treated with SB505124 (Figure 1.11B).

Interestingly, overexpression of endoderm determinant *sox32* even in the background of nodal inhibition resulted in increased chromatin accessibility when compared to embryos treated with SB505124 alone (Figure 1.11C) or when mesoderm is induced (Figure 1.11D). This data suggests that although initial nodal signals are required for activation of mesendoderm fate drivers like *sox32*, change in chromatin accessibility is largely mediated by these factors independent of the level of Nodal signalling. Collectively, these results indicate that transcriptional activation of nodal targets, *gsc*, *sox32* is a key for chromatin remodelling necessary for lineage commitment.

1.3.5 Modulations in the chromatin landscape using chemical inhibitors affect cell fate decisions

Chromatin accessibility changes to permit binding of transcription regulators and perform their target function is governed by many epigenetic processes such as DNA methylation and histone modifications (Beisaw et al., 2018; Signolet and Hendrich, 2015). Histone octamers are heavily modified at the N-terminal tails by the processes of acetylation, methylation, phosphorylation, ubiquitination resulting in activation or repression of target gene expression (Figure 1.12A). The interplay between ‘writers’ and ‘erasers’ play important roles in regulating gene expression programs during the early stages of development (Klose et al., 2006; Shi et al., 2004).

Figure 1.11 Induced endodermal progenitors exhibit higher chromatin accessibility compared to ectoderm and mesoderm independent of upstream nodal signalling.



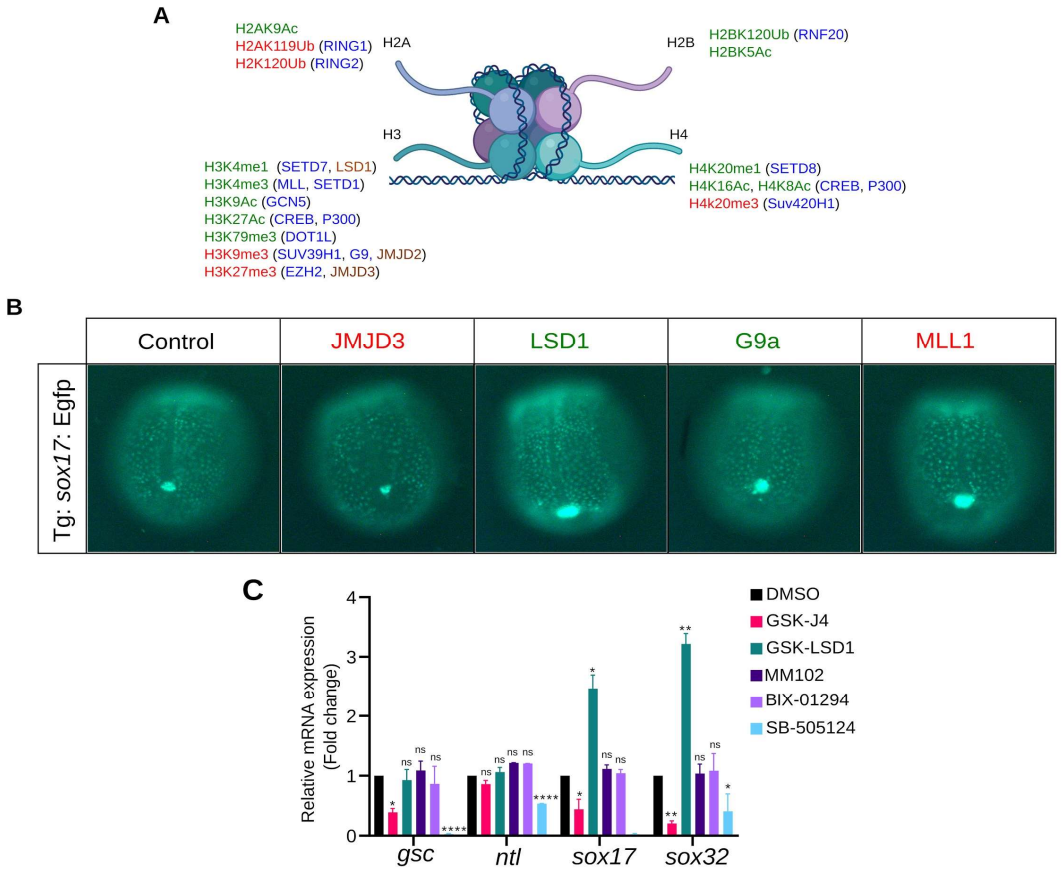
(A) Differential chromatin accessibility profiles using smear plots between wild-type and nodal treated (induced ectoderm) embryos. **(B)** Between induced mesoderm (*ndr2* mRNA + *sox32* MO) and induced ectoderm (SB: SB505124). **(C)** Between induced endoderm (SB505124 + *sox32* mRNA) and induced ectoderm (SB505124). **(D)** Between induced endoderm (SB505124 + *sox32* mRNA) and induced mesoderm (*ndr2* mRNA + *sox32* MO). DiffBind was used to identify differential accessible regions.

To confirm our notion further, that dynamic chromatin remodelling is required for lineage commitment, we decided to modulate chromatin landscape by using specific small molecular inhibitors against various writers and erasers. We targeted activities of

demethylases of H3K4me3; LSD1 (resulting in active chromatin), H3k27me3; JMJD3 (resulting in repressive chromatin) and methyltransferase of H3K9me3; G9a (active), H3K4me3; MLL1 (repressive).

To analyze the effect of chemical inhibition, we treated embryos from Tg:*sox17*:eGFP at the 4 cell stage and visualized $Sox17^+$ cells using fluorescence microscopy (Figure 1.12B). We also measured the mRNA abundance by qRT-PCR (Figure 1.12C). We observed inhibition of JMJD3 activity led to a significant downregulation of endoderm markers (*sox17* and *sox32*) whereas inhibiting the activity of LSD1 led to the upregulation of endoderm markers. Inhibiting the activity of G9a and MLL1 did not cause any effect on the gene expression and the number of $Sox17^+$ cells. In summary, our initial experimentation suggested that changing the chromatin landscape to a permissive state increases commitment towards the endoderm lineage whereas repressive state results in a reduced number of $Sox17^+$ endoderm cells. These results are in concordance with our previous observations that high chromatin accessibility (euchromatin) is required for endoderm lineage commitment whereas active remodelling to repressive state (heterochromatin) could be the underlying principle for the emergence of the mesoderm lineage. Our studies further highlight that activities of the demethylases play a crucial role during lineage segregation.

Figure 1.12 Perturbing activity of histone modifiers results in defective cell fate specification.



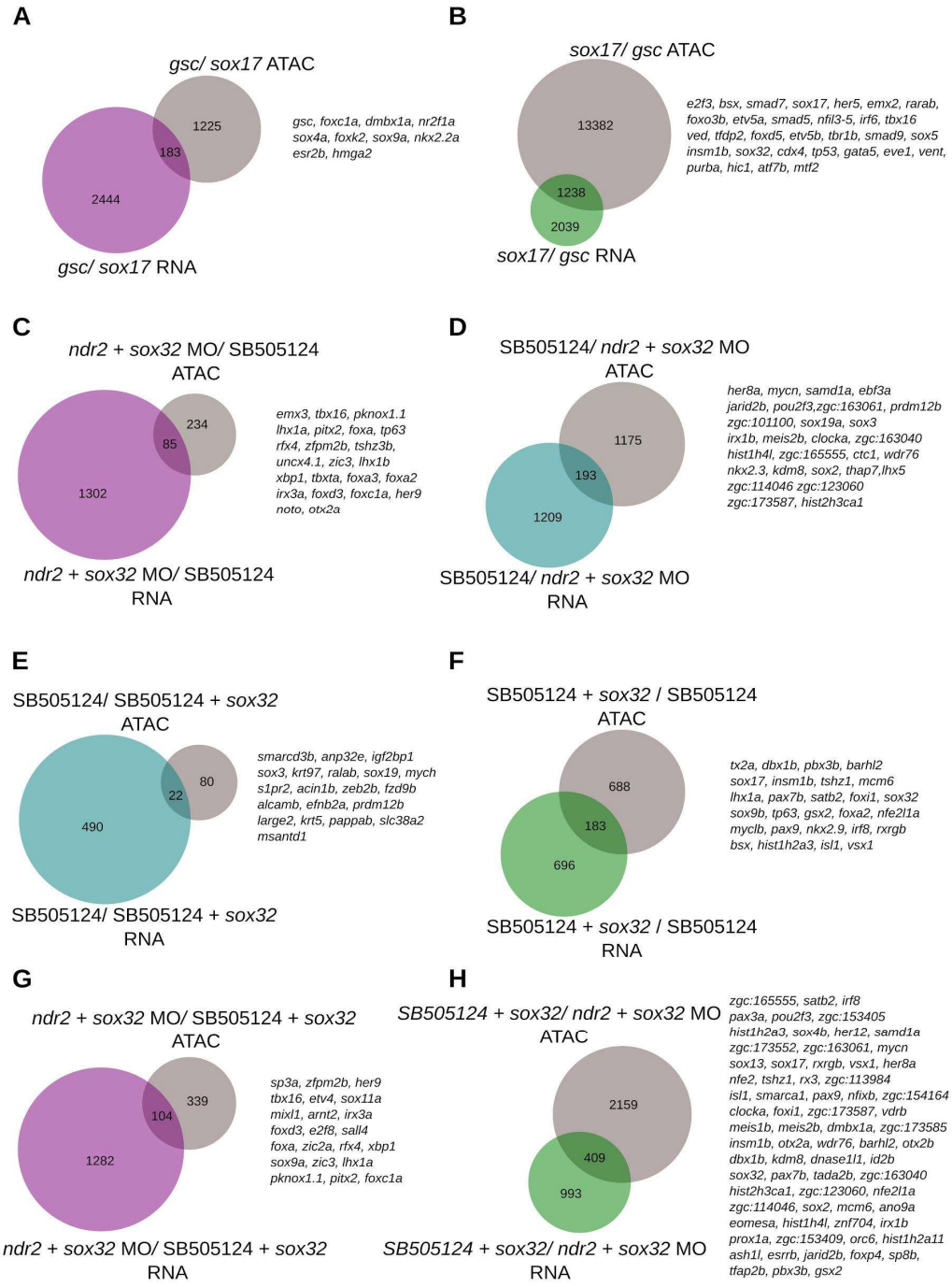
(A) Schematic representation summarizing various histone modifications on N-terminus tails of histones H2A, H2B, H3 and H4. Various writers (blue) and erasers (brown) are listed in brackets. **(B)** Sox17: GFP reporter activity in zebrafish embryos treated with various chemical inhibitors against specific histone modifiers visualized at 10 hpf. Change in chromatin state: activation is marked with green and repression is marked with red upon treatment with respective inhibitors. A bright green cluster of cells marks Kupffer's vesicle in a developing embryo. **(C)** Gene expression levels measured using qPCR for *gsc*, *ntl*, *sox17* and *sox32* in embryos treated with inhibitors GSK-J4 (JMJD3) GSK-LSD1 (LSD1), MM102 (MLL1) BIX-01294 (G9a). SB505124 (TGFß) was used as an internal control for the experiment. * indicates P-value < 0.01, ** P-value < 0.001, **** P-value < 0.0001, ns = non-significant as calculated by the Student's two-tailed t-test, N=2.

1.3.6 Identification of transcription factor complexes driving chromatin accessibility changes and lineage commitment

Chromatin accessibility can be either the driver of gene expression or could be the reflection of gene expression activities. Thus, understanding the correlation between mRNA abundance and chromatin accessibility is necessary. This analysis also helped to identify transcription factor complexes participating in chromatin remodelling. To this, we performed a Venn diagram analysis of genes associated with differential ATAC-seq peaks and differentially expressed genes. For instance, we compared genes that showed higher chromatin accessibility in endogenous Gsc^+ cells with the genes which are expressed at higher levels in a Gsc^+ population when compared to $Sox17^+$ cells (Figure 1.13A). Many Gsc^+ lineage-specific transcription factors such as *gsc*, *foxc1* emerged through this analysis. A similar analysis for endogenous $Sox17^+$ cells highlighted many differentially expressed endoderm determinants *sox17*, *ved*, *her5*, *smad5* and *gata5* etc.

Next, we repeated this analysis for induced populations. Analysis with induced mesoderm population over nodal treated (ectoderm like) embryos identified classical markers of mesoderm showing high accessibility and high mRNA abundance such as *tbx16*, *pitx2* (Figure 1.13C). Whereas reverse analysis revealed ectoderm drivers *sox3*, *sox19a* and *Pou2f3* were significantly enriched in ectoderm like cells (Figure 1.13D). Venn analysis between induced endoderm and induced ectoderm population (Figure 1.13E and 1.13F) as well as a comparison between induced endoderm over induced mesoderm allowed us to identify a bunch of endoderm determinants such as *sox17*, *sox2*, *eomesa* and *foxi1* (Figure 1.13G and 1.13H).

Figure 1.13 Correlation analysis between chromatin accessibility and transcriptional activity in endogenous and induced cell fate progenitors



(A) Correlation analysis for chromatin accessibility and transcriptome activity for genes showing higher enrichment in endogenous axial mesoderm compared to endoderm progenitors depicted by Venn diagrams. **(B)** Correlation analysis for genes showing

higher enrichment in endoderm progenitors compared to axial mesoderm progenitors. **(C and D)** Correlation analysis for genes showing higher enrichment in induced mesoderm progenitors and induced ectoderm respectively when comparing mesoderm (*ndr2* +*sox32* MO) and ectoderm (SB505124). **(E and F)** Correlation analysis for genes showing higher enrichment in induced ectoderm progenitors and induced endoderm respectively when comparing ectoderm (SB505124) and endoderm (SB505124 +*sox32* mRNA). **(G and H)** Correlation analysis for genes showing higher enrichment in induced mesoderm progenitors and induced ectoderm respectively when comparing mesoderm (*ndr2* +*sox32* MO) and endoderm (SB505124 + *sox32* mRNA). Representative genes showing positive correlation are listed next to each comparison.

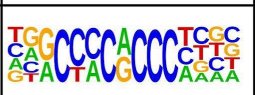
To support observations from the Venn analysis and to predict factors that have the potential to bind on differentially accessible regions, we performed motif enrichment analysis in induced mesoderm and induced endoderm cells respectively (Figure 1.14). We specifically opted for induced population keeping in mind to reduce background from transition state. Again, we identified a bunch of transcription factors whose binding sites were enriched in mesoderm such as Foxh1, Tbr1, Foxa, Foxa2, Eomes in mesoderm whereas Pou5f3, Sox2, Sox3, Sox17, Tead in endoderm population. Collectively, correlation of transcriptome and motif analysis from the ATAC-seq data allowed us to predict factors responsible for chromatin remodelling and further, would be interesting to validate *in vivo*.

Figure 1.14 Differential set of transcription factor motifs drive mesoderm and endoderm specification.

Mesoderm

Sr	Motif sequence	TF	p-value
1		Foxh1	$1e^{-43}$
2		Tbr1	$1e^{-35}$
3		Foxa	$1e^{-30}$
4		Foxa2	$1e^{-21}$
5		Foxm1	$1e^{-30}$
6		Eomes	$1e^{-24}$
7		Lhx2	$1e^{-14}$
8		Foxd3	$1e^{-18}$
9		Sox9	$1e^{-6}$
10		Lef1	$1e^{-9}$

Endoderm

Sr	Motif sequence	TF	p-value
1		Pou5f3	$1e^{-94}$
2		Sox2	$1e^{-92}$
3		Sox3	$1e^{-92}$
4		Sox17	$1e^{-79}$
5		Klf3	$1e^{-4}$
6		Tead1	$1e^{-16}$
7		Tead3	$1e^{-14}$
8		Sp5l	$1e^{-30}$
9		Otx2	$1e^{-4}$
10		YY1	$1e^{-3}$

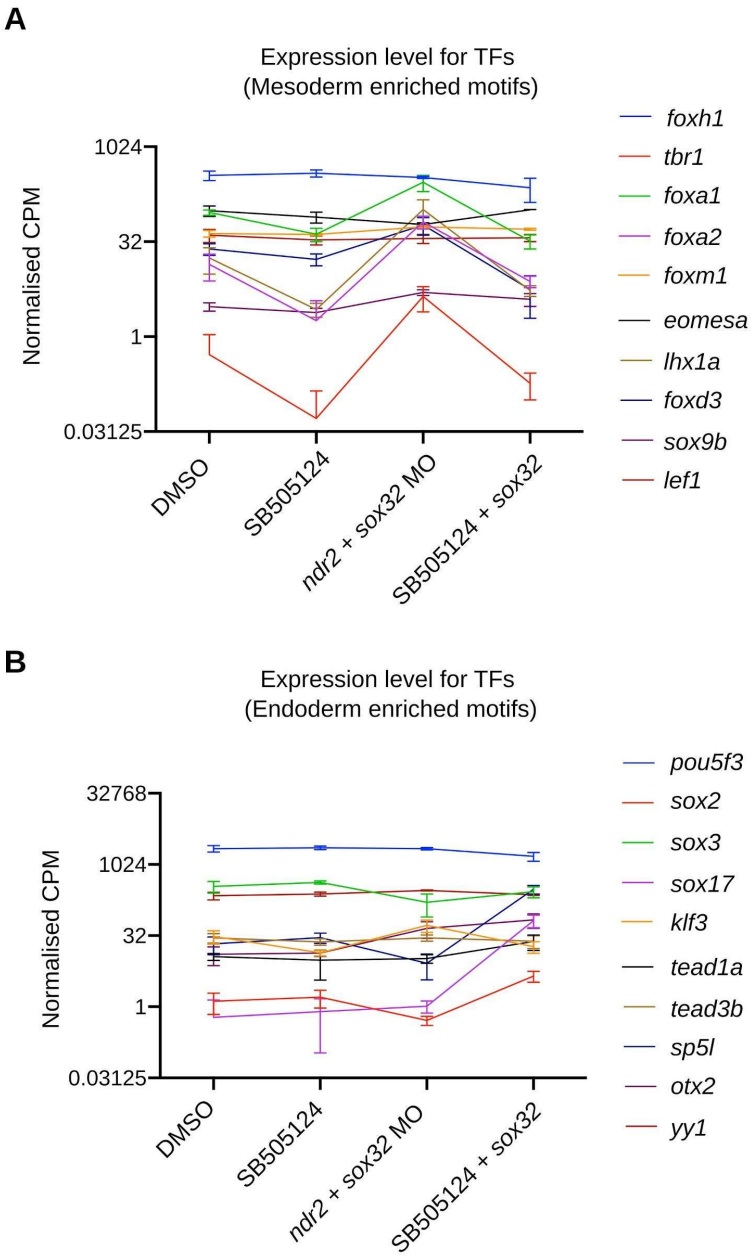
Motif enrichment analysis of gene loci showing higher chromatin accessibility in endogenous and induced mesoderm and endoderm progenitor cells. Enriched motifs are mapped to vertebrate databases for known factors. Top 10 motifs of transcription factors that are expressed at > 5 CPM (Counts per Million) in RNA-seq data under any condition are represented to account for background enrichment.

1.3.7 Nodal signalling regulates expression of factors involved in chromatin remodelling during lineage specification

Nodal signalling has been studied extensively and implicated during cell fate specification in vertebrate embryogenesis (Schier 2003). Mesoderm determinants *gsc*, *ntl* and endoderm determinants *sox17*, *sox32* are direct transcriptional targets of Nodal signalling through activation and binding of phosphorylated Smad2 at the core promoter (Feldman et al., 2000; Gritsman et al., 1999; Sako et al., 2016; Schier, 2003). Moreover, analysis of chromatin accessibility using ATAC-seq of the induced population suggested that overexpression of these nodal targets is sufficient to remodel chromatin and regulate target gene expression. In the process, we also predicted other potential regulators of chromatin accessibility.

To test if the predicted factors are under the regulation of Nodal signalling, we employed a similar experimental strategy to induce cell lineage as elaborated previously (Figure 1.10B) and analyzed mRNA abundance of factors that showed enrichment in mesoderm and endoderm respectively (Figure 1.15A and 1.15B). Gene expression analysis revealed dynamic transcriptional regulation of these factors by nodal signalling. As evident from the analysis, *foxa1*, *foxa2*, *lhx1a* and *tbr1* showed increased expression upon nodal induction while expression of these factors was significantly reduced when embryos treated with the nodal inhibitor, SB505124 (Figure 1.15A). However, other factors, although predicted to be enriched at differentially accessible regions, did not show expression changes upon nodal stimulation or perturbation at least at the mRNA level. Interestingly, none of the predicted factors showed reduced expression upon nodal inhibition. Notably, *sp5l*, *sox17*, *sox3* and *tead1a* showed an increase in the mRNA abundance levels when endoderm is induced (Figure 1.15B).

Figure 1.15 Expression dynamics of transcription factors exhibiting higher motif enrichment in mesoderm and endoderm progenitors upon induced cell fate.



(A) Line plots depicting average expression levels of mesoderm enriched *foxh1*, *tbr1*, *foxa1*, *foxa2*, *foxm1*, *eomesa*, *lhx1a*, *foxd3*, *sox9b* and *lef1* in wild-type and upon nodal inhibition (induced ectoderm), induced mesoderm (*ndr2* mRNA + *sox32* MO) and induced

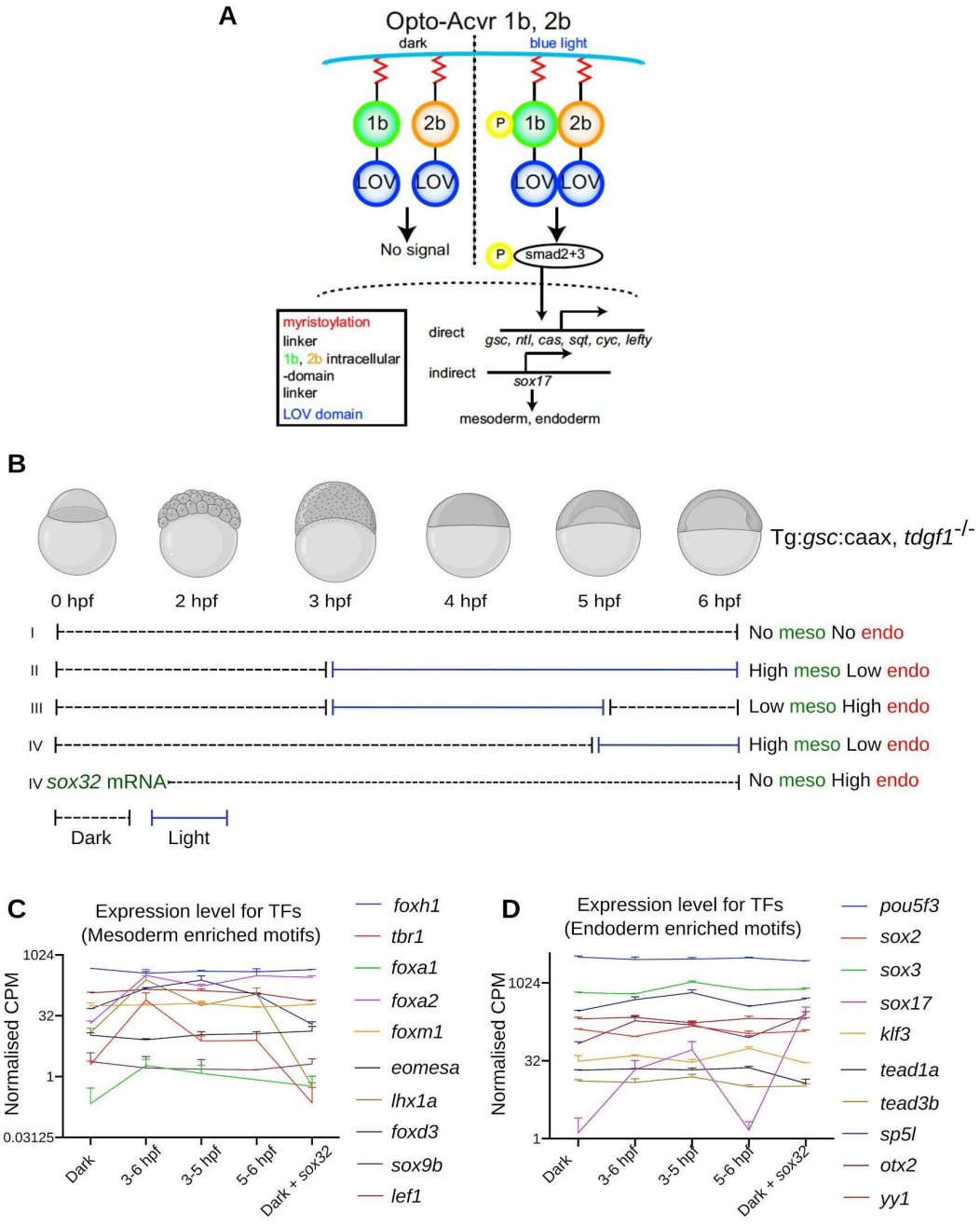
endoderm (SB505124 + *sox32* mRNA) at 4.5 hpf of developing embryo**. (B)** Line plots representing expression dynamics of endoderm enriched *pou5f3*, *sox2*, *sox3*, *sox17*, *klf3*, *tead1a*, *tead3b*, *sp5l*, *otx2* and *yy1*. Y-axis represented average normalized CPM values from three independent experiments (N=3). Y-axis is converted into a log2 scale for representation purposes. Error bar represents standard deviation**.**

To validate these observations using alternative strategies, we employed previously established methods to induce Nodal signals in a light-dependent manner (Sako et al., 2016). In this strategy, one eye pinhead (*oep*$^{-/-}$) mutant embryos were injected with activin receptors (Acvr1a and Acvr2b) fused to the LOV domain. *oep*$^{-/-}$ mutants were used to eliminate the effect of endogenous Nodal signalling. When exposed to blue light, nodal signalling is activated and smad2 gets phosphorylated to regulate downstream signalling (Figure 1.16A). Nodal activation can be achieved spatially or temporally.

Sako et al. suggested that different temporal activation windows of Nodal signalling induce different cell fates (Figure 1.16B). For instance, continuous exposure to blue light from 3 hpf to 6 hpf induces high levels of mesoderm and low levels of endoderm as observed in wild type conditions. Interestingly, limited early exposure from 3 hpf to 5 hpf induces high levels of endoderm cells whereas exposure from 5 hpf to 6 hpf induces high levels of mesoderm. We performed similar experimentation and also induced endoderm by overexpressing *sox32* in *oep*$^{-/-}$ embryos (Figure 1.16B).

Next, to test the expression dynamics of potential regulators of chromatin accessibility we performed bulk mRNA-seq analysis.

Figure 1.16 Optogenetic control of nodal signalling highlights nodal dependent transcription factor activity to induce mesoderm and endoderm progenitors.



(A) Schematic of nodal-optogenetic experimental strategy (reproduced from Sako et al. 2016). LOV domain fused activin receptors are activated only in presence of blue

wavelength light triggering a downstream cascade of nodal signalling inducing mesendoderm cell fate. **(B)** Schematic representation of the experimental strategy to modulate nodal signalling temporally to induce specific cell fate. Embryos obtained from a cross of MZ*oep*$^{-/-}$ and *gsc*:caax: MZ*oep*$^{-/-}$ when incubated under continuous dark conditions fail to induce mesendoderm fate. Exposure to blue light for 3-6 hpf leads to activation of genes responsible for both mesoderm and endoderm. Early exposure (3-5 hpf) leads to induction of endoderm progenitors whereas late exposure (5-6 hpf) induces mesodermal progenitors. Injections of *sox32* mRNA induce endoderm progenitors regardless of upstream nodal activity. **(C)** Line plots depicting average expression levels of mesoderm enriched *foxh1*, *tbr1*, *foxa1*, *foxa2*, *foxm1*, *eomesa*, *lhx1a*, *foxd3*, sox9b and lef1 under various light exposure conditions as depicted in the schematic. **(D)** Line plots representing expression dynamics of endoderm enriched *pou5f3*, *sox2*, *sox3*, *sox17*, *klf3*, *tead1a*, *tead3b*, *sp5l*, *otx2* and *yy1*. Y-axis represented average normalized CPM values from two independent experiments (N=2). Y-axis is converted into a log2 scale for representation purposes. Error bar represents standard deviation.

As suggested in earlier experiments, *tbr1*, *lhx1* and *foxa2* showed increased expression in optogenetically induced mesoderm population whereas expression levels reduced significantly when embryos exposed for 3 hpf to 5 hpf or when injected with *sox32* mRNA (Figure 1.16C). Similarly, *sox2*, *sox3* and *sox17* showed expected expression dynamics in endoderm progenitors.

Taken together, these results allowed us to narrow down the regulators of chromatin remodelling in the specification of both mesoderm and endoderm for future validation using ChIP-seq to gain further insights into the molecular mechanisms in detail.

1.4 Discussion

Coordinated series of morphogenetic events results in the formation of three distinct germ layers: ectoderm, mesoderm and endoderm during the vertebrate gastrulation. Comparative evolutionary studies suggested that mesoderm is the last germ layer to evolve as it is absent in cnidarians and ctenophores (Burke, 2007; Burton, 2008; Martindale et al., 2004; Ryan et al., 2013; Technau and Scholz, 2003). Thus, making it interesting to study the evolution of molecular mechanisms underlying the specification of mesendoderm.

A number of studies have been carried out to understand molecular interplays leading to mesendoderm formation using zebrafish, Xenopus and mouse models (Anderson et al., 2000; Bianchi and Sette, 2011; Keller, 1991; Li et al., 2010; Schier and Talbot, 2005). *In vitro* as well as *ex vivo* models have been established to decipher gene regulatory networks associated with lineage segregation (Martyn et al., 2019; Morgani et al., 2018; Townes and Holtfreter, 1955). Recently, reductionist approaches such as the explant cultures have also been utilized to systematically characterize the function of various signalling pathways during vertebrate gastrulation (Fulton et al., 2020; Schauer et al., 2020). Studies by Wagner et al., and Farrell et al., utilized newly developed single-cell transcriptome technologies for lineage tracing in developing zebrafish embryos (Farrell et al., 2018; Wagner et al., 2018). Similar studies have been conducted in other model systems allowing us to gain evolutionary perspectives underlying lineage segregation (Briggs et al., 2018; Hashimshony et al., 2015; Pijuan-Sala et al., 2019; Siebert et al., 2019). However, detailed characterization of specific lineages during gastrulation has been limited. Recently, Probst *et al.*, have characterized the role of Eomes-dependent segregation of anterior mesoderm and definitive endoderm during mouse gastrulation (Probst et al., 2021). Link *et al.*, utilized quantitative proteomics to identify differences between induced ectoderm and mesendoderm in developing zebrafish embryos (Link et al., 2006). These studies provided crucial indications towards the functions of cytoskeletal proteins such as Ezrin2 in mesendoderm segregation. However, our understanding from this study is limited due to lack of temporal information from endogenous populations and

also inability to differentiate between various subtypes of mesoderm such as axial mesoderm, paraxial mesoderm and true mesoderm (notochord).

To gain further insights, here, we have attempted to characterize lineage-specific transcriptome at the onset of gastrulation by utilizing various transgenic lines. Lineage-specific transcriptome analysis allowed us to identify novel expression patterns of various gene families indicating the possibility of their lineage-specific functions. Further, characterization using PCA analysis revealed interesting dynamicity of gene expression programs. Together, this data suggested that the mesoderm is re-wired to ectoderm or neuroectoderm like the population at the end of gastrulation, whereas the endoderm continues to segregate from other lineages probably to give rise to specialized tissues such as liver, heart, pharynx etc. Farrell *et al*. also predicted such re-wiring between subtypes of mesoderm during the specification of the notochord (Farrell et al., 2018). Importantly, combinatorial analysis of transcriptome at bulk and single-cell level enabled us to identify transient transcriptomic states of lineages before stages of segregation. Sako *et al*. used time-lapse imaging of double transgenics for *gsc* and *sox17* also indicated such transition states and showed the emergence of specific lineage during development. This study highlighted the crosstalk between mesoderm and endoderm and showed negative regulation between the mesoderm marker *gsc* and endoderm marker *sox17* (Sako et al., 2016).

Analysis of chromatin accessibility is a strong suggestive approach to understand the regulation of gene expression by modulating genome architecture than spatial cell-cell signalling. Thus, to identify such regulators of lineage specification, we performed genome-wide accessibility mapping using ATAC-seq for Gsc^{+} and $Sox17^{+}$ cells. Such analysis revealed chromatin to be more accessible in endoderm progenitor cells than mesoderm progenitors at the onset of gastrulation. We further validated these results by using various means to induce these populations. Interestingly, our analysis suggested that endoderm determinant Sox32 is sufficient to drive these changes at chromatin level even in the absence of upstream Nodal-TGFß signalling. Furthermore, the mesoderm

progenitor cells lose chromatin accessibility possibly through active chromatin remodelling.

Nodal signalling plays a crucial determinative role during mesendoderm specification in a developing embryo of multiple species (Flowers et al., 2004; Gritsman et al., 1999; Morokuma et al., 2002; Watanabe et al., 2014). Thus, in this study, we perturbed nodal signalling using chemical inhibitors as well as modulate levels of nodal signalling using optogenetic approaches (Sako et al., 2016). We performed high-throughput gene expression as well as chromatin accessibility analysis to predict various regulators responsible for a change in the chromatin landscape. We narrowed down our search to transcription factors such as Foxa2, Eomesa, Sox2, Pou5f3 and Tead factors. From these, Eomes has been shown to mediate change in chromatin state during specification of neuroectoderm and mesendoderm lineages using mouse embryonic systems (Tosic et al., 2019). Interestingly, Foxa2 has been implicated in endoderm-derived lineage specification during mouse gastrulation and is used as a marker for definitive endoderm (Burtscher and Lickert, 2009; Lee et al., 2019). However, multiple observations from our current study indicate Foxa2 to be a determinant of mesoderm rather than endoderm at least in zebrafish. Thus, our study also facilitates the understanding of the molecular interplays underlying species-specific differences. Moreover, Sox2 and Pou5f3 have been shown to regulate chromatin accessibility during zygotic genome activation in zebrafish (Miao et al., 2020). Although detailed experimental validation through ChIPseq studies for obtaining combinatorial binding patterns is required, our initial experiments enabled us to identify these potential effectors of the gene expression programs.

Taken together, the use of transgenic lines along with genetic manipulation collectively suggested active and differential chromatin remodelling is required for lineage commitment during gastrulation. Through this study, we also attempt to provide novel insights into the emergence of mesoderm specification in triploblastic organisms possibly by fine-tuning the existing gene expression programs. Importantly, based on the appearance of germ layers during evolution and initial results presented in this study that endoderm determinant Sox32 alone is sufficient to bring about the necessary changes in

chromatin landscape allowed us to postulate that endoderm is more likely to be the default fate during initial lineage commitment further to generate specialized mesoderm lineages, the system presumably invests a maximum amount of energy in re-wiring by orchestrating the local chromatin regulatory activities. Further comparative studies across phyla at the chromatin level would be insightful to validate these hypotheses. However, we believe that this study will act as a strong platform enabling researchers to obtain further insights into lineage specification during vertebrate development.

1.5 References

Alexanian, M., Maric, D., Jenkinson, S.P., Mina, M., Friedman, C.E., Ting, C.-C., Micheletti, R., Plaisance, I., Nemir, M., Maison, D. and Kernen, J. (2017). A transcribed enhancer dictates mesendoderm specification in pluripotency. Nat. Commun. *8*,1-19.

Anderson, R., Copeland, T.K., Schöler, H., Heasman, J., and Wylie, C. (2000). The onset of germ cell migration in the mouse embryo. Mech. Dev. *91*, 61–68.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646–3658.

Barone, V., Lang, M., Krens, G.S.F., Pradhan, S.J., Shamipour, S., Sako, K., Sikora, M., Guet, C.C., and Heisenberg, C.-P. (2017). An effective feedback loop between cell-cell contact duration and morphogen signaling determines cell fate specification during zebrafish gastrulation. Mech. Dev. *145*, S18.

Barrero, M.J., Boué, S., and Belmonte, J.C.I. (2010). Epigenetic Mechanisms that Regulate Cell Identity. Cell Stem Cell *7*, 565–570.

Barrero, M.J., Berdasco, M., Paramonov, I., Bilic, J., Vitaloni, M., Esteller, M., and Belmonte, J.C.I. (2012). DNA Hypermethylation in Somatic Cells Correlates with Higher Reprogramming Efficiency. Stem Cells *30*, 1696–1702.

Bedell, V.M., Westcot, S.E., and Ekker, S.C. (2011). Lessons from morpholino-based screening in zebrafish. Brief. Funct. Genomics *10*, 181–188.

Bedzhov, I., Graham, S.J.L., Leung, C.Y., and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. Philos. Trans. R. Soc. Lond. B Biol. Sci. *369,* 20130538.

Beisaw, A., Tsaytler, P., Koch, F., Schmitz, S.U., Melissari, M.-T., Senft, A.D., Wittler, L., Pennimpede, T., Macura, K., Herrmann, B.G.,and and Grote, P.. (2018). BRACHYURY directs histone acetylation to target loci during mesoderm development. EMBO Rep. *19*, 118–134.

Bianchi, E., and Sette, C. (2011). Post-Transcriptional Control of Gene Expression in Mouse Early Embryo Development: A View from the Tip of the Iceberg. Genes *2*, 345–359.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612–616.

Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science *360,6392*.

von Bubnoff, A., and Cho, K.W. (2001). Intracellular BMP signaling regulation in vertebrates: pathway or network? Dev. Biol. *239*, 1–14.

Burke, A.C. (2007). Development and evolution of the vertebrate mesoderm. Developmental Dynamics *236*, 2369–2370.

Burton, P.M. (2008). Insights from diploblasts; the evolution of mesoderm and muscle. J. Exp. Zool. B Mol. Dev. Evol. *310*, 5–14.

Burtscher, I., and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. Development *136*, 1029–1038.

Cao, C., Lemaire, L.A., Wang, W., Yoon, P.H., Choi, Y.A., Parsons, L.R., Matese, J.C., Wang, W., Levine, M., and Chen, K. (2019). Comprehensive single-cell transcriptome lineages of a proto-vertebrate. Nature *571*, 349–354.

Carmany-Rampey, A., and Schier, A.F. (2001). Single-cell internalization during zebrafish gastrulation. Curr. Biol. *11*, 1261–1265.

Chen, S., and Kimelman, D. (2000). The role of the yolk syncytial layer in germ layer patterning in zebrafish. Development *127*, 4681–4689.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods. 2017 Oct;14(10):959-962. doi: 10.1038/nmeth.4396.

David, N.B., and Rosa, F.M. (2001). Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling. Development *128*, 3937–3947.

Davidson, E.H. (1991). Spatial mechanisms of gene regulation in metazoan embryos. Development *113*, 1–26.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., and Diao, Y.(2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Donoghue, P.C.J., Graham, A., and Kelsh, R.N. (2008). The origin and evolution of the neural crest. BioEssays *30*, 530–541.

Dougan, S.T., Warga, R.M., Kane, D.A., Schier, A.F., and Talbot, W.S. (2003). The role of the zebrafish nodal-related genes squint and cyclops in patterning of mesendoderm. Development *130*, 1837–1851.

Erter, C.E., Solnica-Krezel, L., and Wright, C.V.E. (1998). Zebrafish nodal-related 2Encodes an Early Mesendodermal Inducer Signaling from the Extraembryonic Yolk Syncytial Layer. Dev. Biol. *204*, 361–372.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360,*6392.

Feldman, B., Dougan, S.T., Schier, A.F., and Talbot, W.S. (2000). Nodal-related signals establish mesendodermal fate and trunk neural identity in zebrafish. Curr. Biol. *10*, 531–534.

Ferretti, E., and Hadjantonakis, A.-K. (2019). Mesoderm specification and diversification: from single cells to emergent tissues. Curr. Opin. Cell Biol. *61*, 110–116.

Flowers, V.L., Courteau, G.R., Poustka, A.J., Weng, W., and Venuti, J.M. (2004). Nodal/activin signaling establishes oral--aboral polarity in the early sea urchin embryo. Dev. Dyn. *231*, 727–740.

Fulton, T., Trivedi, V., Attardi, A., Anlas, K., Dingare, C., Arias, A.M., and Steventon, B. (2020). Axis Specification in Zebrafish Is Robust to Cell Mixing and Reveals a Regulation of Pattern Formation by Morphogenesis. Curr. Biol. *30*, 3063–3064.

Fussner, E., Ahmed, K., Dehghani, H., Strauss, M., and Bazett-Jones, D.P. (2010). Changes in chromatin fiber density as a marker for pluripotency. Cold Spring Harb. Symp. Quant. Biol. *75*, 245–249.

Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., and Schier, A.F. (1999).

The EGF-CFC protein one-eyed pinhead is essential for nodal signaling. Cell *97*, 121–132.

Grusch, M., Schelch, K., Riedler, R., Reichhart, E., Differ, C., Berger, W., Inglés-Prieto, Á., and Janovjak, H. (2014). Spatio-temporally precise activation of engineered receptor tyrosine kinases by light. EMBO J. *33*, 1713–1726.

Guignard, L., Fiúza, U.-M., Leggio, B., Laussu, J., Faure, E., Michelin, G., Biasuz, K., Hufnagel, L., Malandain, G., Godin, C., et al. (2020). Contact area-dependent cell communication and the morphological invariance of ascidian embryogenesis. Science *369,*6500.

Hashimshony, T., Feder, M., Levin, M., Hall, B.K., and Yanai, I. (2015). Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. Nature *519*, 219–222.

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

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. doi: 10.1016/j.cell.2017.03.024. PMID: 28388407.

Hwang, W.Y., Peterson, R.T., and Yeh, J.-R.J. (2014). Methods for targeted mutagenesis in zebrafish using TALENs. Methods *69*, 76–84.

Kaaij, L., van der Weide, R. H., Ketting, R. F., & de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. Cell reports, 24(1), 1–10.e4. https://doi.org/10.1016/j.celrep.2018.06.003

Keller, R. (1991). Early embryonic development of Xenopus laevis. Methods Cell Biol. *36*, 61–113.

Keller, J.M., Escara-Wilke, J.F., and Keller, E.T. (2008). Heat stress-induced heat shock protein 70 expression is dependent on ERK activation in zebrafish (Danio rerio) cells. Comp. Biochem. Physiol. A Mol. Integr. Physiol. *150*, 307–314.

Kimelman, D., and Griffin, K.J. (2000). Vertebrate mesendoderm induction and patterning. Curr. Opin. Genet. Dev. *10*, 350–356.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and

histone demethylation. Nat. Rev. Genet. *7*, 715–727.

Kondo, M. (2007). Bone morphogenetic proteins in the early development of zebrafish. FEBS J. *274*, 2960–2967.

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

Le Douarin, N.M., Catala, M., and Batini, C. (1997). Embryonic neural chimeras in the study of vertebrate brain and head development. Int. Rev. Cytol. *175*, 241–309.

Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D. (2019). FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. Cell Rep. *28*, 382–393..

Li, M., Zhao, L., Page-McCaw, P.S., and Chen, W. (2016). Zebrafish Genome Engineering Using the CRISPR-Cas9 System. Trends Genet. *32*, 815–827.

Li, S., Quarto, N., and Longaker, M.T. (2010). Activation of FGF signaling mediates proliferative and osteogenic differences between neural crest derived frontal and mesoderm parietal derived bone. PLoS One *5*, e14033.

Link, V., Carvalho, L., Castanon, I., Stockinger, P., Shevchenko, A., and Heisenberg, C.-P. (2006). Identification of regulators of germ layer morphogenesis using proteomics in zebrafish. J. Cell Sci. *119*, 2073–2083.

Little, S.C., and Mullins, M.C. (2006). Extracellular modulation of BMP activity in patterning the dorsoventral axis. Birth Defects Res. C Embryo Today *78*, 224–242.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Maduro, M.F. (2010). Cell fate specification in the C. elegans embryo. Dev. Dyn. *239*, 1315–1329.

Martindale, M.Q., Pang, K. and Finnerty, J.R. (2004). Investigating the origins of triploblasty:mesodermal'gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa). *Development*, *131*,2463-2474.

Martyn, I., Siggia, E.D., and Brivanlou, A.H. (2019). Mapping cell migrations and fates in a gastruloid model to the human primitive streak. Development *146,17*.

McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science *353*, 6298.

Miao, L., Tang, Y., Bonneau, A.R., Chan, S.H., and Kojima, M.L. (2020). Synergistic activity of Nanog, Pou5f3, and Sox19b establishes chromatin accessibility and developmental competence in a context-dependent manner. bioRxiv.

Morgani, S.M., Metzger, J.J., Nichols, J., Siggia, E.D., and Hadjantonakis, A.-K. (2018). Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning. Elife *7,* e32839.

Morokuma, J., Ueno, M., Kawanishi, H., Saiga, H., and Nishida, H. (2002). HrNodal, the ascidian nodal-related gene, is expressed in the left side of the epidermis, and lies upstream of HrPitx. Dev. Genes Evol. *212*, 439–446.

Muñoz-Sanjuán, I., and Brivanlou, A.H. (2001). Early Posterior/Ventral Fate Specification in the Vertebrate Embryo. Developmental Biology *237*, 1–17.

Nelson, A.C., Cutty, S.J., Niini, M., Stemple, D.L., Flicek, P., Houart, C., Bruce, A.E.E., and Wardle, F.C. (2014). Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. BMC Biol. *12*, 81.

Nishida, H., and Stach, T. (2014). Cell lineages and fate maps in tunicates: conservation and modification. Zoolog. Sci. *31*, 645–652.

Nüsslein-Volhard, C. (2012). The zebrafish issue of Development. Development *139*, 4099–4103.

Pei, W., Noushmehr, H., Costa, J., Ouspenskaia, M.V., Elkahloun, A.G., and Feldman, B. (2007). An early requirement for maternal FoxH1 during zebrafish gastrulation. Dev. Biol. *310*, 10–22.

Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., and Reik, W.(2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495.

Pinheiro, D., and Heisenberg, C.-P. (2020). Zebrafish gastrulation: Putting fate in motion. Curr. Top. Dev. Biol. *136*, 343–375.

Pott, S., and Lieb, J.D. (2014). What are super-enhancers? Nat. Genet. *47*, 8–12.

Probst, S., Sagar, Tosic, J., Schwan, C., Grün, D., and Arnold, S.J. (2021). Spatiotemporal sequence of mesoderm and endoderm lineage segregation during mouse

gastrulation. Development *148*,1.

Ramel, M.-C., and Hill, C.S. (2012). Spatial regulation of BMP activity. FEBS Lett. *586*, 1929–1941.

Rankin, S.A., McCracken, K.W., Luedeke, D.M., Han, L., Wells, J.M., Shannon, J.M., and Zorn, A.M. (2018). Timing is everything: Reiterative Wnt, BMP and RA signaling regulate developmental competence during endoderm organogenesis. Dev. Biol. *434*, 121–132.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. *35*, W193–W200.

Rivera-Pomar, R., and Jäckle, H. (1996). From gradients to stripes in Drosophila embryogenesis: filling in the gaps. Trends Genet. *12*, 478–483.

Rodaway, A., and Patient, R. (2001). Mesendoderm. Cell *105*, 169–172.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983–994.

Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., Smith, S.A. and Putnam, N.H. (2013). The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution.Science, *342*,6164.

Sako, K., Pradhan, S.J., Barone, V., Inglés-Prieto, Á., Müller, P., Ruprecht, V., Čapek, D., Galande, S., Janovjak, H., and Heisenberg, C.-P. (2016). Optogenetic Control of Nodal Signaling Reveals a Temporal Pattern of Nodal Signaling Regulating Cell Fate Specification during Gastrulation. Cell Rep. *16*, 866–877.

Schauer, A., Pinheiro, D., Hauschild, R., and Heisenberg, C.-P. (2020). Zebrafish embryonic explants undergo genetically encoded self-assembly. Elife *9,* e55190.

Schier, A.F. (2003). Nodal signaling in vertebrate development. Annu. Rev. Cell Dev. Biol. *19*, 589–621.

Schier, A.F., and Talbot, W.S. (2005). Molecular genetics of axis formation in zebrafish. Annu. Rev. Genet. *39*, 561–613.

Schier, A.F., Neuhauss, S.C., Helde, K.A., Talbot, W.S., and Driever, W. (1997). The one-eyed pinhead gene functions in mesoderm and endoderm formation in zebrafish and interacts with no tail. Development *124*, 327–342.

Schneider, S., Steinbeisser, H., Warga, R.M., and Hausen, P. (1996). Beta-catenin translocation into nuclei demarcates the dorsalizing centers in frog and fish embryos. Mech. Dev. *57*, 191–198.

Shen, M.M. (2007). Nodal signaling: developmental roles and regulation. Development *134*, 1023–1034.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Siebert, S., Farrell, J.A., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., and Juliano, C.E. (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. Science *365,*6451.

Signolet, J., and Hendrich, B. (2015). The function of chromatin modifiers in lineage commitment and cell fate specification. FEBS J. *282*, 1692–1702.

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

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., and Junker, J.P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nat. Biotechnol. *36*, 469–473.

Stark, R., Brown, G., and Others (2011). DiffBind: differential binding analysis of ChIP-Seq peak data. R Package Version *100*.

Streisinger, G., Walker, C., Dower, N., Knauber, D., and Singer, F. (1981). Production of clones of homozygous diploid zebra fish (Brachydanio rerio). Nature *291*, 293–296.

Szutorisz, H., and Dillon, N. (2005). The epigenetic basis for embryonic stem cell pluripotency. Bioessays *27*, 1286–1293.

Szutorisz, H., Canzonetta, C., Georgiou, A., Chow, C.-M., Tora, L., and Dillon, N. (2005). Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage. Mol. Cell. Biol. *25*, 1804–1820.

Technau, U., and Scholz, C.B. (2003). Origin and evolution of endoderm and mesoderm. Int. J. Dev. Biol. *47*, 531–539.

Tirosh-Finkel, L. (2006). Mesoderm progenitor cells of common origin contribute to the head musculature and the cardiac outflow tract. Development *133*, 1943–1953.

Tosic, J., Kim, G.J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L. and Arnold, S.J. (2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. Nat. Cell Biol. *21*,1518-1531.

Townes, P.L., and Holtfreter, J. (1955). Directed movements and selective adhesion of embryonic amphibian cells. J. Exp. Zool. *128*, 53–120.

Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science *360*, 981–987.

Watanabe, H., Schmidt, H.A., Kuhn, A., Höger, S.K., Kocagöz, Y., Laumann-Lipp, N., Ozbek, S., and Holstein, T.W. (2014). Nodal signalling determines biradial asymmetry in Hydra. Nature *515*, 112–115.

Zinski, J., Tajer, B., and Mullins, M.C. (2018). TGF-β Family Signaling in Early Vertebrate Development. Cold Spring Harb. Perspect. Biol. 10, a033274.

Chapter 2

Characterizing the role of chromatin remodeler Satb2 during the early events of vertebrate gastrulation

2.1 Introduction

Cambrian explosion resulted in the appearance of major differences in body plans of organisms belonging to different phyla (Valentine et al., 1999). The evolution of such complex tissue structures and body plans has been achieved without a corresponding increase in the number of genes and is therefore attributed to the establishment of higher-order gene regulatory networks (GRNs) (Davidson, 2001, 2010; Davidson et al., 2002). GRNs are thought to be established through timely incorporation of novel gene families either *de novo* or by gene duplication and diversification and/or re-arrangement of existing genetic elements to gain new functions (Davidson, 2001). To unravel the molecular mechanisms underlying the evolution of these complex processes, identification and functional characterization of such co-evolved proteins is necessary. Dynamic chromatin organization is a key for differential tissue-specific gene regulation. One of the protein families which exhibits the potential to modulate chromatin landscape in a tissue-specific manner is special AT-rich binding proteins, SATB1 and SATB2. Interestingly SATB family proteins are highly conserved across vertebrates (Sheehan-Rooney et al., 2010) and therefore their function can be used as a paradigm to delineate the molecular principles of vertebrate evolution. Moreover, the use of early vertebrate systems such as zebrafish and Xenopus to study the function of SATB proteins would be instrumental in providing more insights into basic body plan determination.

2.1.1 Chromatin Organization and tissue-specific gene expression

The transcription of genes is predominantly regulated by cis-regulatory elements such as promoters and enhancers. CTCF and large molecular complexes like polycomb group (PCG), cohesin complex, MLL and SAGA complex have been extensively implicated in the regulation of gene expression during embryonic development by mediating long-range interactions (Andrey and Mundlos, 2017). A few reports have also suggested the role of long non-coding RNAs in mediating long-range interactions in human cells (Bonasio et al., 2010). However, how these long-range interactions are established was

not clear until the advent of high-resolution genomic studies such as HiC and ATAC-seq. Interestingly, HiC studies revealed long-range interacting regions of transcriptionally active and inactive domains which are called A/B compartments (Lieberman-Aiden et al., 2009). Moreover, these methods also revealed that certain chromatin elements, although separated by regions of low interactions (boundary elements), display a high degree of the association through internal interaction. They are often classified as topologically associating domains (TADs) and are shown to facilitate enhancer-promoter interactions (Gómez-Díaz and Corces, 2014). Internal interactions within TADs are often referred to as sub-TADs and interactions between multiple TADs results in Meta-TADs. It has been observed that these Meta-TAD structures play a much more significant role than individual TADs in tissue-specific gene regulation (Andrey and Mundlos, 2017; Zhan et al., 2017).

2.1.2 Long-range interactions and dynamic chromatin architecture during development

Interestingly, dynamic changes in methylation and histone modifications are observed more on the enhancer elements than on the promoter of developmentally regulated genes (Kaaij et al., 2016). Kaaij et al., in their study about characterizations of enhancers in zebrafish, reported an interesting observation that in contrast to studies in other mammalian species, active enhancers are hyper-methylated while hypo-methylated enhancer regions are mostly associated with transcription factors which are functional at later stages of embryonic development (Kaaij et al., 2016). 4C experiments suggested constant interaction of these enhancer elements with respective promoters throughout embryonic development are primed by H3K4me1. This observation makes overall regulation even more complex to understand but interesting to decode further. In light of this, it is important to understand how the genome is dynamically organized throughout development.

To this, a recent study from the same group, Kaaij et al., where they have generated HiC maps at various embryological stages of zebrafish contributed to a large extent (Kaaij et al., 2018). As seen in mammals, the zebrafish genome is also compartmentalized into

two: A (active) and B (inactive). Compartment A is gene dense, highly transcribed and associated with high enrichment of H3K4me3, H3K4me1 and H3K27Ac. It is also observed that TSS of constitutively and ubiquitously expressed genes are associated more with TAD boundaries. In this study, authors found that prior to ZGA, the genome is highly structured with the formation of topologically associating domains (TADs). However, along with the burst of zygotic transcription, the genome loses structural attributes. Lost chromatin architecture re-establishes as embryos develop and a completely structured genome can be seen again at 24 hpf highlighting the importance of higher-order structural changes required during the period of gastrulation. However, only limited knowledge is available about the identity of the extrinsic and intrinsic factor which regulate such dynamic chromatin architecture.

2.1.3 Chromatin remodelers

Chromatin organizers such as cohesin, condensin and CTCF are extensively associated with the higher-order genome organization. CTCF is known to be enriched at the TAD boundaries (Gómez-Díaz and Corces, 2014). However, from the experiments conducted by Kaaij et al., loss of chromatin architecture post-ZGA does not seem to be dependent on CTCF binding efficiency at the TAD boundaries (Kaaij et al., 2018). Studies in the mammalian system suggest cohesin be one of the crucial factors for TAD formation as loss of cohesin leads to a loss in TAD organisation (Gómez-Díaz and Corces, 2014). In zebrafish, cohesin subunit Rad21 is important for embryonic development. Depletion of Rad21 delays ZGA in zebrafish. Moreover, It has been observed that Rad21 binding sites overlap significantly with highly translated factor, Pou5f1, before ZGA (Meier et al., 2018). Although in mammals, these studies suggest the potential role of chromatin modifiers in development, more systematic studies need to be performed to understand this mode of regulation in a comprehensive manner.

2.1.4 SATB family proteins as chromatin organizers

In eukaryotic nuclei, chromatin is dynamically compartmentalized and organized based on functional compartments of the genome. Such higher-order chromatin architecture

facilitates the regulation of gene expression by allowing transcription factor accessibility and long-range enhancer function through chromatin looping. Special AT-rich Binding Protein (SATB) family of chromatin organizers, namely the homologs SATB1 and SATB2, regulate gene expression by binding to AT-rich sequences within the matrix attachment regions (MARs) and regulatory sequences at periodic distances and anchoring them to the nuclear matrix (Dickinson et al., 1997; Yasui et al., 2002). Both SATB1 and SATB2 have similar protein structures, with an N-terminal ubiquitin-like domain (ULD) to mediate the oligomerization of these proteins into dimers and tetramers, which allows subsequent binding to DNA segments through two CUT domains and one downstream Homeobox domain (Wang et al., 2012). The two homologs, SATB1 and SATB2 show high structural similarity, barring a few differences (Britanova et al., 2005; Naik and Galande, 2019).

SATB1, the first characterized homolog of the SATB family, can bind to either the histone deacetylase HDAC1 of the NuRD complex (corepressor) or PCAF HAT (coactivator) depending on its phosphorylation and acetylation states on the N-terminal domain (Kumar et al., 2006). SATB1 binding to MARs is dissociated as a result of its proteolysis by caspase 6 mediated cleavage upstream of the CUT domain (Galande et al., 2001). SATB1 is specifically expressed in thymocytes (Cai et al., 2003) and is involved in the repression of several genes during T cell development (Kohwi-Shigematsu et al., 1997), along with several other phenomena including regulation of embryonic stem cell differentiation through downregulation of pluripotency factors such as Nanog and Klf4 (Savarese et al., 2009) as well as regulation of tumour progression. SATB1 has been well associated with a range of cancer types (Naik and Galande, 2019)– predominantly breast cancer (Han et al., 2008) and colorectal cancer progression (Brocato and Costa, 2015; Mir et al., 2016).

2.1.5 Role of SATB2 during early embryonic development

SATB2 binds to multiple sites in the regulatory regions of genes and influences chromatin architecture thereby regulating target gene expression (Dobreva et al., 2003, 2006; Gyorgy et al., 2008). In humans, on chromosome 2, *SATB2* spans 191 kilobases of genomic DNA with 11 exons. SATB2 protein consists of 733 amino acids (predicted

molecular weight of 82.5 kDa) with two CUT domains located at amino acids 352-437 and 482-560, and a homeobox domain at amino acids 614-677 (FitzPatrick et al., 2003). Interestingly, the caspase cleavage site observed in SATB1 is not conserved in SATB2 (FitzPatrick et al., 2003), Instead, it possesses two independent tetrapeptide SUMO acceptor sites at two internal lysine residues (positions 233 and 350) (Dobreva et al., 2003). Together, these studies highlight differential mechanisms employed by SATB1 and SATB2 to achieve target functions. One of these lysine residues is located at a site reminiscent of the caspase cleavage site in SATB1, while the other lysine residue is positioned upstream of the CUT domains, which are conserved in both proteins (Dobreva et al., 2003). PIAS1 has been reported as the specific E3 SUMO ligase for SATB2 SUMO modification. The reversibility of sumoylation suggests that the regulation of SATB2 through SUMO modification is dynamic (Dobreva et al., 2003; Tan et al., 2010). Reports have shown that sumoylation of SATB2 reduces transcription activation and alters subnuclear localization. SATB2 is well expressed in immature B cells, brain and kidney, and is found to increase expression of immunoglobulin μ by binding to MAR sequences flanking the enhancer of the gene encoding its heavy chain (Dobreva et al., 2003).

SATB2 has also been implicated in the regulation of embryonic stem cells through interaction with Nanog and Klf4 (Savarese et al., 2009). In zebrafish, compromising Satb2 activity using morpholinos resulted in severe developmental defects including epiboly arrest which likely resulted from the aberrant exocytosis and endocytosis (Ahn et al., 2010). However, the role of SATB2 in cell fate specification events during the early stages of embryogenesis is lacking and necessitates exploration using alternative developmental models.

Here, we combined loss of function and genome-wide occupancy analysis for Satb2 to characterize its function(s) during early cell fate specification. We observed that loss of Satb2 impaired the specification of neuroectoderm progenitors and resulted in early developmental arrest. We also observed that expression levels of many key endoderm determinants were upregulated in the absence of Satb2. By comparing the genome-wide occupancy using ChIP-seq analyses we show that both neuroectodermal and

endodermal genes are direct targets of Satb2, suggesting Satb2 has the potential to act as both activator and repressor depending on its target genes and the cell populations in which it is active. Expression of Satb2 was restricted by increased Wnt signalling, highlighting novel mechanisms for involvement of the latter in promoting mesendodermal fates during zebrafish gastrulation.

2.2 Methods

2.2.1 Experimental models

Zebrafish (Danio *rerio*) were maintained as described previously (Westerfield, 2000). All the experimental procedures were carried out following the guidelines from the institutional animal ethics committees at IISER Pune and IST Austria. Embryos were raised at 28-31^{o}C in E3 medium and staged as described previously (Kimmel et al., 1995).

2.2.2 Phylogenetic analysis of SATB proteins

Homologous sequences for SATB proteins were extracted from GeneTree data from Ensembl (http://www.ensembl.org/Multi/GeneTree/Image?gt=ENSGT00390000008096) and were further curated using NCBI BLASTP in selected organisms for their true copy in their genome to avoid inclusion of splice variants. Curated sequences were aligned using MUSCLE with default parameters. Alignments were further trimmed using trimAl on auto mode which are embedded in the online server (https://ngphylogeny.fr) (Lemoine et al., 07 02, 2019). A phylogenetic tree was generated using the randomized accelerated maximum likelihood method (RAxML) method available at CIPRE Science Gateway server (Miller et al., 2012). Next, PROTCAT was used as a substitution model with DAYHOFF substitution matrix. DVE CAEL was used as an outgroup and analysis was performed with 1000 bootstraps. The resulting phylogenetic tree was visualized using iTOL and tree branches are colour coded based on different groups (Letunic and Bork, 2016). The representative animal's silhouette images were collected from (http://phylopic.org/) and used in the annotation.

2.2.3 Morpholino (MO) and RNA antisense oligonucleotide (ASO) injections

For loss of function studies using MOs and ASOs, embryos were arranged in agarose moulds and injections were performed at the 1 cell stage as described previously (Westerfield, 2007). Glass capillaries (WPI) were pulled using a needle puller (P-97, Sutter Instruments) and mounted on a microinjection system (PV820, World Precision Instruments). 4 ng of MO mix and 50 pg of ASO mix were used in each knockdown experiment.

The following morpholino sequences targeting translation initiation (MO1) and targeting exon splicing (MO2) were synthesized from Genetools and a 1:1 mixture was used for all knockdown experiments. Five base mismatch morpholino was used as a control for all the experiments. RNA antisense oligonucleotides were custom synthesized from IDT as described previously (Pauli et al., 2015). ASOs were injected as a mixture in a 1:1 ratio for knockdown experiments. Phenotypic changes were captured at indicated time points using an Olympus stereo microscope.

For rescue experiments, 5 pg of morpholino resistant Satb2 version was injected at 1 cell stage followed by injections of 4 ng morpholinos. To generate morpholino resistant clones, oligo containing 7 mismatches spanning the first 24 bases of the coding region were used for amplification using pCS2-*satb2* plasmid as a template.

satb2 MO1 (targeting translation initiation) (5'→3')	GACTCTCTCCTCCACCACGCTCCAT
satb2 MO2 (targeting splicing) (5'→3')	TGTGAAGTGCCTGATGAGAAAAGAA
satb2 mismatch control MO (control MO) (5'→3')	GAGTGTCTCGTCCACGACCCTCCAT
Satb2 Antisense oligonucleotide #1 (5'→3')	mU*mC*mU*mC*mC*U*C*C*A*C*C*A*iMe-dC*G*C*mU*mC*mC*mA*mU
Satb2 Antisense oligonucleotide #2 (5'→3')	mC*mA*mG*mC*mC*A*G*A*G*G*A*U*A*A*G*mG*mA*mC*mC*mU

Key: mA,mU,mC,mG: 2'O-methyl (2'OMe)-modified RNA nucleotides
iMe-dC: methylated deoxy-cytosine * phosphorothioate bond

2.2.4 mRNA sequencing and differential gene expression analysis

For Satb2 knock-down studies 4 ng of morpholino mix and 50 pg of RNA antisense oligonucleotides (ASO) mix were injected at 1 cell stage and embryos were harvested at 4.5 hpf. Approximately 500 ng of Total RNA with RIN > 9 was used as a starting material to prepare mRNA sequencing libraries using Ultra II RNA seq kit (NEB). All cDNA samples were amplified for 10-15 cycles depending on cycle number estimation by qPCR. Amplified libraries were purified using two rounds of 0.8X volume of Hi-prep PCR purification kit (Magbio Genomics, USA). The concentration of libraries was estimated using the Qubit DNA HS system (Thermo Scientific, USA). Finally, all the libraries were pooled and subjected to high throughput sequencing using 100 bp PE chemistry on Hiseq 2500 (Macrogen Inc, Korea).

2.2.5 Antibody production

All the procedures were performed as per the approved guidelines from the ethical committee at the national toxicology centre (NTC), Pune. To generate polyclonal antibodies, anti-zebrafish Satb2 (c-term): CIPSSGAEENPQANTGSGNNGP peptide was synthesized commercially (Apeptide, China). Antibodies were produced in New Zealand white rabbits, as per the protocols from the laboratory of Tony Hyman, MPI CBG, with modification as below (https://hymanlab.mpi-cbg.de/hyman_lab/general/). Briefly, the required amount of peptides were conjugated with KLH using glutaraldehyde followed by subsequent dialysis to remove glutaraldehyde. Conjugated peptides were mixed with Freud's complete adjuvant (Sigma Aldrich) for the first immunization. Rabbits were immunized intradermally. Further, after every 21 days, rabbits were immunized using peptides mixed with Freud's incomplete adjuvant until sufficient titre for the antibody was obtained.

Antisera were purified by a peptide affinity column prepared using sulfoLink coupling resin according to the manufacturer's instructions (Thermo Fisher Scientific) and stored in 50% glycerol solution at -20 º

2.2.6 Genome-wide occupancy analysis for zebrafish Satb2

I. ChIP setup:

To map the genome-wide occupancy of zebrafish Satb2, we performed ChIP-sequencing at desired stages of embryogenesis as mentioned in the results section and figure legends. To validate the use of in-house raised antibody for successful ChIP-seq experiment, we injected embryos with 25 pg of 3xFLAG-Satb2 at 1 cell stage and harvested embryos at 4.5 hpf for further processing. ChIP was performed using anti-FLAG antibody (Sigma) and anti-zebrafish Satb2 antibody (generated in-house as described above) and subjected to high throughput sequencing as described below. A correlation matrix was generated to confirm the reproducibility of the two methods. For all stage-specific ChIP experiments, anti-zebrafish Satb2 was used to capture endogenous genomic binding regions.

Briefly, ChIP was performed using a modified protocol (Leichsenring et al., 2013). Approximately, 2000 embryos per stage were harvested in 1x E3 medium and homogenized using glass homogenizer loose piston in the presence of 1mM PMSF and immediately fixed with 1% methanol free formaldehyde (Thermo) at room temperature for 12 minutes with constant shaking. Fixation was stopped by the addition of glycine to a final concentration of 0.125 M and incubation at room temperature for an additional 5 min. Fixed cells were centrifuged at 500x g for 5 min and washed thrice with 1X ice-chilled PBS by centrifugation at 4^{o}C for 5 min each. Cells were further lysed in a 6 times bed volume cell lysis buffer (10 mM Tris-HCl pH7.5, 10 mM NaCl, 0.5% (v/v) NP-40) for 10 minutes on ice and subjected to homogenization in Dounce homogenizer for 10 times with loose piston followed by 3 times with tight piston. Nuclei were collected by centrifugation at 2500x g, washed with ice-cold PBS and resuspended in 8 times the pellet volume in the nuclei lysis buffer (50 mM Tris-HCl pH7.5, 10 mM EDTA, 1% (w/v) SDS, protease inhibitor cocktail (Roche,11873580001) and incubated further on ice for 30 minutes. The samples were sonicated using the following Covaris S2 sonication conditions: 20% duty cycle, intensity= 5, cycle per burst= 200, Time= 40 cycles of 30 seconds ON 30 seconds OFF to obtain an average size of 200-300 bp. The samples were centrifuged at 12000 rpm for 10 minutes and the supernatant containing chromatin was stored at -80^{o}C till further use. Prior to the ChIP setup, the supernatant was precleared using a mixture of 1:1 Dynabeads protein A: G beads for 2 hrs at 4^{o}C. For each ChIP replicate and Mock reaction, 100 µg of chromatin was diluted 1:10 with ChIP dilution buffer (16.7 mM Tris-HCl pH7.5, 167 mM NaCl, 1.2 mm EDTA, 0.01% (w/v) SDS, 1X PIC) and incubated with either 5 µg of anti-FLAG antibody or 20 µg of anti-Satb2 antibody was used. For Mock ChIP reactions, an equal amount of Rabbit IgG (cat number #31235 Invitrogen) was used.

The chromatin-antibody complex was incubated overnight at 4°C with constant rotating. Following day, Chromatin-Antibody-complex was captured using 100 µl pre-blocked (with IgG-free BSA and t-RNA) Dynabeads® Protein A: G mix for 4 hrs at 4^{o}C. Beads were washed with ice-cold buffers in the following order: 7 minutes 4 times with low salt buffer (20 mM Tris HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), 10 minutes twice with high salt buffer (20 mM Tris HCl pH 8.0, 200 mM NaCl, 2 mM EDTA,

0.1% SDS, 1% Triton X-100), 10 min once with LiCl buffer (0.25 M LiCl, 1mM EDTA, 10 mM Tris HCl pH 8.0, 1% NP-40, 1% Sodium deoxycholate) and 10 minutes twice with TE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA). The TE buffer was removed completely and 150 µL of the elution buffer (0.1 M NaHCO3, 1% SDS) was added and vortexed gently to solubilize the beads, followed by incubation at 65^{o}C for 30 min at 1000 rpm. The eluate was collected in a fresh tube and the process was repeated with the addition of 150 µL of the elution buffer. To 300 µL of the eluates and 10% input, 20 µL of 5 M NaCl and 2 µL of RNAseA (10mg/ml) was added and the samples were incubated at 65^{o}C overnight with constant shaking at 700-800 rpm. Next, 20 µL of 1 M Tris pH 8.0, 10 µL of 0.5 M EDTA and 2 µL of Proteinase K (20 mg/ml) was added and the samples were further incubated at 42^{o}C for 1 h at 700-800 rpm. Samples were purified by standard phenol: chloroform extraction method and DNA was precipitated overnight with an equal volume of 100% isopropanol in the presence of Glycoblue (Ambion) at -20^{o}C. DNA pellets were eluted in nuclease-free water and quantified using Qubit 4 fluorometer (Thermo Scientific) before proceeding with further analysis.

II. Library preparation, sequencing and data analysis:

An equal amount of DNA (~5 ng) was used as an input for library preparation and libraries were prepared using Nugen ultra2 ovation kit (Nugen, Austria GmbH). The number of cycles for amplification of adapter-ligated libraries was estimated by the qPCR method as described in the datasheet provided by the manufacturer. Final libraries were purified using HiPrep PCR clean up system (Magbio Genomics, USA). Library concentration was determined using Qubit and average fragment size was estimated using DNA HS assay on bioanalyzer 2100 (Agilent Technologies, USA) before pooling libraries at equimolar ratio. Sequencing reads (100 bp PE) were obtained on the HiseqX platform at Macrogen Inc, Korea.

Sequencing reads were trimmed using TrimmomaticPE for Truseq2:PE adapters and read with quality greater than Phred 33 were retained (Bolger et al., 2014). The quality of sequencing reads was determined using fastQC. High-quality sequencing reads were aligned to zebrafish danRer10 genome version using default parameters of BWA (Li and Durbin, 2009). Aligned reads were subsampled to 40 million for each sample using

BBmap. Correlation between each replicate was estimated using multiBamSummary and replicates showing very high Pearson correlation (> 0.7) were used for further analysis. Peak calling was performed using macs2 with default parameters and q value 0.05. A consensus set of peaks from biological replicates were extracted using a custom R script from Roman Cheplyaka.

Bigwig files were generated first using bamCoverage normalizing to RPKM and then subtracting Input signals using bamCompare utilities from deepTools 3.3.2 (Ramírez et al., 2016) and used for visualization with Integrative Genomics Viewer (IGV). Satb2 peaks in regulatory genes were extracted by intersecting with Ensembl of ATAC seq peaks (GSE106428, GSE130944, GSE101779). Promoter bound peaks were assigned as +/- 2 Kb from the TSS of the genes. Peak annotation to the nearest gene was performed using annotate.pl utility from Homer. Motif discovery for a given set of peaks was performed using findMotifsGenome.pl script from Homer.

2.2.7 Inhibitor treatment

Wild-type AB/TU embryos were dechorionated and incubated in media containing either 3 μM BIO (GSK3β inhibitor) or 30 μM of SB505124 (Nodal inhibitor) from 2-4 cell stage. Embryos were grown at 28 ^{o}C till the desired developmental stage is reached.

2.2.8 Gene expression analysis using qPCR

Embryos were collected at the 6 hpf and were lysed in RNA iso plus (DSS-Takara, India) to extract total RNA. Each sample consisted of at least 10 embryos per experiment, and all experiments were repeated at least three times independently otherwise mentioned in the figure legends. cDNA was synthesized using PrimeScript RT kit (DSS-Takara, India), following the manufacturer's instructions.

Quantitative real-time PCR was performed using TB green premix ExTaq II (DSS-TAKARA) using gene-specific primers as listed below on ViiA7 thermal cycler (Applied Biosystems). Changes in threshold cycles were calculated by subtracting the Ct values of the gene of interest from that of housekeeping control used for qRT-PCRs [Ct(target

genes) – Ct(*ef1a*)]. ΔCt values of specific target genes from the experimental samples were then subtracted from their respective control samples to generate ΔΔCt values. The fold changes were calculated using the formula 2^ (-(ΔΔCt).

Target site	Forward primer (5'→3')	Reverse primer (5'→3')
Danio_satb2	CAAGAGTTTGGTCGCTGGTA	CGCTGGGCTAATACACAGAA
Danio_ef1a	CTTCTCAGGCTGACTGTGC	ACGATCAGCTGTTTCACTCC
Danio_chordin	TCTGCAAATTCGGGAAAAAC	GTTTCTGCAGCTCAGCAGTG
Danio_axin2	AGAAGATGACCCACGTCCAC	CCCGAGAAGAAAATGCAAAG

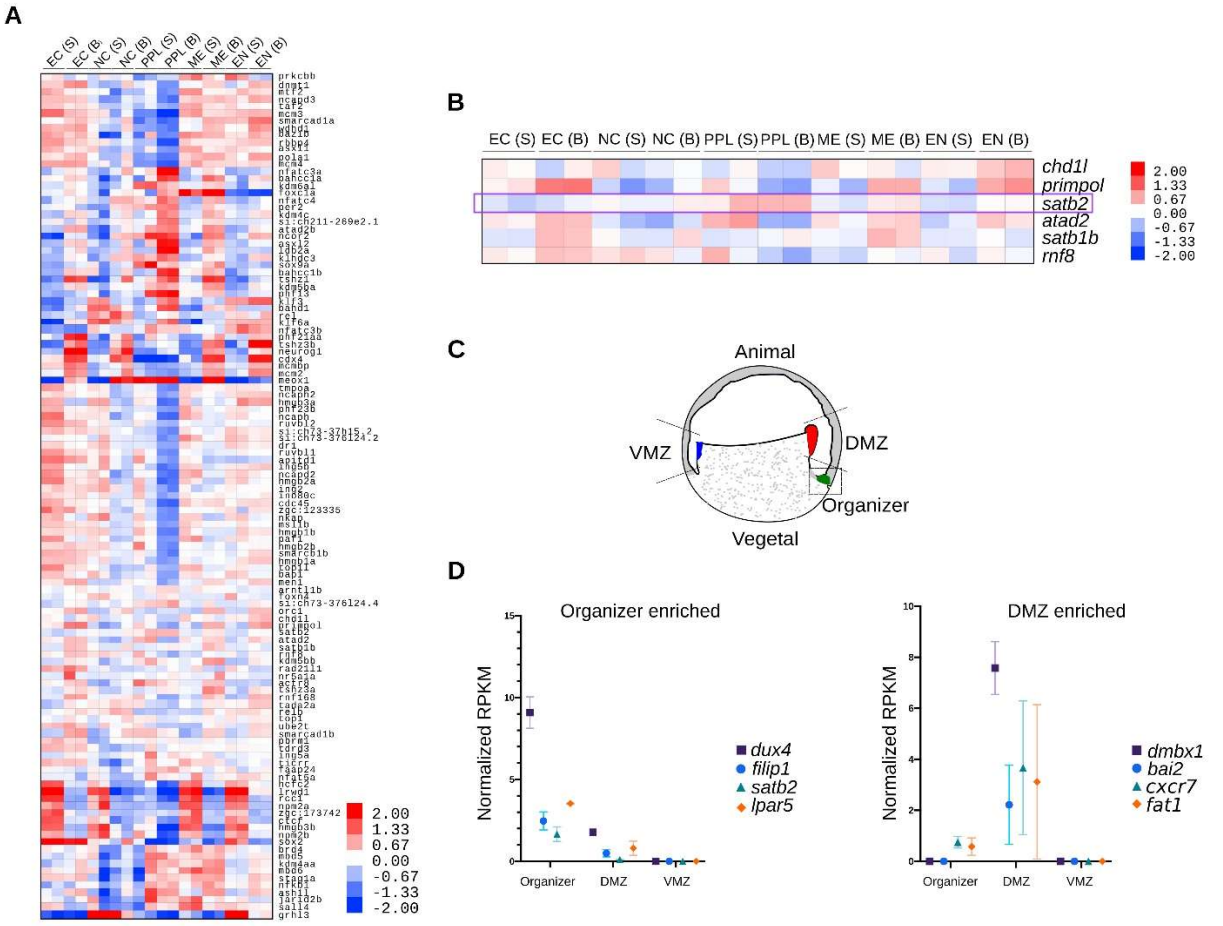
2.3 Results

2.3.1. Identification of lineage enriched chromatin organizers during gastrulation

In a previous section (Chapter 1), we generated lineage-specific transcriptome and characterized chromatin landscape during gastrulation in zebrafish. These studies also highlighted the importance of active chromatin remodelling during germ layer segregation. However, chromatin organizers which have the potential to remodel the chromatin landscape during the process of germ layer segregation remain unidentified. To explore this, we extracted a list of factors that had been previously implicated in chromatin remodelling by using the query GO term 'Chromatin remodelling' (GO:0006338) and clustered them in a lineage-specific manner (Figure 2.1A).

Interestingly, we observed enrichment for one such factor *satb2* in the prechordal plate (PPL) tissue (Figure 2.1B). Notably, *satb2* shows higher specificity compared to the other known chromatin organizers and is expressed at higher levels only in PPL. However, the closely related family member *satb1a* is expressed in multiple cell lineages such as ectoderm, notochord and paraxial mesoderm. Both Satb1 and Satb2 are known to regulate chromatin landscape by modulating long-range loop extrusions (Cera et al., 2019; Gyorgy et al., 2008). However, we did not observe such lineage-specific enrichment for another family member *satb1a*. This suggested that during gastrulation, *satb2* may have a specific function in conferring prechordal plate identity, a tissue structure equivalent to a dorsal organizer in Xenopus.

Figure 2.1 *satb2* is highly enriched in the dorsal organizer region in zebrafish and Xenopus.



(A) Lineage-specific expression patterns for chromatin organizers and chromatin remodelers during stages of gastrulation (GO:0006338). (B) Enlarged snapshot of selected region highlighting expression pattern of *satb2* and *satb1* in sorted populations as indicated. Scale bar represents a normalized score based on counts per million values from bulk mRNA-seq experiments (blue: low to red: high). Two independent experiments are clustered next to each other highlighting consistency between the replicates (N=2). (C) Schematic representation of a sagittal section of *Xenopus laevis* embryo at the gastrulation stage. The Blue region demarcates the ventral marginal zone (VMZ), Red region represents the dorsal marginal zone (DMZ) and the green region denotes dorsal organizer. (D) Normalised expression levels (RPKM) of selected genes enriched in the

organizer and dorsal marginal zone (DMZ). Values in the ventral marginal zone (VMZ) are used as a negative control.

In 2018, Popov *et al*. characterized tissue-specific transcriptome in Xenopus *laevis* (Figure 2.1C) and identified *satb2* as a previously uncharacterized, enriched factor in dorsal organizer when compared to dorsal marginal zone (DMZ) and ventral marginal zone (VMZ) (Figure 2.3D). Together, these findings suggest a possible conserved role of Satb2 during vertebrate embryogenesis.

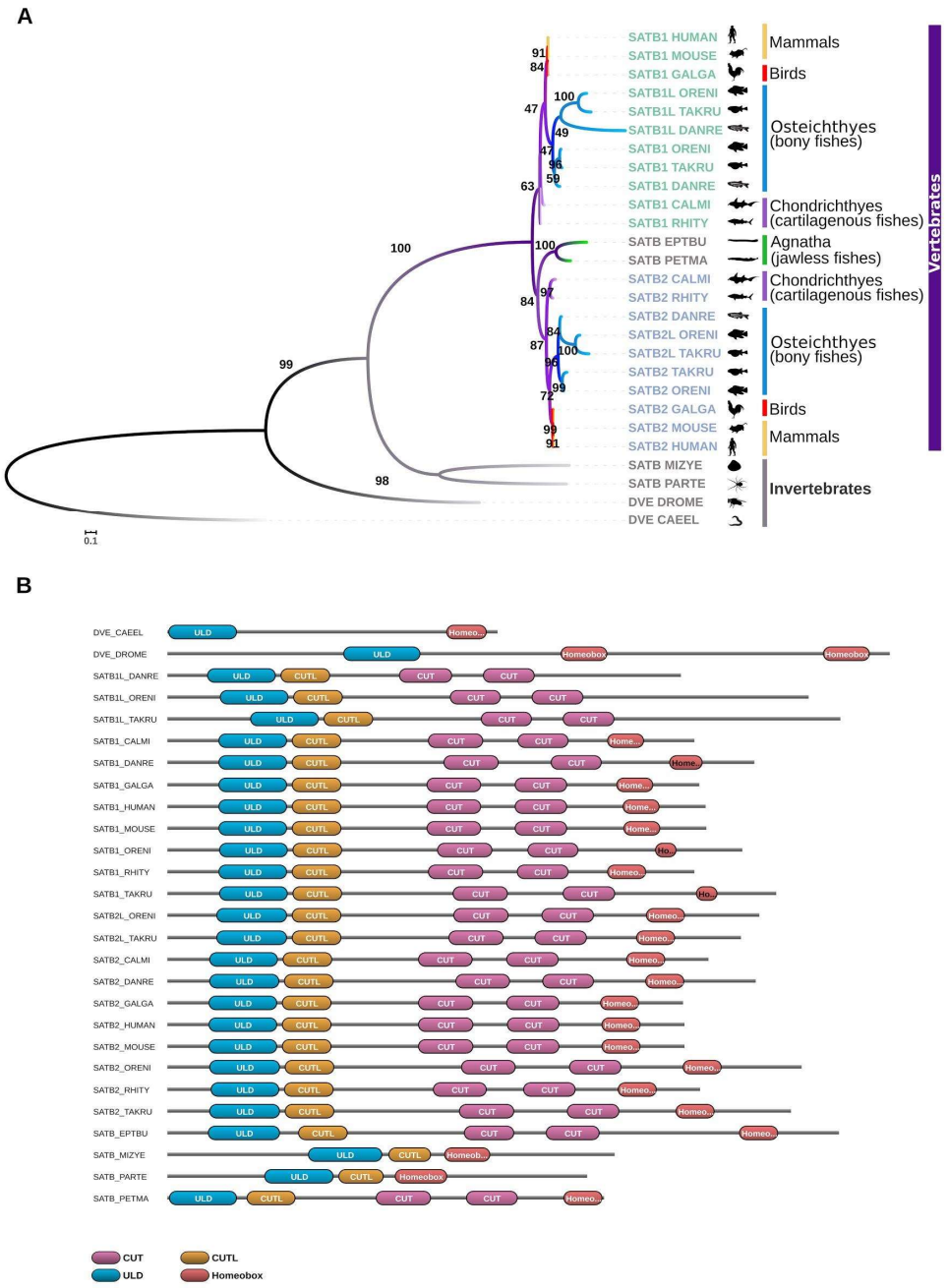
2.3.2 SATB family proteins diverged with the evolution of jawed fish and SATB2 is an ancient member of the family

To gain detailed insights into possible structural conservation among the SATB family proteins, we extracted the homologous sequence stretches by comparing the individual protein sequences. These sequences were curated based on their genomic copy numbers. This step was essential to exclude splice variants and make use of sequences resulting only from genome duplication for further analysis. SATB family proteins were earlier thought to be restricted only to the vertebrate lineage (Bürglin and Affolter, 2016; Bürglin and Cassata, 2002; Takatori and Saiga, 2008). However, a comprehensive understanding of the evolution and divergence of the SATB proteins (SATB1 and SATB2) is lacking. A distantly related homolog of SATB, Dve was identified in the fruit fly and roundworm and was classified into COMPASS superclass of proteins as they do not possess characteristic CUT domain (Bürglin and Cassata, 2002; Fuß and Hoch, 1998; Nakagoshi et al., 1998). Interestingly, our phylogenetic analysis revealed the presence of SATB homologs possessing all three types of functional domains in invertebrates, *Parasteatoda tepidariorum* (spider) and *Mizuhopecten yessoensis* (A molluscan-scallop) (Figure 2.2A). These homologs also contain CUT domains (Figure 2.2B). However, the interdomain distance is minimal in these organisms, the functional significance of which is not yet known. Agnatha, a class of vertebrate jawless fishes, exhibit a single form of SATB proteins and the presence of both SATB1 and SATB2 proteins is apparent only from jawed fish onwards.

Another noteworthy observation is the duplication of SATB1 and SATB2 sequences in most of the cartilaginous and bony fishes corroborating with the 2R genome duplication. The only exception being a single copy of Satb2 in zebrafish and this could be due to a selective loss in this organism. This particular observation can be further explained by allotetraploidization, a mechanism of asymmetrical paralogue retention during vertebrate genome duplication (Simakov et al., 2020). Collectively, our phylogenetic analysis suggests that ancestral SATB proteins originated in invertebrates and diverged into SATB1 and SATB2 with the evolution of jawed fish.

Moreover, our extensive domain architecture analysis revealed zebrafish Satb1 protein harbours a putative homeobox domain whereas Satb2 possesses a conserved homeobox. Based on the phylogenetic affinities, homologs of SATB proteins observed in invertebrates and jawless fishes exhibit more similarity with SATB2, suggesting amongst the two, SATB2 is closer to invertebrate homologs and therefore may play a significant role in the initial developmental transitions or body plan determination. Hence, we further focused our study on dissecting the molecular function of SATB2 during lineage segregation.

Figure 2.2 Phylogenetic and domain architecture analysis of SATB proteins.



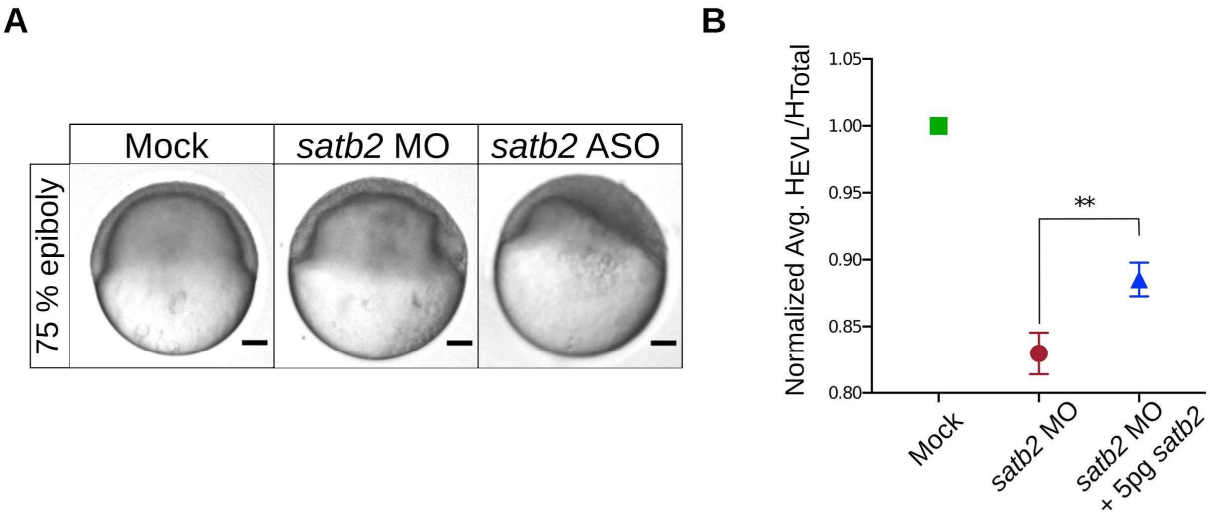
(A) Phylogenetic tree depicting evolution and divergence of SATB family proteins. SATB1 homologs are highlighted in green, SATB2 in blue whereas ancestors of SATB proteins are in grey. Organisms belonging to different classes are highlighted with different colours. Numbers on branching points represent bootstrap values. Organism labels;

HUMAN- *Homo sapiens*, MOUSE- *Mus musculus*, GALGAL- *Gallus gallus*, ORENI- *Oreochromis niloticus*, TAKRU- *Takifugu rubripes*, DANRE- *Danio rerio*, CALMI- *Callorhinchus milii*, RHITY- *Rhincodon typus*, EPTBU- *Eptatretus burger*, PETMA- *Petromyzon marinus*, MIZYE- *Mizuhopecten yessoensis*, PARTE- *Parasteatoda tepidariorum*, DROME- *Drosophila melanogaster* and CAEEL- *Caenorhabditis elegans*. Additional copies of SATB1 and SATB2 were denoted with SATB1L (SATB1-like) and SATB2L (SATB2-like). **(B)** Comparative analysis of domain architecture of SATB1 and SATB2 for various species across evolution. ULD domain is represented in blue, CUT-like domain in yellow, CUT domains in purple and Homeobox domain in red. Homeobox domain with black text indicates putative or predictive Homeobox.

2.3.3 Satb2 is required during the early developmental window in zebrafish

Previous reports have suggested that Satb2 is maternally deposited and the loss of Satb2 leads to early developmental defects such as epiboly arrest (Ahn et al., 2010). However, the precise function of Satb2 in cell lineage commitment has been largely unexplored. To explore this, we chose to conduct loss of function studies using transient knockdown methods. We employed two different strategies by injecting morpholinos (MOs) and RNA antisense oligonucleotides (ASOs) at 1 cell stage of zebrafish embryos. Both these strategies differ in their mode of action. Morpholinos primarily block translation or splicing, whereas RNA antisense oligonucleotides are single-strand RNA-DNA hybrids that have the potential to degrade transcripts by that inducing catalytic degradation of complementary RNAs via RNase H in a cellular fraction (Pauli et al., 2015).

Figure 2.3 Loss of Satb2 function leads to developmental arrest during zebrafish gastrulation.



(A) Lateral view of zebrafish embryos at 75% epiboly stage injected with 4 ng of morpholinos and 50 pg of RNA antisense oligonucleotide against *satb2*. **(B)** Dot plot representing a ratio of the average height of EVL from animal pole to the total height of the embryo normalized to Mock control (n=30). ** signifies P-value < 0.005 as determined by student's t-test.

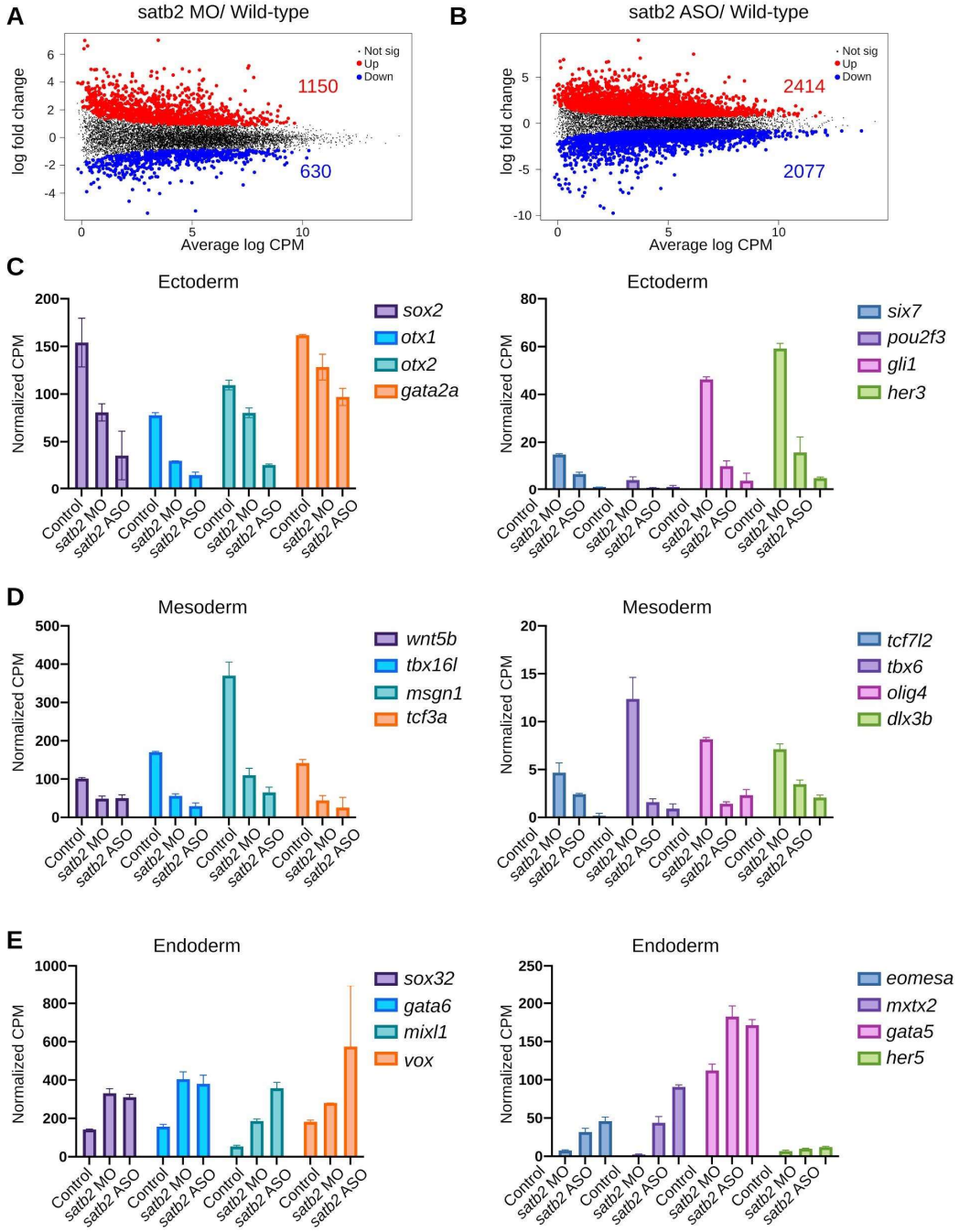
Both these approaches led to the loss of function of Satb2 resulting in similar defects in zebrafish embryos. In concordance with previous reports, Satb2 loss resulted in a developmental arrest at epiboly stages (Figure 2.3A). The effect of satb2 knockdown using morpholinos can be rescued partially by co-injecting morpholino resistant versions of zebrafish *satb2* mRNA (Figure 2.3B). It is also observed that the effect of knockdown using ASOs was much stronger than the use of MOs presumably due to the effect of a transcriptional hindrance.

2.3.4 Satb2 exhibits differential regulation on different germ layers

To investigate the cause underlying these developmental defects in detail, we performed gene expression profiling of embryos injected with MO mix and ASO mix and compared it with that of control embryos. As previously observed with developmental phenotypes, the effect on gene expression was higher in embryos injected with ASO mix both in terms of a number of deregulated genes and an extent of deregulation (Figure 2.4A and 2.4B). the mRNA-seq analysis further revealed an interesting pattern of gene expression regulation by Satb2. We observed reduced abundance for genes involved in neural ectoderm and mesoderm whereas key marker genes for endoderm were significantly upregulated upon Satb2 depletion.

To exemplify, ectodermal and neural ectoderm markers such as *sox2, otx1, otx2, gata2a, six7, pou2f3, gli1* and *her3* showed a significant reduction in the level of RNA abundance (Figure 2.4C). Mesodermal markers such as *wnt5b, tbx16l, msgn1, tcf3a, tcf7l2, tbx6, olig4* and *dlx3b* also showed a significant decrease in the level of RNA expression upon loss of Satb2 (Figure 2.4D). However, endodermal genes *sox32, gata6, mixl1, vox, eomesa, mxtx2, gata5* and *her5* were significantly upregulated upon injection with MOs and ASOs (Figure 2.4E). Together, these observations strengthen the notion that Satb2 plays an important function in the process of cell fate specification by regulating key lineage specification genes during zebrafish gastrulation.

Figure 2.4 Satb2 is required for specification of neural ectoderm and mesoderm fate and prevents commitment to endoderm lineage.

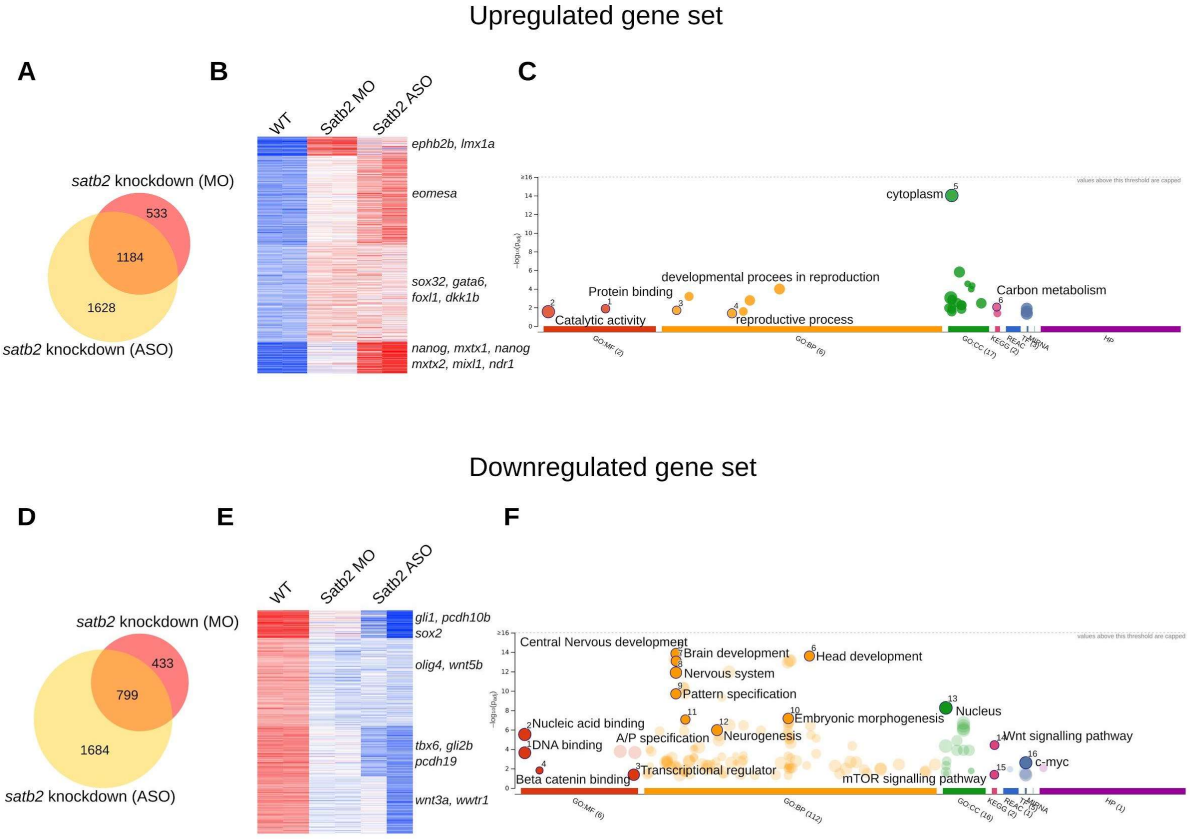


(A) Differential gene expression profiles using mRNA-seq upon *satb2* knockdown using morpholinos and **(B)** RNA antisense oligonucleotides. Bar plots representing normalized CPM values obtained from RNA-seq analysis for known ectoderm markers **(C)** Mesoderm

markers **(D)** and Endoderm markers **(E)**. Y-axis represented average normalized CPM values from three independent experiments (N=2). Error bar represents standard deviation. (p-value < 0.05 as determined by Robinson and Smyth statistical tests of EdgeR).

To identify the subset of deregulated genes that exhibit a similar expression pattern upon knockdown using either MO mix and ASO mix, we performed Venn analysis and extracted intersections of upregulated (Figure 2.5A) and downregulated genes (Figure 2.5D). K-means clustering again highlighted similar gene regulation of markers of three germ layers by Satb2 (Figure 2.5B and Figure 2.5E). Further, gene ontology analysis of a commonly downregulated subset of genes suggested enrichment for pathways such as pattern specification, embryonic morphogenesis, anterior-posterior axis specification etc. supporting the hypothesis that Satb2 is essential for cell fate specification and gastrulation.

Figure 2.5 Satb2 positively regulates pattern specification and induces neural fates during gastrulation.



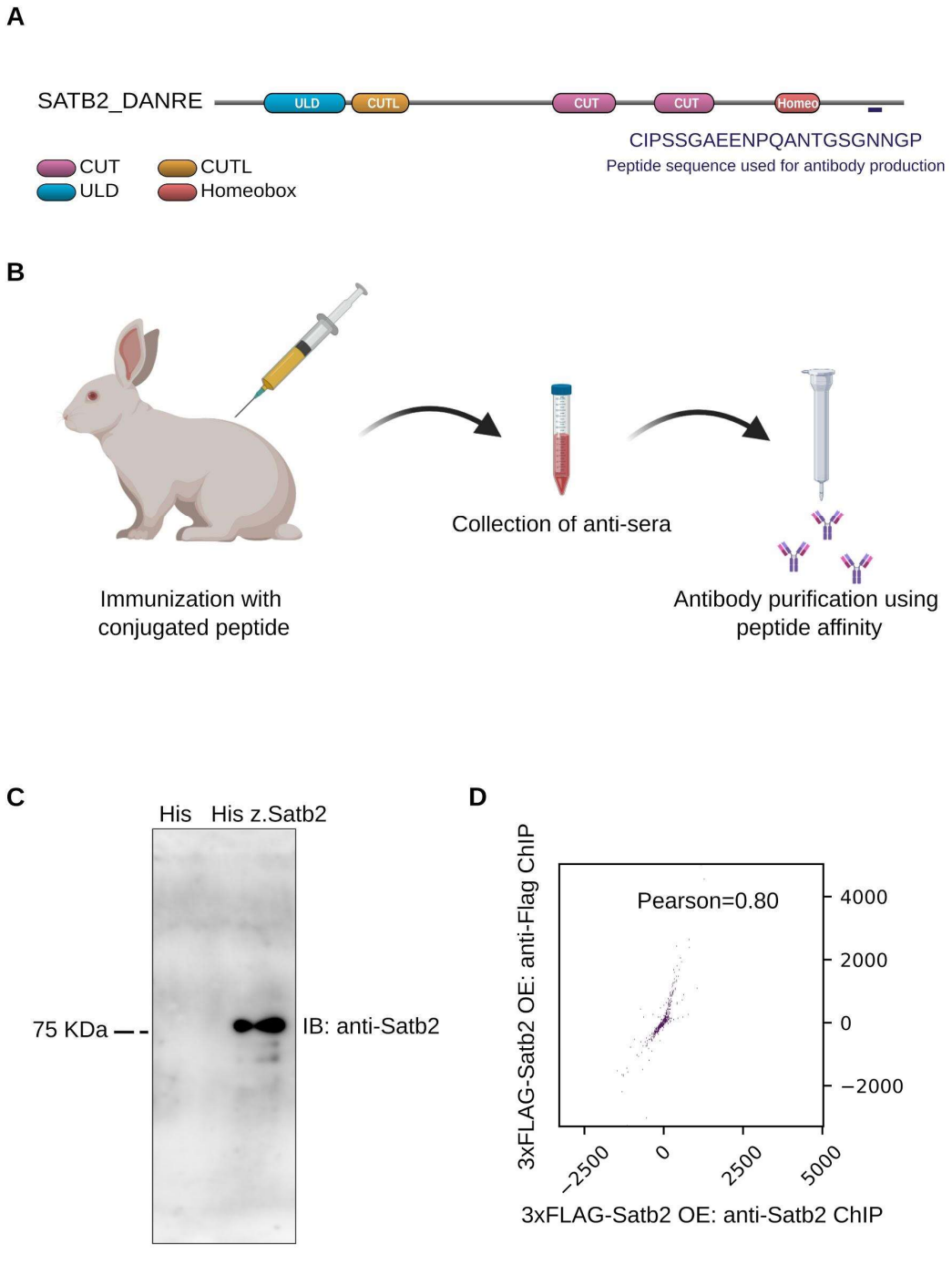
(A) Shared upregulated gene targets as compared to wild type control identified using Venn analysis of morpholino mediated and RNA antisense oligonucleotide mediated knockdown. **(B)** Heatmap representing normalized CPM values for common upregulated gene sets (Red: high, Blue: Low). **(C)** The associated dot plot represents gene ontology analysis for upregulated gene candidates using gProfiler. **(D)** Shared downregulated gene targets as compared to wild type control identified using Venn analysis of morpholino mediated and RNA antisense oligonucleotide mediated knockdown. **(E)** Heatmap representing normalized CPM values for common upregulated gene sets (Red: high, Blue: Low). **(F)** The associated dot plot represents gene ontology analysis using gProfiler for downregulated gene candidates.

2.3.5 Genome-wide occupancy analysis revealed direct transcriptional targets of Satb2.

To comprehend transcriptional regulatory mechanisms by Satb2, we performed chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) at 6 hpf. Since no ChIP-seq grade antibody was available for zebrafish Satb2, we generated a polyclonal antibody against a peptide spanning the N-terminal region of Satb2 (Figure 2.6A). Antiserum was further purified using peptide affinity columns to enrich antibody titre (Figure 2.6B). Quality control of antibody was performed using immunoblotting for recombinant purified zebrafish His-Satb2 protein (Figure 2.6C). Next, to test whether generated antibody can be successfully used for the ChIP-seq experiment, we overexpressed 3xFLAG-Satb2 from the one-cell stage and performed ChIP-seq at the dome stage using either anti-Flag antibody or in house generated anti-Satb2 antibody. We observed a very high correlation (Pearson correlation 0.80) between ChIP signal intensities across these two conditions (Figure 2.6D).

Next, we performed a ChIP-seq experiment in duplicates and identified genomic regions with Satb2 binding at 6 hpf. To be able to distinguish functional Satb2 binding sites, we intersected consensus peaks with an ensemble of accessible regions of chromatin across various developmental stages (GSE106428, GSE130944, GSE101779). Through this analysis, we could identify 43855 potential functional binding sites for Satb2 across the genome. Genomic distribution analysis for these peaks suggested Satb2 occupies both promoter and non-promoter regions (intergenic + intronic + exonic + TTS). However, about 90% of these sites were present in non-promoter genomic regions (Figure 2.7A).

Figure 2.6 Generation and characterization of anti-zebrafish Satb2 antibody.

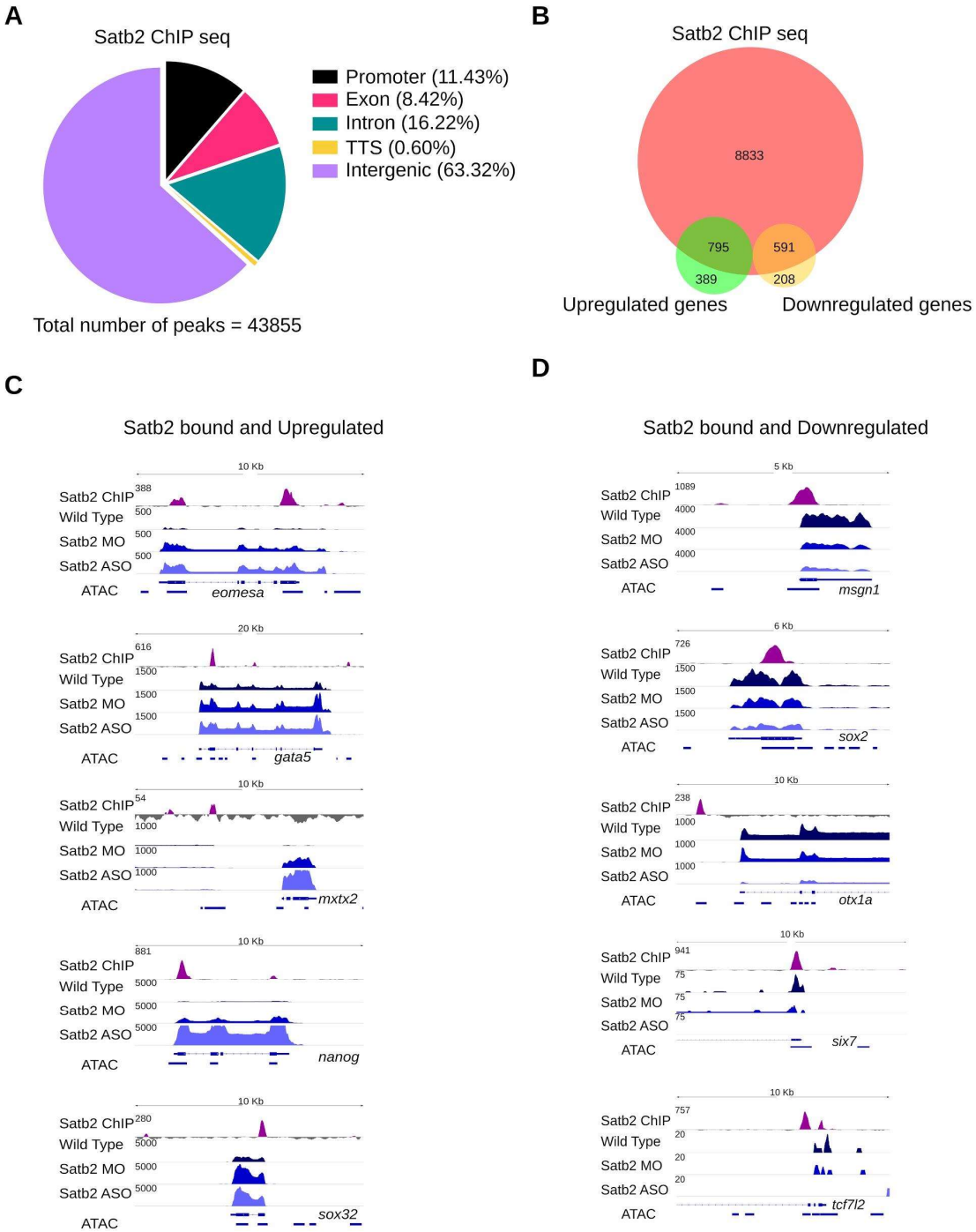


(A) Schematic of zebrafish Satb2 domain architecture highlighting C-terminal region used for peptide synthesis (blue block). Peptide sequence is denoted in N-C terminal

orientation. **(B)** The workflow of polyclonal antibody production followed by affinity purification. **(C)** Immunoblot analysis of recombinant control 6XHis-tag and His-tagged zebrafish Satb2 protein using lab generated zebrafish specific anti-Satb2 antibody. **(D)** Pearson correlation analysis of ChIP-sequencing datasets obtained using anti-FLAG antibody and anti-SATB2 antibody showing a high level of correlation.

Next, we annotated Satb2 ChIP peaks to the nearest genes. Intersection with differentially expressed gene set upon loss of function of Satb2 using Venn analysis revealed, Satb2 potentially occupies regulatory regions of both upregulated (795) and downregulated (591) genes (Figure 2.7B). As discussed earlier, we observed Satb2 negatively regulates endoderm marker genes such as *eomesa*, *gata5*, *mxtx2* and *sox32* by directly binding to core promoter or TSS (Figure 2.7C). Similarly, direct occupancy was observed on genetic loci of ectoderm or paraxial mesoderm marker genes such as *msgn1*, *otx2*, *sox2*, *six7* and *tcf7l2* which are positively regulated by Satb2 (Figure 2.7D). Collectively, our results suggest that Satb2 can potentially function as both activator and repressor of transcription based on activities of its target genes during vertebrate gastrulation.

Figure 2.7 Genome-wide occupancy analysis identifies Satb2 bound direct transcriptional targets.



(A) Genomic distribution of 43855 Satb2 peaks at 6 hpf. **(B)** Venn diagram 795 upregulated and 591 down-regulated genes are directly bound by Satb2 at 6 hpf. **(C)**

Integrative genomic viewer snapshots for Satb2 ChIP seq and corresponding RNA-seq for wild type, Satb2 morpholino and Satb2 RNA antisense oligonucleotide injected embryos for upregulated (left panel) and **(D)** downregulated genes

---

2.3.6 Characterizing genome-wide transcriptional activities of Satb2

Correlation analysis between Satb2 ChIP-seq and differential gene expression analysis highlighted the dual function of Satb2 both as an activator and as a repressor during fate commitment. To understand molecular mechanisms by which Satb2 differentiate its target activity, we first correlated Satb2 occupancy with histone modification(s) occupancy across the genome (Figure 2.8A). K-means clustering around +/- 2 Kb of Satb2 binding sites allowed us to identify 3 distinct clusters. Genomic sites belonging to each cluster were further subjected to motif analysis. This allowed us to predict potential interactors of Satb2 resulting in differential transcriptional activity (Figure 2.8B). Further, genomic distribution analysis helped us to understand the regulatory nature of transcriptional activity associated with each cluster either through direct regulation at the transcriptional start site (TSS) or through long-range mediated interactions (Figure 2.8C).

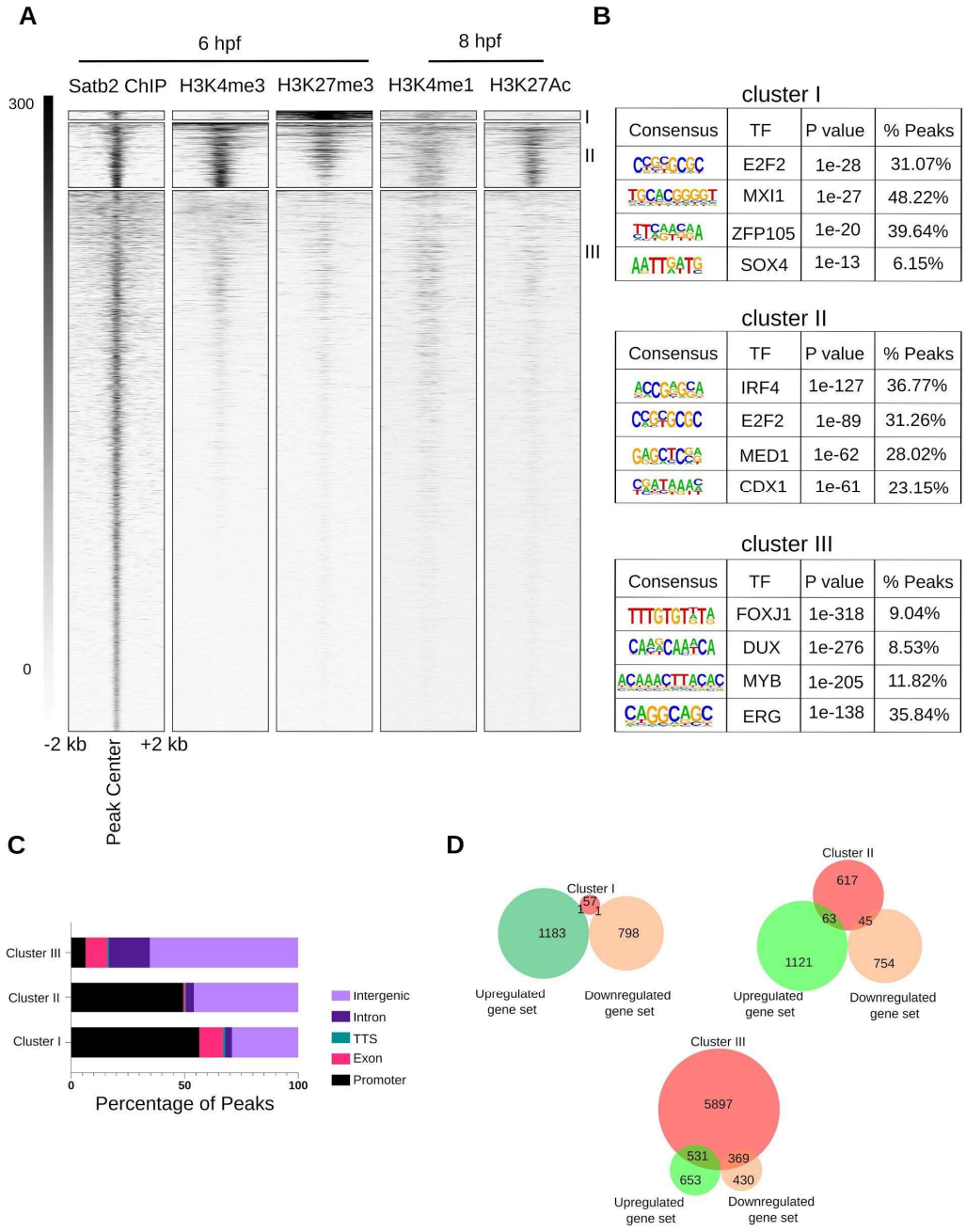
Cluster 1 denotes a very small subset of peaks that are enriched with high occupancy of H3K27me3 (mark of repressive chromatin). We observed the enrichment of transcriptional factors including E2F2, Mxi1 and Sox2. Interestingly, there was no significant overlap between genes belonging to cluster 1 and DE genes (Figure 2.8D), suggesting genomic occupancy associated with this particular cluster could be non-functional targets of Satb2 at least during early gastrulation.

Cluster 2 signifies genomic sites which show higher enrichment for activating (H3K4me3) than repressive marks (H3K27me3). These sites were also enriched for both enhancer associated marks H3K4me1 and H3K27Ac. Notably, we observed many of these sites to be associated with both core promoter and distal intergenic regions in an equal

proportion. A significant number of DE genes overlapped with this cluster. Interestingly, as in the case of cluster 1, E2F2 is enriched at these genomic sites. However, these sites are functionally regulated by Satb2 in contrast to sites associated with cluster 1.

Finally, analysis of cluster3 highlighted bivalent genes marked by both activation and repressive marks to a similar extent (Figure 2.8A). This bivalent nature of these genomic sites (mainly proximal regulatory sites) suggested that they are poised for transcriptional regulation during gastrulation. Motif enrichment analysis identified co-occupancy with homeobox proteins such as Foxj1, Dux and other transcriptional activators like Erg and Myb (Figure 2.8B). Interestingly, we found a maximum overlap of DE genes with cluster 4 suggesting Satb2 in association with other transcriptional regulators potentially target bivalent, developmentally poised genes to participate in the regulation of germ layer specification.

Figure 2.8 Characterizing Genome-wide occupancy of Satb2 bound direct transcriptional targets.



(A) Heatmap showing signal density distribution for Satb2 ChIP signal, H3K4me3, H3K27me3 (6 hpf), H3K4me1 and H3K27Ac (8 hpf) around +/- 2 Kb of Satb2 peaks. **(B)** Consensus binding sequences enriched respectively in cluster 1, cluster 2 and cluster 3. **(C)** Genomic distribution of cluster-specific peaks highlighting identification of promoter

and distal element mediated gene regulation. **(D)** Venn analysis of genes associated with each cluster and deregulated genes upon loss of Satb2 at 6 hpf.

2.3.7 Satb2 shares genomic sites with distinct mesendoderm fate determinants to drive target gene expression programs

Various developmental molecular signals including Nodal signalling activate target mesendoderm genes. Briefly, upon activation, phosphorylated Smad2 shuttles into the nucleus and interacts with other transcriptional regulators such as Eomesa, Ta, Tbx16, Mixl1, Mxtx2 and Nanog (Figure 2.9A) (Schier, 2009). Multiple studies have highlighted the role of these factors during cell fate specification (Griffin et al., 1998; Ho and Kane, 1990; Picozzi et al., 2009; Ryan et al., 1996; Talbot et al., 2020). Thus, to gain further insights into transcriptional regulatory activities by Satb2 particularly during mesendoderm specification, we correlated Satb2 occupancy with the binding profiles of these well-characterized regulators of mesendoderm specification. Although these datasets are from different developmental stages, in part allowed us extrapolation to understand gene regulatory networks during mesendoderm specification.

Clustering signal intensities from genome-wide occupancy analysis experiments for these factors around +/- 2 Kb of Satb2 ChIP peaks, highlighted two distinct regulatory modules. The first module showed shared genomic occupancy with endoderm driver factors, Eomesa and Mixl1. Moreover, we also observed shared occupancy with previously well-characterized mesodermal T-box protein Tbx16. Recently, Tbx16 is also implicated in regulating endoderm fates during zebrafish development (Nelson et al., 2017).

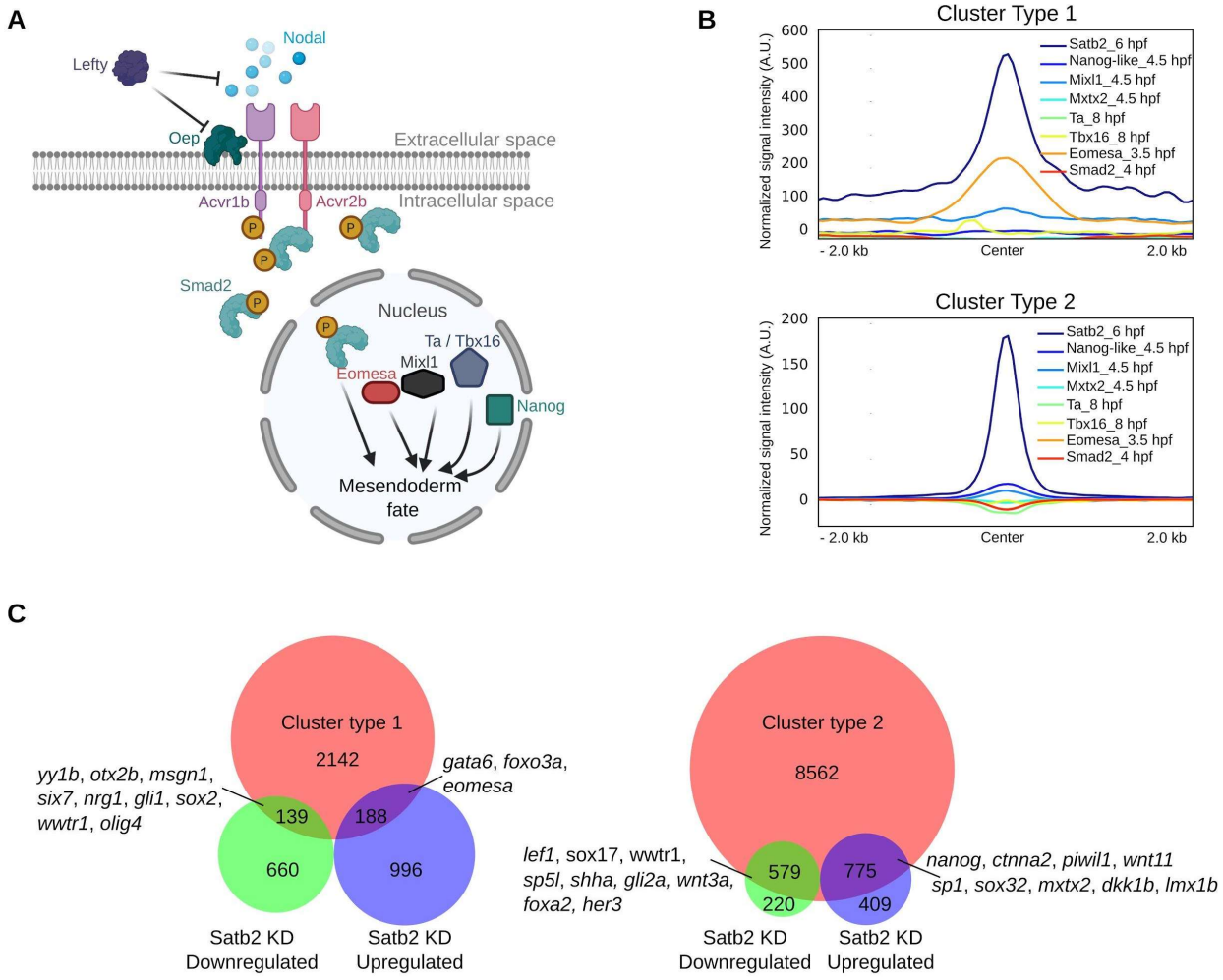
The second cluster exhibits shared occupancy only with Nanog to a certain extent (Figure 2.9B). For the rest of the factors, we did not find any significant correlation with the occupancy of Satb2. Instead, a negative correlation for Satb2 was observed with Eomesa, Ta/Tbx16 and Smad2. These results indicated that Satb2 participates in two distinct

regulatory modules depending on shared occupancy with mesendoderm drivers leading to differential target activities.

Next, we correlated differential gene expression dataset upon loss of Satb2 with each cluster. Genes belonging to cluster 1 were enriched in the processes such as early neurogenesis, neural crest specification and embryonic morphogenesis. However, genes belonging to cluster 2 were mostly associated with the development of vital organs of endoderm origin as well as various biosynthetic pathways. Expectedly, Satb2 positively regulates genes involved in neurogenesis and negatively regulates endoderm fate specification.

Collectively, these observations prompted us to hypothesize that during blastula stages endoderm determinant suppresses premature expression of neuroectodermal lineages and perhaps during gastrulation Satb2 negatively regulates these endoderm markers and occupy the same cis-regulatory modules to positively facilitate expression of neuroectodermal lineages.

Figure 2.9 Satb2 shares genomic sites with distinct mesendoderm fate determinants to drive target gene expression programs.



(A) Schematic representation of signalling inputs required for mesendoderm specification during gastrulation. Upstream nodal signalling is highlighted. Another signalling such as Bmp, Wnt, Fgf required for specification are not included for representation purpose. **(B)** Signal intensity profiles from ChIPseq experiments centred around +/- 2 Kb Satb2 bound regions at 6 hpf. Corresponding stages for each dataset have been mentioned in labels. Two types of clusters were identified and represented separately. **(C)** Venn analysis of cluster type 1 and cluster type 2 with downregulated and upregulated gene sets upon loss of Satb2. Some key genes belonging to each intersection are highlighted.

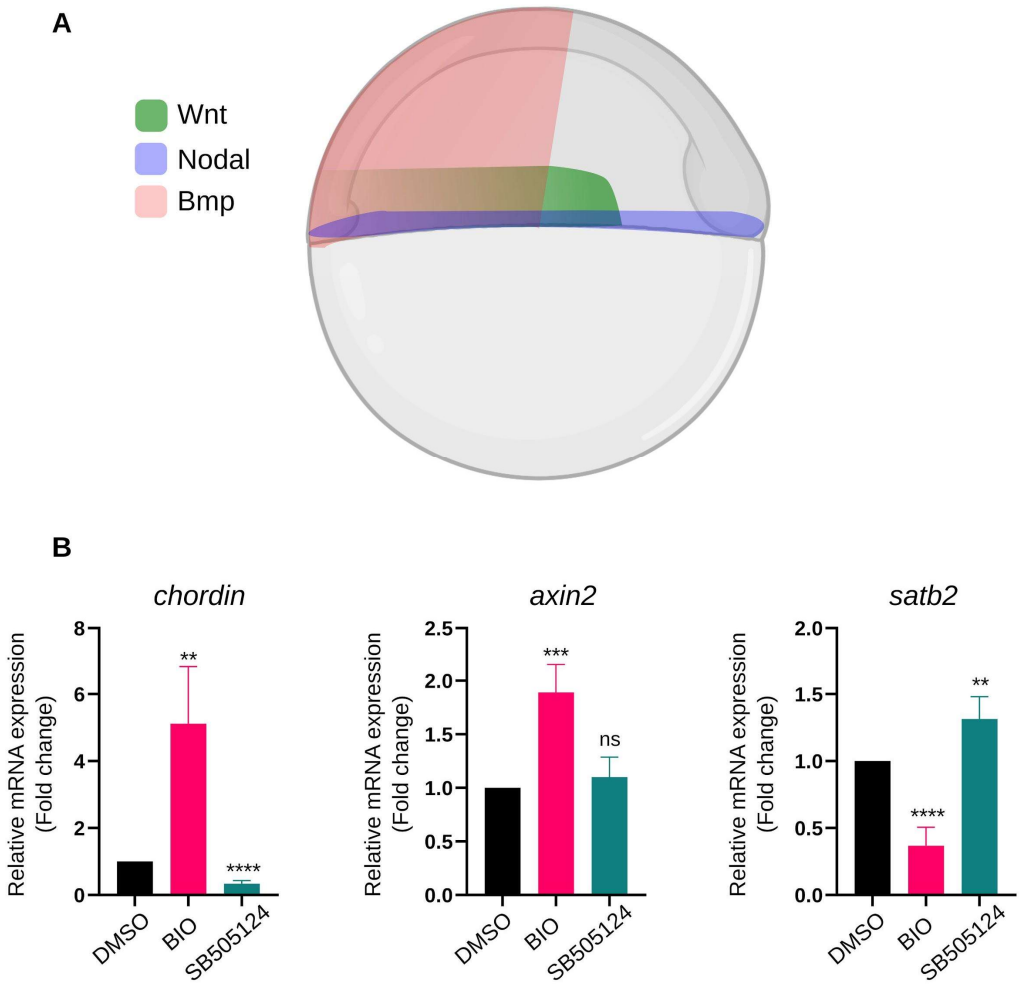
2.3.8 Interplay between Wnt signalling and Satb2 is essential for mesendoderm specification

After establishing a molecular mechanism by which Satb2 differentially regulates various lineages during zebrafish gastrulation, we wondered how the expression and activity of Satb2 are regulated. Early molecular mapping in combination with genetic studies allowed researchers to map signalling activities of key developmental signalling pathways such as Wnt, Bmp, nodal etc (Figure 2.10A) (Schier and Talbot, 2005). Bmp and Shh signalling have been shown to regulate the expression of Satb2 in pharyngeal arches during the late embryonic stages of zebrafish (Sheehan-Rooney et al., 2013).

Here, we tested if Wnt and Nodal signalling potentially regulate the expression of Satb2 during gastrulation. To achieve this, we treated zebrafish embryos with either Wnt activator BIO or Nodal inhibitor SB505124. Embryos were harvested at 6 hpf and mRNA abundance was estimated by qPCR. Treatment with Wnt activator resulted in significant upregulation of both *chordin* and *axin2*. However, treatment with Nodal inhibitor, SB505124 led to reduced expression of *chordin* but not *axin2*, as the latter is not the target of Nodal signalling. Interestingly, *satb2* levels were significantly downregulated when embryos were treated with BIO, suggesting a negative correlation between active Wnt signalling and expression of *Satb2.* We did not observe any massive change in expression of *satb2* upon treatment with Nodal inhibitors.

Collectively, these data indicated that regulation by Wnt signalling is necessary to determine the activity domain of Satb2 leading to the specification of mesendoderm.

Figure 2.10 Interplay between Wnt signalling and Satb2 is essential for mesendoderm specification.



(A) Schematic of activity of signalling pathways during axis patterning in zebrafish. Dorsal to the right, animal pole to the top. Gradients of Nodal (purple), Wnt (green) and Bmp (red) signals are depicted as in 6 hpf. **(B)** Relative mRNA abundance of *chordin*, *axin2* and *satb2* upon hyperactivation of wnt signalling using BIO and inhibiting nodal signalling using SB505124 at 6 hpf.

2.4 Discussion

Chromatin modifiers actively participate in cell fate specification and lineage commitment during embryogenesis by modulating genome architecture to regulate tissue-specific GRNs (Signolet and Hendrich, 2015). However, how the global changes in genome architecture are mechanistically connected to individual loci has remained elusive. Equally unknown are the individual players that participate as crucial molecular links between these processes. In this study, we have investigated the function of chromatin organizer protein SATB2 in establishing such higher-order GRNs, essential for gene regulatory transitions.

Lineage-specific transcriptome analysis generated in our study (Chapter 1) corroborated very well with the data from tissue-specific transcriptome analysis in Xenopus (Popov et al., 2017). We observed that Satb2 is enriched in dorsal organizer tissues indicating potential conserved functions of Satb2 in cell lineage commitment during gastrulation at least in early vertebrates. It would be interesting to characterize such spatial activity of Satb2 during mammalian embryogenesis.

Extensive Gene Tree and phylogenetic analysis provided the first evidence of the presence of complete SATB homolog proteins in invertebrates. Our analysis indicated that SATB1 and SATB2 diverged from the ancestral orthologs along with the evolution of jawed fish and acquired different functions in the mammalian systems. Systematic functional studies of ancestral orthologs will further contribute to our understanding of the evolution of mammalian SATB proteins and in turn the evolution of higher-order GRNs. The domain architecture analysis further highlighted the conserved nature of Satb2 across vertebrates and argues for possible functional conservation of this ancient member of the family. Collectively, these observations provided us with a strong motivation to study the function of Satb2 during the early stages of gastrulation.

Indeed, loss of function using two different methods (MOs and ASOs) resulted in severe developmental arrest during early gastrulation. Gene expression analysis highlighted the

differential regulation potential of Satb2 by acting as an activator as well as a repressor of key developmental genes. Interestingly, Satb2 facilitates the expression of genes involved in neural progenitors and simultaneously suppresses the expression of endoderm genes. Gene ontology analysis further strengthens these observations by suggesting positive regulation of processes like pattern specification, neurogenesis amongst others. Correlation between gene expression analysis and genome-wide occupancy analysis confirmed direct regulation of these molecular targets by Satb2.

Histone modification levels serve as an indication of local chromatin state at a given developmental time (Bannister and Kouzarides, 2011). Integrating ChIP-seq data for H3K4me3 (active mark), H3K27me3 (repressive mark), H3K4me1 (active enhancer) and H3K27Ac (active mark) allowed us to identify distinct clusters and gene modules that are differentially regulated by Satb2 during development. We observed that Satb2 potentially targets genes that exhibit a bivalent chromatin state. These genes were mainly associated with key developmental processes such as pattern determination. The bivalent chromatin state of these developmental genes suggests that their expression can be switched on or switched off in a context-dependent manner. Indeed, our study indicated that Satb2 regulates this cluster of genes differently based on the potential interaction with different transcription factors as well as by changing the location of Satb2 binding across the genome.

Intrigued by the differential regulation by Satb2, we decided to integrate Satb2 genomic occupancy with cis-regulatory modules associated with well-characterized mesendoderm drivers such as Smad2, Eomesa, Ta, Tbx16, Mxtx2, Mixl1 and Nanog. Smad2 is the primary effector of Nodal signalling which associates with other TFs including Eomesa to facilitate gene expression of mesendoderm specification genes (Nelson et al., 2014; Schier, 2009). Eomesa is a known determinant of definitive endoderm during mouse embryonic development and its function is conserved in zebrafish (Arnold et al., 2008; Talbot et al., 2020; Tosic et al., 2019). However, many studies have also indicated the role of Eomesa in the specification of mesoderm (Ryan et al., 1996, Picozzi et al., 2009). Studies by Nelson et al. characterized new functions for Ta and Tbx16 in the development

of endoderm derived organs such as liver and gut in addition to their known role in mesoderm specification in zebrafish and Xenopus (Griffin et al., 1998; Ho and Kane, 1990; Nelson et al., 2017). In a study using zebrafish, Hong et al. suggested the requirement of Nodal-related ligands and Mxtx2 for induction of mesendoderm lineages (Hong et al., 2011). Moreover, the interplay between Nanog-Mxtx2 and Eomesa-Mxtx2 have been implicated extensively in endoderm formation (Xu et al., 2012, 2014). Notably, Mixl1 is also characterized as an endodermal regulator (Kikuchi et al., 2000). Collectively, these studies suggest an involvement of higher-order gene regulatory networks which operate in an interlinked but distinct manner during mesendoderm specification. Integrative analysis of ChIP-seq and transcriptome analysis revealed that Satb2 regulates gene expression based on its association with CRMs regulated by these driver genes. We observed two distinct clusters, first, Satb2 shares genomic sites with Eomesa and Tbx16 and second, Satb2 is associated with Nanog but exhibits anticorrelation with the occupancy of other transcription factors. Collectively, this data allowed us to predict the negative interplay between Satb2 and Eomesa leading to differential regulation of genes involved in neurogenesis and endoderm specification.

Mesendoderm specification is induced by gradient activities of various developmental signalling pathways such as Wnt and Nodal. Eomesa, Ta (T in mouse), Mixl1 are molecular targets of Wnt signalling (Choi et al., 2015; Herrmann et al., 1990; Nelson et al., 2014). Interestingly Nanog suppresses the activity of nuclear beta-catenin to restrict expression of latter to dorsal tissue (Cheng et al., 2015; Davidson et al., 2012; He et al., 2020). Both Satb1 and Satb2 have been shown to regulate various physiological processes through modulation of Wnt signalling (Mir et al., 2016; Notani et al., 2010; Yu et al., 2017). Collectively, this prompted us to study the interplay between Wnt signalling and Satb2 during gastrulation. Strikingly, the activation of Wnt signalling using a small molecular weight inhibitor of GSK3β resulted in substantial downregulation of *satb2* expression. These results again highlighted the activity of transcriptional repressor activity of Satb2 is restricted to dorsal organizer tissue by active Wnt signalling. Further molecular characterization of Wnt/Satb2 functional interaction would be instrumental to our understanding of spatial activation mesendoderm specification genes.

2.5 References

Ahn, H.-J., Park, Y., Kim, S., Park, H.-C., Seo, S.-K., Yeo, S.-Y., and Geum, D. (2010). The expression profile and function of Satb2 in zebrafish embryonic development. Mol. Cells *30*, 377–382.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646–3658.

Arnold, S.J., Huang, G.-J., Cheung, A.F.P., Era, T., Nishikawa, S.-I., Bikoff, E.K., Molnár, Z., Robertson, E.J., and Groszer, M. (2008). The T-box transcription factor Eomes/Tbr2 regulates neurogenesis in the cortical subventricular zone. Genes Dev. *22*, 2479–2484.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612–616.

Britanova, O., Akopov, S., Lukyanov, S., Gruss, P., and Tarabykin, V. (2005). Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. Eur. J. Neurosci. *21*, 658–668.

Brocato, J., and Costa, M. (2015). SATB1 and 2 in colorectal cancer. Carcinogenesis *36*, 186–191.

Bürglin, T.R., and Affolter, M. (2016). Homeodomain proteins: an update. Chromosoma *125*, 497–521.

Bürglin, T.R., and Cassata, G. (2002). Loss and gain of domains during evolution of cut superclass homeobox genes. Int. J. Dev. Biol. *46*, 115–123.

Cai, S., Han, H.-J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nat. Genet. *34*, 42–51.

Cera, I., Whitton, L., Donohoe, G., Morris, D.W., Dechant, G., and Apostolova, G. (2019). Genes encoding SATB2-interacting proteins in adult cerebral cortex contribute to human cognitive ability. PLoS Genet. *15*, e1007890.

Cheng, P., Sun, X., Yin, D., Xu, F., Yang, K., Qin, L., Dong, Y., Guo, F., Chen, A., Zhang, W. and Huang, H.(2015). Nanog down-regulates the Wnt signaling pathway via β-catenin phosphorylation during epidermal stem cell proliferation and differentiation.Cell Biosci. *5*,1-10.

Choi, S.C., Choi, J.H., Cui, L.H., Seo, H.R., Kim, J.H., Park, C.Y., Joo, H.J., Park, J.H., Hong,

S.J., Yu, C.W. and Lim, D.S.(2015). Mixl1 and Flk1 are key players of Wnt/TGF-β signaling during DMSO-induced mesodermal specification in P19 cells.J. Cell Physiol. *230*,.1807-1821.

Davidson, E.H. (2001). Genomic Regulatory Systems: In Development and Evolution (Elsevier).

Davidson, E.H. (2010). The Regulatory Genome: Gene Regulatory Networks In Development And Evolution (Elsevier).

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C. and Otim, O.(2002). A genomic regulatory network for development.Science, *295*,1669-1678.

Davidson, K.C., Adams, A.M., Goodson, J.M., McDonald, C.E., Potter, J.C., Berndt, J.D., Biechele, T.L., Taylor, R.J., and Moon, R.T. (2012). Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. Proc. Natl. Acad. Sci. U. S. A. *109*, 4485–4490.

Dickinson, L.A., Dickinson, C.D., and Kohwi-Shigematsu, T. (1997). An Atypical Homeodomain in SATB1 Promotes Specific Recognition of the Key Structural Element in a Matrix Attachment Region. **J**. Biol. **Chem**. *272*, 11463–11470.

Dobreva, G., Dambacher, J., and Grosschedl, R. (2003). SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin μ gene expression. Genes Dev. *17*, 3048–3061.

Dobreva, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Fariñas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell *125*, 971–986.

FitzPatrick, D.R., Carr, I.M., McLaren, L., Leek, J.P., Wightman, P., Williamson, K., Gautier, P., McGill, N., Hayward, C., Firth, H. and Markham, A.F.(2003). Identification of SATB2 as the cleft palate gene on 2q32–q33. Hum. Mol. Genet. *12*,2491-2501.

Fuß, B., and Hoch, M. (1998). Drosophila endoderm development requires a novel homeobox gene which is a target of Wingless and Dpp signalling. Mech. Dev. *79*, 83–97.

Galande, S., Dickinson, L.A., Mian, I.S., Sikorska, M., and Kohwi-Shigematsu, T. (2001). SATB1 cleavage by caspase 6 disrupts PDZ domain-mediated dimerization, causing detachment from chromatin early in T-cell apoptosis. Mol. Cell. Biol. *21*, 5591–5604.

Gómez-Díaz, E., and Corces, V.G. (2014). Architectural proteins: regulators of 3D genome organization in cell fate. Trends Cell Biol. *24*, 703–711.

Griffin, K.J., Amacher, S.L., Kimmel, C.B., and Kimelman, D. (1998). Molecular identification of spadetail: regulation of zebrafish trunk and tail mesoderm formation by T-box genes. Development *125*, 3379–3388.

Gyorgy, A.B., Szemes, M., de Juan Romero, C., Tarabykin, V., and Agoston, D.V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. Eur. J.

Neurosci. *27*, 865–873.

Han, H.-J., Russo, J., Kohwi, Y., and Kohwi-Shigematsu, T. (2008). SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. Nature *452*, 187–193.

He, M., Zhang, R., Jiao, S., Zhang, F., Ye, D., Wang, H., and Sun, Y. (2020). Nanog safeguards early embryogenesis against global activation of maternal β-catenin activity by interfering with TCF factors. PLoS Biol. *18*, e3000561.

Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. Nature *343*, 617–622.

Ho, R.K., and Kane, D.A. (1990). Cell-autonomous action of zebrafish spt-1 mutation in specific mesodermal precursors. Nature *348*, 728–730.

Hong, S.-K., Jang, M.K., Brown, J.L., McBride, A.A., and Feldman, B. (2011). Embryonic mesoderm and endoderm induction requires the actions of non-embryonic Nodal-related ligands and Mxtx2. Development *138*, 787–795.

Kaaij, L.J.T., Mokry, M., Zhou, M., Musheev, M., Geeven, G., Melquiond, A.S.J., de Jesus Domingues, A.M., de Laat, W., Niehrs, C., Smith, A.D., et al. (2016). Enhancers reside in a unique epigenetic environment during early zebrafish development. Genome Biol. *17*, 146.

Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. Cell Rep. *24*, 1–10.e4.

Kikuchi, Y., Trinh, L.A., Reiter, J.F., Alexander, J., Yelon, D., and Stainier, D.Y. (2000). The zebrafish bonnie and clyde gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors. Genes Dev. *14*, 1279–1289.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.

Kohwi-Shigematsu, T., Maass, K., and Bode, J. (1997). A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. Biochemistry *36*, 12005–12010.

Leichsenring, M., Maes, J., Mossner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 Transcription Factor Controls Zygotic Gene Activation In Vertebrates. Science *341*, 1005–1009.

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., and Gascuel, O. (07 02, 2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res. *47*, W260–W265.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. *44*, W242–W245.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. and Sandstrom, R.(2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome.Science *32*6,289-293.

Meier, M., Grant, J., Dowdle, A., Thomas, A., Gerton, J., Collas, P., O'Sullivan, J.M., and Horsfield, J.A. (2018). Cohesin facilitates zygotic genome activation in zebrafish. Development *145,1*.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2012). The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources. In Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the Campus and beyond, (Chicago, Illinois, USA: Association for Computing Machinery), 1–8.

Mir, R., Pradhan, S.J., Patil, P., Mulherkar, R., and Galande, S. (2016). Wnt/β-catenin signaling regulated SATB1 promotes colorectal cancer tumorigenesis and progression. Oncogene *35*, 1679–1691.

Naik, R., and Galande, S. (2019). SATB family chromatin organizers as master regulators of tumor progression. Oncogene *38*, 1989–2004.

Nakagoshi, H., Hoshi, M., Nabeshima, Y.-I., and Matsuzaki, F. (1998). A novel homeobox gene mediates the Dpp signal to establish functional specificity within target cells. Genes Dev. *12*, 2724–2734.

Nelson, A.C., Cutty, S.J., Niini, M., Stemple, D.L., Flicek, P., Houart, C., Bruce, A.E.E., and Wardle, F.C. (2014). Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. BMC Biol. *12*, 81.

Nelson, A.C., Cutty, S.J., Gasiunas, S.N., Deplae, I., Stemple, D.L., and Wardle, F.C. (2017). In Vivo Regulation of the Zebrafish Endoderm Progenitor Niche by T-Box Transcription Factors. Cell Rep. *19*, 2782–2795.

Notani, D., Gottimukkala, K.P., Jayani, R.S., Limaye, A.S., Damle, M.V., Mehta, S., Purbey, P.K., Joseph, J., and Galande, S. (2010). Global regulator SATB1 recruits beta-catenin and regulates T(H)2 differentiation in Wnt-dependent manner. PLoS Biol. *8*, e1000296.

Pauli, A., Montague, T.G., Lennox, K.A., Behlke, M.A., and Schier, A.F. (2015). Antisense Oligonucleotide-Mediated Transcript Knockdown in Zebrafish. PLoS One *10*, e0139504.

Pavan Kumar, P., Purbey, P.K., Sinha, C.K., Notani, D., Limaye, A., Jayani, R.S., and Galande, S. (2006). Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. Mol. Cell *22*, 231–243.

Picozzi, P., Wang, F., Cronk, K., and Ryan, K. (2009). Eomesodermin requires transforming growth factor-beta/activin signaling and binds Smad2 to activate mesodermal genes. J. Biol. Chem. *284*, 2397–2408.

Popov, I.K., Kwon, T., Crossman, D.K., Crowley, M.R., Wallingford, J.B., and Chang, C. (2017). Identification of new regulators of embryonic patterning and morphogenesis in Xenopus gastrulae by RNA sequencing. Dev. Biol. *426*, 429–441.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Ryan, K., Garrett, N., Mitchell, A., and Gurdon, J.B. (1996). Eomesodermin, a key early gene in Xenopus mesoderm differentiation. Cell *87*, 989–1000.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev. *23*, 2625–2638.

Schier, A.F. (2009). Nodal morphogens. Cold Spring Harb. Perspect. Biol. *1*, a003459.

Schier, A.F., and Talbot, W.S. (2005). Molecular genetics of axis formation in zebrafish. Annu. Rev. Genet. *39*, 561–613.

Sheehan-Rooney, K., Pálinkášová, B., Eberhart, J.K., and Dixon, M.J. (2010). A cross-species analysis of Satb2 expression suggests deep conservation across vertebrate lineages. Dev. Dyn. *239*, 3481–3491.

Sheehan-Rooney, K., Swartz, M.E., Lovely, C.B., Dixon, M.J., and Eberhart, J.K. (2013). Bmp and Shh Signaling Mediate the Expression of satb2 in the Pharyngeal Arches. PLoS One *8*.

Signolet, J., and Hendrich, B. (2015). The function of chromatin modifiers in lineage commitment and cell fate specification. FEBS J. *282*, 1692–1702.

Simakov, O., Marlétaz, F., Yue, J.X., O'Connell, B., Jenkins, J., Brandt, A., Calef, R., Tung, C.H., Huang, T.K., Schmutz, J. and Satoh, N. (2020). Deeply conserved synteny resolves early events in vertebrate evolution. Nat. Ecol. Evol., *4*, 820-830

Takatori, N., and Saiga, H. (2008). Evolution of CUT class homeobox genes: insights from the genome of the amphioxus, Branchiostoma floridae. The International Journal of Developmental Biology *52*, 969–977.

Talbot, C.D., Walsh, M.D., Cutty, S.J., Elsayed, R., Bruce, A.E.E., Wardle, F.C., and Nelson, A.C. (2020). Eomesodermin is functionally conserved between zebrafish and mouse in spite of different mutant phenotypic severities, and controls left/right organiser formation via interlocking feedforward loops. *bioRxiv*.

Tan, J.-A.T., Song, J., Chen, Y., and Durrin, L.K. (2010). Phosphorylation-Dependent Interaction of SATB1 and PIAS1 Directs SUMO-Regulated Caspase Cleavage of SATB1. Molecular and Cellular Biology *30*, 2823–2836.

Tosic, J., Kim, G.J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L. and Arnold, S.J.(2019). Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. Nat. Cell Biol.

21, 1518-1531.

Valentine, J.W., Jablonski, D., and Erwin, D.H. (1999). Fossils, molecules and embryos: new perspectives on the Cambrian explosion. Development *126*, 851–859.

Wang, Z., Yang, X., Chu, X., Zhang, J., Zhou, H., Shen, Y., and Long, J. (2012). The structural basis for the oligomerization of the N-terminal domain of SATB1. Nucleic Acids Res. *40*, 4193–4202.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene.

Westerfield, M. (2007) The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 5th Edition. University of Oregon Press, Eugene.

Xu, C., Fan, Z.P., Müller, P., Fogley, R., DiBiase, A., Trompouki, E., Unternaehrer, J., Xiong, F., Torregroza, I., Evans, T. and Megason, S.G.(2012). Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway.Dev. Cell *22*, 625-638.

Xu, P., Zhu, G., Wang, Y., Sun, J., Liu, X., Chen, Y.-G., and Meng, A. (2014). Maternal Eomesodermin regulates zygotic nodal gene expression for mesendoderm induction in zebrafish embryos. J. Mol. Cell Biol. *6*, 272–285.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. Nature *419*, 641–645.

Yu, W., Ma, Y., Ochoa, A.C., Shankar, S., and Srivastava, R.K. (2017). Cellular transformation of human mammary epithelial cells by SATB2. Stem Cell Res. *19*, 139–147.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res. *27*, 479–490.

Chapter 3

Investigating the role of chromatin organizer Satb2 during early stages of neural fate commitment

3.1 Introduction

Another example of the Cambrian explosion which is also a characteristic of the evolution of vertebrates is the formation of craniofacial structures like skeletal tissues and jaws. The evolution of these structures is vital for prey-predator interactions (Gans and Northcutt, 1983). The SATB family exhibits co-divergence at the branch of evolution of jawed vertebrates (Chapter 2, Figure 2.2). Moreover, SATB2 exhibits a high level of conservation across vertebrates at the amino acid sequence level and is also an ancient member of the family based on phylogenetic analysis (Sheehan-Rooney et al., 2010) (Chapter 2, Figure 2.2A).

Studies using an early vertebrate model system, zebrafish (Danio *rerio*) suggest that Satb2 is maternally deposited and widely expressed during the early stages of embryogenesis (Ahn et al., 2010). Loss of function studies using morpholinos against Satb2 resulted in an early developmental arrest (Ahn et al., 2010) (Chapter 2, Figure 2.3). During the stages of organogenesis, BMP and SHH signalling have been shown to regulate the expression of *satb2* (Bonilla-Claudio et al., 2012; Sheehan-Rooney et al., 2013). Loss of BMP and SHH signalling results in a significant reduction of *satb2* expression in neural crest cells of zebrafish (Sheehan-Rooney et al., 2013). Neural crest cells (NCCs), a specialized type of multipotent cells, largely contribute to the development of cartilage, pigmentation cells, bones and connective tissue of craniofacial structures in developing embryos (Douarin et al., 1999; Le Douarin and Dupin, 2003). However, studies highlighting the molecular interplay between Satb2 and transcription factors regulating NC specification and differentiation have been limited, hence would be instrumental in decoding the higher-order GRNs underlying genetic diseases.

3.1.1 Craniofacial patterning and skeletal morphogenesis

Craniofacial development is a complex process consisting of the formation of head, jaws, skeletal morphogenesis and tooth development. Defects incorrect specification leads to craniofacial abnormalities and is amongst the most prevalent birth defects. Abnormalities in cleft lip and cleft palate are the most common congenital anomalies affecting approximately 1 in every 700 live births (World Health Organization [WHO], 2001). Thus understanding the aetiology and the molecular mechanisms that regulate craniofacial patterning is crucial towards both surgical correction and medical treatment.

Craniofacial development requires coordinated regulatory interactions between a number of distinct cellular sources including cranial ectoderm, cranial endoderm, mesoderm and neural crest cells (Trainor and Krumlauf, 2000). Many studies focusing on such coordinated interactions between neurogenic placode-derived organs and cartilage or bone have made significant advances in our understanding of developmental coordination in head structures and hold significant evolutionary importance suggesting co-evolution of neural components with cranial structures (reviewed in Adameyko and Fried, 2016). However, further studies focusing on delineating regulatory networks which are responsible for the establishment of such neural components would be instrumental in understanding underlying mechanisms of craniofacial development (Adameyko and Fried, 2016).

3.1.2 Neural crest cells and craniofacial development

Neural crest cells (NCCs) are a major source of neural components amongst others that participate in the formation of craniofacial structures (Kague et al., 2012). Neural crest cells are vertebrate-specific, transient cell populations that originate and delaminate from the dorsal neural tube (Simões-Costa and Bronner, 2015). NCCs are a multipotent group of cells that contribute to melanocytes, neurons and glia of the peripheral nervous system (specifically the dorsal root and parasympathetic ganglion), cartilage, bone, adipose tissue, tendons, dermal fibroblasts, smooth muscles, pericytes, pigmentation, sensory neurons and many more cell types which contributes to the developing head (Douarin et al., 1999; Dupin et al., 2006; Le Douarin et al., 1999). NCCs also possess a unique property of interacting with cellular components originated from all three germ layers (mesoderm, endoderm and ectoderm) required for morphogenetic events which are essential for craniofacial development (Couly et al., 2002; Marcucio et al., 2011; Rinon et al., 2007; Van Ho et al., 2011).

During early embryogenesis in vertebrates, specific gene regulatory inputs confer multipotency and migratory capabilities to neural crest cells (Bronner-Fraser, 2008; Sauka-Spengler et al., 2007; Yu, 2010). The high migratory potential of these cells explains how these cells can contribute to the development of multiple tissue structures and organs of different origins. NCCs are also responsible for the appearance of jaw structures, organization of cephalic organs in

sensory organs of vertebrates and are regarded as one of the unique features of vertebrate evolution (Gitton et al., 2010; Minoux and Rijli, 2010; Trainor, 2005). Neural crest cells which undergo delamination and migrate to the pharyngeal arches will initiate and organize craniofacial development responding to gene regulatory inputs which determines tissue patterning and cell differentiation such as FGF, Notch, Shh, Wnt, BMP, PDGF and retinoic acid (Cornell and Eisen, 2000; Garnett et al., 2012; Liem et al., 1995; Marchant et al., 1998; Mayor et al., 1997; Steventon and Mayor, 2009; Trainor, 2013; Villanueva et al., 2002).

Early studies in mouse and chick chimaeras highlighted the instructive competence of the neural crest in the formation of dental primordia in the developing jaws (Mitsiadis et al., 2003). The experiments performed here demonstrate that mouse neural crest cells can utilize the potential of chick ectoderm to form dental primordia again highlighting crosstalk between different germ layers and neural crest cells to shape developing head structures.

3.1.3 Role of chromatin modifiers in neural crest specification

Neural crest cells (NCCs), a specialized type of multipotent cells, largely contribute to the development of cartilage, pigmentation cells, bones and connective tissue of craniofacial structures in developing embryos (Gans and Northcutt, 1983; Le Douarin and Dupin, 2003; Le Douarin et al., 1999). Exit from pluripotency and entry into differentiation programmes is accompanied by dramatic changes in the chromatin landscape. These changes are gradually acquired by changing the histone modification status through the activity of chromatin modifiers. Previous sections have highlighted the role of various signalling pathways in neural crest specification. However, how active chromatin remodelling affects chromatin landscape and in turn, neural crest specification remains under investigation. Recent advances in molecular tools such as Chromatin Immunoprecipitation followed by high throughput sequencing (ChIP-seq), Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) has been instrumental in characterizing chromatin landscape of developing cell type. Recently, these tools are being used for identifying enhancers and characterizing enhancer mediated regulation of neural crest cells. Typically, the local chromatin landscape around enhancers and promoters are tightly regulated by several epigenetic modifiers during NC induction and migration events (Rada-Iglesias et al., 2012). Classical marker genes of neural crest such as Sox10, Sox9, Snai2, Foxd3 and Pax3 are shown to be regulated through enhancer mediated long-range interactions

in multiple vertebrate systems (Betancur et al., 2010a, 2010b; Dutton et al., 2001; Garnett et al., 2012; Hu et al., 2014; Li et al., 2002; Lister et al., 2006; Stewart et al., 2006; Thisse et al., 1995). Interestingly, recent studies in zebrafish also suggest that these Sox9 and Foxd3 genes are required to maintain local chromatin landscape and loss of function of these genes results in severe craniofacial abnormalities (Lukoseviciute et al., 2018; Rainger et al., 2014). Lukoseviciute *et al*. suggest that foxd3 regulates neural crest specification through its bimodal activity and differentially regulating chromatin landscape (Lukoseviciute et al., 2018). Tien *et al*. conducted similar studies in frogs and reported that the interaction between Polycomb Repressive Complex (PRC) interacts and SNAI1/2 proteins control H3K27 trimethylation on target promoters of NC specifying genes (Tien et al., 2015). Additionally, DNMT3B binds to and methylates Sox10 promoter whereas JMJD2A (KDM4A) facilitates removal of repressive marks (H3K9me3) at the promoter of Sox10 (Matsukawa et al., 2015). Such a switch in molecular networks is also observed for regulation of other NC specifiers such as Snai2, Foxd3 and Sox3 (Strobl-Mazzulla and Bronner, 2014; Strobl-Mazzulla and Bronner-Fraser, 2011). Thus, characterizing functions of additional chromatin modifiers during NC specification and migration would be pivotal in understanding craniofacial development.

3.1.4 Zebrafish as a model organism to understand Craniofacial patterning and skeletal morphogenesis

Craniofacial abnormalities include multiple malformations in the skull, jaw, face and teeth development (OMIM). These abnormalities are genetic with no preventive measures available currently. Thus, modelling these malformations in other vertebrate developmental systems is a necessity of the time to understand underlying molecular mechanisms.

Recently, *Danio rerio* (Zebrafish) have been extensively used to model the genetic and environmental basis of these defects and to study contributions from different cell type sources of neural components such as neural crest cells in craniofacial development (Raterman et al., 2020). The adult zebrafish skull is made up of 43 cartilage derived bones and 30 ossifying bones. This number is significantly higher than mammalian counterparts (Cubbage and Mabee, 1996). Development of these structures can be traced back to as early as 5 dpf during larval stages of zebrafish development and have been used to study skeletal anatomy extensively (Piotrowski et al., 1996; Schilling et al., 1996).

In zebrafish, migration of neural crest cells marks the early stages of craniofacial development. Induction of the neural crest begins during the late stages of gastrulation (Kimmel et al., 1995) and differs from other vertebrates and non-teleost fish in certain aspects (Rocha et al., 2020). For instance, in sarcopterygians, proliferation, invagination sets up a process called 'open neurulation'. This process is characterized by the opening of the dorsal neural tubes. However, In zebrafish, the convergence of the neural ectoderm cells replaces the process of invagination and produce the neural keel followed by the neural tube, through the process of delamination (Lowery and Sive, 2004; Rocha et al., 2020). Neural crest development is a multistep process as NC induction, NC specification and finally migration of neural crest to destination. Following induction of the neural crest, cells undergo epithelial to mesenchymal transition, often referred to as delamination in the context of neural crest development, before initiation of migration. The migratory routes of cranial and trunk neural crest cells differ significantly. Neural crest cell (NCC) populations are specified along the anterior-posterior axis of the embryo, hence, cranial neural crest migrates prior to that of trunk neural crest which contributes to the peripheral nervous system. The migration of both cranial and trunk NCCs is regulated by a combination of chemo-attractants and repulsive cues (Theveneau and Mayor, 2012). Typically, the development of pharyngeal arches in zebrafish is studied to understand the development of mammalian inner ear and jaw structures whereas neurocranium is studied to model the development of mammalian palette (Mork and Crump, 2015). Simple genetics, large sample size and ability to trace back development to very early stages using high-resolution imaging makes zebrafish a versatile model to understand craniofacial development and complement previous studies performed using other vertebrates and mammalian model systems such as chick and mouse.

3.1.5 Role of Satb2 in neural crest function and craniofacial development

Satb2 is located at a gene-poor locus on chromosome 2q32-q33 in humans and exhibits haploinsufficiency during cleft palate abnormalities (FitzPatrick et al., 2003). Cleft palate is the result of the failure of fusion of the secondary palate during human embryo development. Patients with cleft palate often harbour translocations and/or deletions at the 2q32-q33 region. SATB2 null mutations result in amplification of these defects and craniofacial development, such as defects in jaw formation (Britanova et al., 2005, 2006) as well as defects in cognition (Leoyklang et al., 2007). Another example of association between a loss of SATB2 function and

congenital defects is a condition known as the Pierre-Robin disorder often characterized by cleft palate, micrognathia, and glossoptosis, often accompanied by obstructive apnea (Rainger et al., 2014). Further, craniofacial abnormalities along with neurodevelopmental defects including behavioural issues and compromised speech ability are typical characteristics of the SATB2 associated syndrome in humans (Thomason et al., 2019; Zarate and Fish, 2017; Zarate et al., 2017). Mouse knockout models have been instrumental in characterizing the molecular functions of Satb2. Constitutive homozygous knockout mutants of Satb2 are embryonic lethal exhibiting severe defects in neural development, craniofacial patterning and skeletal morphogenesis (Alcamo et al., 2008; Britanova et al., 2005; Dobreva et al., 2006). During skeletogenesis, Satb2 physically interacts with RUNX2, ATF4 and SKI to regulate target gene expression (Baranek et al., 2012; Dobreva et al., 2006). Satb2 functions during these processes by mediating long-range interactions via recruiting HDAC1 and MTA at the distal enhancer regions (Gyorgy et al., 2008). In developing mouse brain, loss of function of Satb2, further results in the defective corpus callosum and an increase in subcortical connections (Alcamo et al., 2008).

SATB2 proteins exhibit a high level of conservation across vertebrates at the amino acid sequence level (Sheehan-Rooney et al., 2010). Such evolutionary conservation suggests conserved functions for SATB2 across vertebrate species. In zebrafish, satb2 is maternally deposited and widely expressed during the early stages of embryogenesis. However, as the embryo develops, *satb2* expresses strongly in the pronephric duct at 24 hpf, the branchial arches at 36 hpf, the ganglion cell layer of the retina, fins at 48 hpf and in neural structures at 72 hpf. In vertebrates, the pharyngeal arches involve a contribution from all three primary germ layers (Frisdal and Trainor, 2014). The pharyngeal arch is a major source of neural crest cells which contributes largely to craniofacial development. Therefore we hypothesize that Satb2 contributes to the regulation of cranial neural crest development essential for the development of the jaw and palatal regions through its transcriptional activity. Loss of function models in zebrafish would contribute largely to our understanding of early events in teleost's craniofacial development which is potentially regulated by Satb2.

3.2 Methods

3.2.1 Experimental models

Zebrafish (Danio *rerio*) were maintained as described previously (Westerfield, 2000). All the experimental procedures were carried out following the guidelines from the institutional animal ethics committee at IISER Pune and IST Austria. Embryos were raised at 23-31^{o}C in E3 medium and staged as described previously (Kimmel et al., 1995). Experiments with wild type SWR/J mice models (*Mus musculus*) were carried out at TIFR, Mumbai adhering to Institutional ethics guidelines.

3.2.2 CRISPR/Cas9 mutant generation

Synthetic oligonucleotides targeting exon1 of *satb2* gene using CRISPR-Cas9 method were designed using CHOPCHOP and cross-validated using CRISPRscan (Montague et al., 2014; Moreno-Mateos et al., 2015). Oligonucleotides required to generate sgDNA synthesized commercially as listed below (IDT, USA).

sgRNA oligo sequence targeting *satb2* (bold letters)	ATTTAGGTGACACTATA**GGGCCCCGTTGTGACGACTG** TTTTAGAGCTAGAAATAGCAAG
Constant oligo for synthesis of sgDNA	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAA CGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAA AAC
STOP cassette with homology arms (lower case)	cagccccactggtcaccgcagGTCATGGCGTTTAAACCTTAATT AAGCTGTTGTAGtcgtcacaacggggccccac

sgDNA template was generated using gene-specific oligonucleotides and constant oligonucleotide by polymerase chain reaction for 30 cycles using Q5 high fidelity polymerase (NEB, USA). PCR product was further purified using Magbio HiPrep PCR cleanup system (Magbiogenomics, USA) and quantified using Nanodrop 2000. sgRNA was synthesized *in vitro*

as per the suggested protocol using Ampliscribe T7 Flash Transcription kit (Lucigen, USA). 6xHis-Cas9 protein was synthesized as described previously (Gagnon et al., 2014). Briefly, plasmid coding for 6x-His-Cas9 was transformed in Rosetta DE3 Novagen competent cells (Merck Millipore, USA). Protein production was induced using an autoinduction medium (Formedium, UK) for 36 hrs at 18 ^{o}C with constant shaking at 200 RPM. Bacterial cells were harvested and lysed in 20 mM Tris pH 8.0, 30 mM Imidazole, 500 mM NaCl followed by sonication with 20 secs ON 20 secs OFF pulses for 30 mins using a probe-based sonicator, Sonics Vibra cell (Sonics, USA). Lysates were cleared using high-speed centrifugation and incubated with HisPur cobalt resin (Thermo Scientific, USA) for 1 hr. Beads were further washed three times with lysis buffer and eluted in 20 mM Tris pH 8, 200 mM Imidazole, 500 mM NaCl. Eluted fractions were further dialyzed in 20 mM Tris, 200 mM KCl, 10 mM $MgCl_2$ and stored as 5 μl aliquots at -80 ^{o}C for further use.

The sgRNA-Cas9 protein complex (sgRNA 200 ng/ul, Cas9 600 ng/ul and STOP cassette 1 μM) was injected at the one-cell stage of zebrafish embryos. Founders were obtained by genotyping using oligonucleotides flanking the target site and maintained as heterozygous.

3.2.3 Genotyping of Satb2 mutants

The embryos used for all the experiments were a result of $satb2^{+/-}$ incrosses. As a result, genotyping of the embryos after every experiment was necessary. DNA from embryos/larvae at desired stages of development were extracted in 50 μl of genomic DNA extraction buffer containing 10 μg proteinase K at 55^{o}C for 16 hrs (10 mM Tris-HCl pH 8.0, 50 mM KCl, 0.3% Tween-20 and 0.3% NP40). Proteinase K was inactivated by incubating at 95 ^{o}C for 20 mins followed by snap chilling. For adult fish, genotyping was performed using fin clips as described (Westerfield, 2000). PCR was carried out using forward primer: 5'GGAGGAGAGAGTCCTCGACTG3', Reverse primer: 5'GTTGCAGCATGTTTCAGATGAT3' with paq polymerase 5000 (Agilent, USA) and resulting PCR products were electrophoresed on 3% agarose gel. Gel images were captured using a G-Box gel documentation system (Syngene, UK).

3.2.4 Antibody production

All the procedures were performed as per the approved guidelines from the ethical committee at the national toxicology centre (NTC), Pune. To generate polyclonal antibodies, anti-zebrafish Satb2 (C-term): CIPSSGAEENPQANTGSGNNGP and anti-human SATB2 (C-term):

CQQSQPAKESSPPREEAP peptides were synthesized commercially (Apeptide, China). Antibodies were produced in New Zealand White Rabbits, as per the protocols from the laboratory of Tony Hyman, MPI-CBG with modification as below (https://hymanlab.mpi-cbg.de/hyman_lab/general/). Briefly, the required amounts of peptides were conjugated with KLH using glutaraldehyde followed by subsequent dialysis to remove glutaraldehyde. Conjugated peptides were mixed with Freud's complete adjuvant (Sigma Aldrich) for the first immunization. Rabbits were immunized intradermally. Further, after every 21 days, rabbits were immunized using peptides mixed with Freud's incomplete adjuvant until sufficient titre for the antibody was obtained.

Antisera were purified by a peptide affinity column prepared using sulfoLink coupling resin according to the manufacturer's instructions (Thermo) and stored in 50% glycerol solution at -20^{o}C.

3.2.5 Genome-wide occupancy analysis for zebrafish Satb2

I. ChIP setup:

To map the genome-wide occupancy of zebrafish Satb2, we performed ChIP-sequencing at desired stages of embryogenesis as mentioned in the results section and figure legends. To validate the use of in-house raised antibody for successful ChIP-seq experiment, we injected embryos with 25pg of 3xFLAG-Satb2 at 1 cell stage and harvested embryos at 4.5 hpf for further processing. ChIP was performed using anti-FLAG antibody (Sigma) and anti-zebrafish Satb2 antibody and subjected to high throughput sequencing as described below. A correlation matrix was generated to confirm the reproducibility of the two methods. For all stage-specific ChIP experiments, anti-zebrafish Satb2 was used to capture endogenous genomic binding regions.

Briefly, ChIP was performed using a modified protocol (Leichsenring et al., 2013). Approximately, 2000 embryos per stage were harvested in 1x E3 medium and homogenized using glass homogenizer loose piston in the presence of 1mM PMSF and immediately fixed with 1% methanol free formaldehyde (Thermo) at room temperature for 12 minutes with constant shaking. Fixation was stopped by the addition of glycine to a final concentration of 0.125 M and incubation at room temperature for additional 5 minutes. Fixed cells were centrifuged at 500 x g for 5 mins and washed thrice with 1x ice-chilled PBS by centrifugation at 4 degrees for 5 minutes each. Cells were further lysed in a 6 times bed volume cell lysis buffer (10 mM Tris-HCl pH7.5, 10mM NaCl, 0.5% (v/v) NP-40) for 10 minutes on ice and subjected to homogenization in

Dounce homogenizer for 10 times with loose piston followed by 3 times with tight piston. Nuclei were collected by centrifugation at 2500g, washed with ice-cold PBS and resuspended in 8 times the pellet volume in the nuclei lysis buffer (50 mM Tris-HCl pH7.5, 10mM EDTA, 1% (w/v) SDS, protease inhibitor cocktail (Roche, #11873580001)) and incubated further for 30 minutes on ice. The samples were sonicated using Covaris S2 under the following sonication conditions: 20% duty cycle, intensity = 5, cycle per burst = 200, Time = 40 cycles of 30 seconds ON 30 seconds OFF to obtain an average size of 200-300 base pairs. The samples were centrifuged at 12000 rpm for 10 minutes and the chromatin containing supernatant was stored at -80 ^{o}C till further use. Prior to the ChIP setup, the supernatant was precleared using a mixture of 1:1 Dynabeads protein A: G beads for 2 hrs at 4 ^{o}C. For each ChIP replicate and Mock reaction, 100 µg of chromatin was diluted 1:10 with ChIP dilution buffer (16.7 mM Tris-HCl pH7.5, 167 mM NaCl, 1.2mm EDTA, 0.01% (w/v) SDS, 1xPIC) and incubated with either 5µg of anti-FLAG antibody or 20µg of anti-Satb2 antibody was used. For Mock ChIP reactions, an equal amount of Rabbit IgG (#31235 Invitrogen) was used. The chromatin-antibody complex was incubated overnight at 4°C with constant rotating. Following day, Chromatin-Antibody-complex was captured using 100µl pre-blocked (with IgG-free BSA and t-RNA) Dynabeads® Protein A: G mix for 4 hrs at 4 ^{o}C. Beads were washed with ice-cold buffers in the following order: 7 minutes 4 times with low salt buffer (20 mM Tris HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), 10 minutes twice with high salt buffer (20 mM Tris HCl pH 8.0, 200 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), 10 minutes once with LiCl buffer (0.25M LiCl, 1mM EDTA, 10mM Tris HCl pH 8.0, 1% NP40, 1% Sodium deoxycholate) and 10 minutes twice with TE buffer (10mM Tris HCl pH 8.0, 1mM EDTA). The TE buffer was removed completely and 150µL of the elution buffer (0.1M NaHCO3, 1% SDS) was added and vortexed gently to solubilise the beads, followed by incubation at 65 ^{0}C for 30 minutes at 1000rpm. The eluate was collected in a fresh tube and the process was repeated with the addition of 150 µL of the elution buffer. To 300 µL of the eluates and 10% input, 20 µL of 5 M NaCl and 2 µL of RNAse A (10 mg/ml) was added and the samples were incubated at 65 ^{o}C overnight with constant shaking at 700-800 rpm. Next, 20 µL of 1 M Tris pH 8.0, 10 µL of 0.5M EDTA and 2 µL of Proteinase K (20 mg/ml) was added and the samples were further incubated at 42 ^{o}C for 1 hour at 700-800 rpm. Samples were purified by standard phenol: chloroform extraction method and DNA was precipitated overnight with an equal volume of 100% isopropanol in the presence of Glycoblue (Ambion) at -20 ^{o}C. DNA pellets

were eluted in nuclease-free water and quantified using Qubit fluorometer (Thermo) before proceeding with further analysis.

II. Library preparation, sequencing and data analysis:

An equal amount of DNA (~5 ng) was used as an input for library preparation and libraries were prepared using Nugen ultra2 ovation kit (Nugen). The number of cycles for amplification of adapter-ligated libraries was estimated by the qPCR method as described in the datasheet provided by the manufacturer. Final libraries were purified using HiPrep PCR clean up system (MagBio). Library concentration was determined using Qubit and average fragment size was estimated using DNA HS assay on bioanalyzer 2100 (Agilent) before pooling libraries at equimolar ratio. Sequencing reads (100bp PE) were obtained on the HiseqX platform at Macrogen Inc, Korea.

Sequencing reads were trimmed using TrimmomaticPE for Truseq2:PE adapters and read with quality greater than Phred 33 were retained (Bolger et al., 2014). The quality of sequencing reads was determined using fastQC. High-quality sequencing reads were aligned to zebrafish danRer10 genome version using default parameters of BWA (Li and Durbin, 2009). Aligned reads were subsampled to 40 million reads in each sample using BBmap. Correlation between each replicate was estimated using multiBamSummary and replicates showing very high Pearson correlation (> 0.7) were used for further analysis. Peak calling was performed using macs2 with default parameters and q value 0.05. A consensus set of peaks from biological replicates were extracted using a custom R script from Roman Cheplyaka (https://ro-che.info/articles/2018-07-11-chip-seq-consensus). Bigwig files were generated first using bamCoverage normalizing to RPKM and then subtracting Input signals using bamCompare utilities from deepTools 3.3.2 (Ramírez et al., 2016) and used for visualization with Integrative Genomics Viewer (IGV). Satb2 peaks in regulatory genes were extracted by intersecting with Ensembl of ATAC seq peaks (GSE106428, GSE130944, GSE101779). Further, Satb2 peaks were classified as putative enhancer bound peaks by intersecting Satb2 peaks with a consolidated dataset of H3K4me1 (GSE32483, GSE74231). Promoter bound peaks were assigned as +/- 2 kb from the TSS of the genes. Peak annotation to the nearest gene was performed using annotate.pl utility from Homer. Motif discovery for a given set of peaks was performed using findMotifsGenome.pl script from Homer and gene ontology was performed using webGestalt utilizing KEGG and Panther databases.

3.2.6 ChIP Western

The efficiency of ChIP for mouse SATB2 and validation of mammalian SATB2 antibody was performed by ChIP western assay. Briefly, ChIP was performed with in house anti-human Satb2 antibodies. After bead washing and removal of excess TE buffer, beads were resuspended in 1x PBS and treated with 2 µl of DNAse (10 mg/ml) for 30 mins at 37^{o}C. Further, beads were washed with 1x PBS twice and boiled in 2X Laemmli buffer (0.25 M Tris-Cl pH 6.8, 1% SDS, 1% β-mercaptoethanol, 15% glycerol) at 98 ^{o}C for 5 minutes and electrophoresed on 7.5% of SDS-PAGE gel. Proteins were transferred to the 0.45 µM PVDF membrane (Millipore) by wet transfer method applying 0.6 A current for 2.5 hrs at 4 ^{o}C. Non-specific sites on the membrane were blocked using 3% BSA and further incubated with anti-SATB2 antibody (1:1000 Abcam ab34735) in 0.3% BSA overnight at 4 ^{o}C with constant rocking. The next day, the membrane was washed with 1X TBST (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% Tween-20) 4 times for 7 minutes each and incubated with anti-Rabbit HRP conjugated antibody (1:10000 STAR124P Bio-Rad) in 1x TBST for 45 min at room temperature. After removing excess of secondary antibody by repeated washes with 1x TBST, the signal was developed with Clarity western ECL substrate and captured using LAS 4000 system (GE Healthcare).

3.2.7 Micro-CT image acquisition and analysis

Satb2$^{+/-}$ fish were incrossed and were grown until 4 months post-fertilization. Fish were fin-clipped and genotyped using the protocol described above. Screened fish (N=5 per group) were fixed in a 3.7% paraformaldehyde and 1% glutaraldehyde in 1X PBS solution, fish were fixed for 72 hours at room temperature and then transferred to 1% of fixative solution for storage. Fish were imaged first using an SLR camera for documenting the overall morphology. The internal bony structures were then imaged using a Quantum GX microCT imaging system (Perkin Elmer, USA). Image acquisition settings were as follows, Temperature: 23^{o}C, Humidity: 57%, Voltage: 90 kVp, Current: 88 uA, X-Ray dosage: 449 mGy, FoV: 36mm, Recon: 36 mm, Voxel size: 72 um, Scan mode: high resolution, Time: 4 minutes. Images were further processed with ImageJ (FIJI) for adjusting brightness and contrast.

3.2.8 Alcian blue - Alizarin red double staining

Alcian blue: Alizarin red acid-free double staining was performed as described previously (Walker and Kimmel, 2007). All solutions were stored, and all steps were carried out at room temperature. Briefly, embryos from satb2$^{+/-}$ incrosses were grown to 15 dpf and fixed with 4% PFA for 2 hours. After dehydration for 10 min in 50% ethanol, larvae were stained overnight in a double staining solution made from a mixture of Alcian blue: Alizarin red stock solution in a 100:1 ratio respectively. Alcian blue stock solution for cartilage staining contained 0.1% Alcian blue powder (Sigma Aldrich) in 70% ethanol and 60 mM $MgCl_2$ final concentrations. Alizarin red (Sigma Aldrich) stock solution for bone staining was made at a final concentration of 0.5% in MQ water. After staining, the solution was removed by a quick rinse with MQ water followed by washes with a 1:1 solution of 3% H_2O_2 and 2% KOH for 20 min with the tube caps kept open. Larvae were then cleared with successive washes with increasing concentration of glycerol (20-50%) and 0.25% KOH solutions. Larvae were finally stored in a 50% glycerol and 0.1% KOH storage solution until imaging. At the time of imaging, larvae were indexed and temporarily mounted with the help of 75% glycerol and 0.1% KOH solution and imaged using an Olympus light microscope. After 3 subsequent washes with PBST to remove the bone stain, selected larvae were genotyped for representation.

3.2.9 3'mRNA gene expression assay and differential gene expression analysis

Single embryos were harvested in 100 µl RNA iso-plus total RNA extraction reagent (DSS Takara, India) at 80% epiboly, 6 somites and 14 somites stage of zebrafish embryonic development. Samples were homogenized by vigorous vortexing to assure complete lysis followed by the addition of 15 µl of $CHCl_3$. The aqueous layer was separated by centrifugation at 12000x g for 15 min at 4 ^oC. The aqueous layer was collected in a separate tube and an equal volume of isopropanol containing GlycoBlue co-precipitant (Invitrogen, USA) was added. The aqueous fraction was stored at -20 ^oC until further processing. Meanwhile, the organic fraction was processed for DNA extraction by precipitating with 150 µl of 100% ethanol and incubated for 10 mins at RT. Samples were centrifuged for 10 min at high speed and the pellet was washed with 70% ethanol (twice). After drying, the DNA pellet was resuspended in low TE and used for genotyping as mentioned before. Once the genotype for the single embryo is confirmed,

corresponding RNA fractions were processed further. We pooled 2-3 samples to obtain sufficient amounts of RNA to prepare libraries.

Extracted RNA samples were quantified using Nanodrop 2000 and RNA integrity was determined using Agilent bioanalyzer 2100. Only samples with RIN > 8 were used to prepare sequencing libraries using Quantseq 3' mRNA seq library kit (Lexogen GMBH, Austria) as per the instructions from the manufacturer. Briefly, 300 ng of RNA was used for polyA capturing and subjected to first and second-strand synthesis. All cDNA samples were amplified for 10-15 cycles depending on cycle number estimation by qPCR (PCR add on kit, Lexogen GMBH, Austria). Amplified libraries were purified using two rounds of 0.8x volume of Hi-prep PCR purification kit. The concentration of libraries was estimated using the Qubit DNA HS system. Finally, all the libraries were pooled and subjected to high throughput sequencing using 76 bp SE chemistry on Nextseq 550 sequencer.

Sequencing reads were trimmed for quality using Trimmomatic and aligned to daRer10 genome assembly using STAR aligner. Counts for each gene feature was estimated using the FeatureCounts package from Rsubread. Differential expression analysis was performed for replicates using EdgeR. A volcano plot for significantly differentially expressed genes was generated using the PlotVolcano tool from the Galaxy server. Gene ontology analysis for upregulated and downregulated genes were performed using a web-based tool, gProfiler (Reimand et al., 2007).

3.2.10 Synthesis of riboprobes for Whole-mount in situ hybridization

Synthetic oligonucleotides against target mRNAs (listed below) with T7 RNA polymerase promoter sequence on the 5' end of the reverse primers were obtained commercially (Sigma Aldrich, India). cDNA templates for corresponding stages were used for PCR amplification. PCR reactions were purified using NEB Monarch DNA Gel Extraction Kit or HighPrep PCR Clean-up System beads. Further, Riboprobes were synthesized using Roche DIG RNA labelling kit at 37^{o}C for 3 hrs followed by DNAse treatment for 15 mins. Probes were further purified using Bio-Spin 30 Tris Columns (Biorad, Germany). The digoxigenin-labelled probes were 400-700 nucleotides in length and stored at -20 ^{o}C at 1 μg/ml.

For *chrd* WISH experiments amplified region was cloned into TOPO II dual promoter vector (Thermo Scientific, USA). To generate an antisense probe, the plasmid was linearized using BamHI and probe synthesis reaction was carried out using T7 RNA polymerase promoter as described above.

Target gene	Forward primer (5'→3')	Reverse primer (5'→3')	Product size
id2a	AGGCGAGTCTTTTCAACGAA	gagtaatacgactcactataggGATATTTGACGGGACGCTGAG	688 bp
sox9a	CGCAGAATCTCCTCGACCC	gagtaatacgactcactataggGGGGACTGGCCTGAGTGTTCG	700 bp
zic1	TTCTTTTTCGCAATCGGGGC	gagtaatacgactcactataggGAAGGTTTTTCACCTGTGTGTGT	687 bp

3.2.11 Whole-mount *in situ* hybridization analysis

Whole-mount *in situ* hybridization (WISH) was performed as described previously (Thisse et. al., 2008). Fertilized embryos from satb2$^{+/-}$ incross were raised until the desired stage of development was reached. Embryos were fixed in 4% paraformaldehyde in 1x PBS overnight at

4 ^{o}C with gentle rocking. Next day, embryos were hand dechorionated with a pair of fine forceps and incubated in 100% MeOH overnight at -20^{o}C. Embryos were further downgraded in methanol: PBST solution (0.1% Tween-20 in 1x PBS pH 7.4). Following successive rehydration to PBST, embryos were treated with 5ug/ul proteinase K solution in PBST for 2 min for 14ss stage embryos. The reaction was stopped by a subsequent 20 min incubation in 4% PFA. For the Dome stage, embryos were not treated with proteinase K. After washes in PBST, samples were pre hybridized for 2-5 hrs and incubated overnight at 70 ^{o}C in approximately 500 ng of the respective probe. The next day, embryos were washed with SSC buffers and further blocked in blocking buffer (2% v/v normal goat serum, 2 mg/ml BSA in PBST) for 3-4 hrs at room temperature followed by overnight incubation at 4 ^{o}C in anti-DIG-AP Fab fragments (1:5000) (Roche 1093274). Embryos were washed with PBST followed by an alkaline tris buffer. Staining was performed with BM Purple substrate (Roche). After staining, the colour reaction was stopped using a stop solution containing 20% Tween-20 and 05M EDTA in 1x PBS (pH 5.5). Embryos were further washed with PBST and stored at 4^{o}C until imaging. Embryos for imaging were mounted on agarose moulds and imaged using Z-stacking mode on a Leica DFC450C microscope. Post imaging, embryos were genotyped using the standard protocol discussed above. Representative images were further processed using ImageJ for adjusting brightness and contrast.

3.2.12 Molecular cloning and protein expression

For recombinant expression of zebrafish Satb2 protein, coding sequences were amplified using Q5 high fidelity DNA polymerase (NEB, USA) and cloned into a 6xHis-PET28-Strep vector (Kind gift from Dr Thomas Pucadyil, IISER Pune) using Gibson assembly. PCR product was treated with DpnI and transformed into *E. coli* DH5α bacterial strain for *in-vivo* ligation. The resulting clones were screened by colony PCR. The recombinant protein was expressed in an autoinduction medium (Formedium, UK) for 36 hrs at 18 ^{o}C with constant shaking at 200 RPM. Bacterial cells were harvested and lysed in 20 mM Tris pH 8.0, 30 mM Imidazole, 500 mM NaCl followed by sonication with 20 secs ON 20 secs OFF pulses for 30 mins using a probe-based sonicator, Sonics Vibra cell (Sonics, USA). Lysates were cleared using high-speed centrifugation and incubated with HisPur cobalt resin (Thermo Scientific, USA) for 1 hr. Beads were further washed three times with lysis buffer and eluted in 20 mM Tris pH 8, 200 mM Imidazole, 500 mM NaCl. Eluted fractions were further dialyzed in 20 mM Tris, 200 mM KCl, 10 mM $MgCl_2$.

3.2.13 ATAC-seq and data analysis

Single embryos at 14 somite stages from the incross of *satb2*$^{+/-}$ were hand dechorionated and harvested in 50 µl of ice-chilled 1x DPBS (Invitrogen). 5 µl of cell suspension was used for identifying genotypes of each embryo. Cell suspension from 3 embryos was pooled together and processed for Omni-ATAC seq as described previously (Corces et al., 2017) with modification from Amanda Ackermann lab. Briefly, cells were washed with 1x DPBS and resuspended in cell lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM KCl, 0.1 % NP40, 0.1% Tween20 and 0.01% Digitonin) and incubated on ice for 3 minutes. Further, cells were washed with a wash buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM KCl and 0.1% Tween20) by centrifugation at 500 for 10 mins 4^{o}C. The supernatant was discarded and the pellet was resuspended in 25 µl 2x Tagmentation buffer (Illumina, catalogue # 15027866), 16.5 µl DPBS, 0.5 µl 10% Tween 20, 0.5 µl 1% Digitonin, 5 µl nuclease and 2.5 µl Tn5 transposase enzyme (TDE1, Illumina, catalogue #15027865) and incubated for 28 minutes at 37^{o}C. After the tagmentation reaction, DNA was isolated using the DNA clean up kit (Zymo). Purified DNA was used as an input to generate a library by amplifying with 2x Q5 DNA polymerase mix (NEB) and indexing primers. Optimal cycles were determined using qPCR analysis. Amplified libraries were purified using Agencourt ampure XP beads to remove adapters and larger fragments.

Sequencing reads (41 bp PE) were obtained on NExtseq 550 at IISER Pune and trimmed for Nextera adapters using default parameters of Trimmomatic PE. Trimmed reads were aligned to danRer10 using default parameters of Bowtie2 (Langmead and Salzberg, 2012). Briefly, BAM files were subsampled to 55 million reads in each sample using bbmap and sorted by name. paired-end bed files were obtained using bedtools bamtobed. Reads were displaced by +4 bp and -5 bp. Peak calling was performed using macs2 callPeak -f BEDPE -q 0.05 --nomodel --extsize 200 --gsize 1.3e9 --keep-dup 2 parameters. Consensus peaks were obtained using a custom R script used for ChIPseq analysis. BigWig files were generated using bamCoverage (deepTools). Peaks were annotated to the nearest gene using Homer and classified into promoter (+/- 2 Kb) and non-promoter regions. K-means clustering was performed around +/- 2 Kb of Satb2 peak centre using deepTools. Clusters were annotated using Homer and gene ontology analysis was performed using webGestalt. Motif analysis was performed using findMotifsGenome.pl from Homer.

3.2.14 Differential chromatin accessibility analysis (DiffBind)

The differential chromatin accessibility analysis of Satb2 mutant and wild type embryos at 14 somite stages was performed using the DiffBind (Stark and Brown, 2011). Significantly differentially accessible peaks were identified using the Deseq2 package and only sites with FDR < 0.05 and fold change of > Log2 (+/- 1.5) were used for further analysis. Differential accessible sites were annotated to the nearest gene using Homer. Core promoter was defined as +/- 2 Kb from the TSS.

3.2.15 Nucleosome occupancy analysis

Nucleosome occupancy analysis was carried out using the nucleoATAC suite with default parameters for Satb2 mutant and wild type ATACseq datasets. Nucleosome fuzziness scores were obtained and used for calculating the difference in nucleosome phasing upon loss of function of Satb2.

3.2.16 Protein extraction and immunoblotting

To confirm the absence of Satb2 in zebrafish larvae, they were manually dechorionated, deyolked and harvested at 48 hpf by boiling in 2X Laemmli buffer (0.25M Tris-Cl pH 6.8, 1% SDS, 1% β-mercaptoethanol, 15% glycerol) at 98^{o}C for 5 minutes. The proteins in the resultant lysate were resolved by electrophoresis on 7.5% of SDS-PAGE gel. Proteins were transferred to the 0.45 µM PVDF membrane (Millipore) by wet transfer method at 0.6A for 2.5 hrs at 4^{o}C. Non-specific sites on the membrane were blocked using 3% BSA and further incubated with in-house anti-SATB2 antibody (1:500, this study) in 0.3% BSA overnight at 4^{o}C with constant rocking. The next day, the membrane was washed with 1X TBST (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% Tween-20) 4 times for 7 minutes each and incubated with anti-Rabbit HRP conjugated antibody (1:10000 STAR124P Bio-Rad) in 1x TBST for 45 minutes at room temperature. After removing excess of secondary antibody by repeated washes with 1x TBST, the signal was developed with Clarity western ECL substrate and captured with LAS 4000 system (GE healthcare).

3.2.17 Statistical analysis

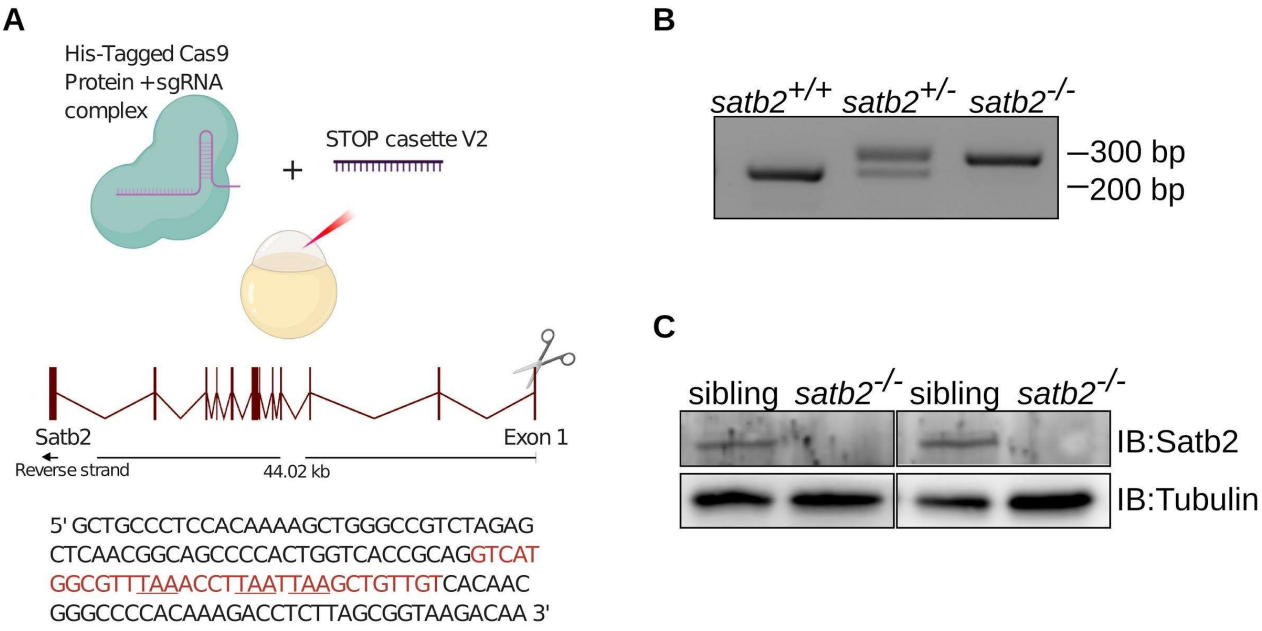
GraphPad Prism (GraphPad, USA) was used for statistical analysis as indicated in legends.

3.3 Results

3.3.1 Generation of *satb2* mutant using CRISPR-Cas9 system

While SATB2 has been shown to regulate multiple developmental processes, molecular mechanisms underlying its activity remain largely unexplored (Alcamo et al., 2008; Britanova et al., 2008; Dobreva et al., 2006; Döcker et al., 2014; Gong et al., 2014; Leone et al., 2008, 2015; Savarese et al., 2009). In zebrafish embryos, compromising Satb2 activity using morpholinos resulted in severe developmental defects including epiboly arrest (Ahn et al. 2010, Chapter 2). These phenotypes were attributed to aberrant exocytosis and endocytosis (Ahn et al. 2010). The broad and possibly pleiotropic nature of the early epiboly arrest phenotype induced by the morpholinos is not highly informative about the precise function of Satb2. This further underscores the limitations of using morpholinos for understanding the functions of Satb2 during organogenesis. Thus, to uncouple the maternal versus zygotic requirement of Satb2, we introduced a mutation in *satb2* locus using the CRISPR-cas9 system (Gagnon et al., 2014) by inserting a STOP cassette into the first exon, leading to premature termination (Figure 3.1A and Figure 3.1B). The absence of Satb2 in the mutant fish was confirmed by immunoblot analysis (Figure 3.1C).

Figure 3.1 Generation of *satb2* mutant using CRISPR-Cas9 system.



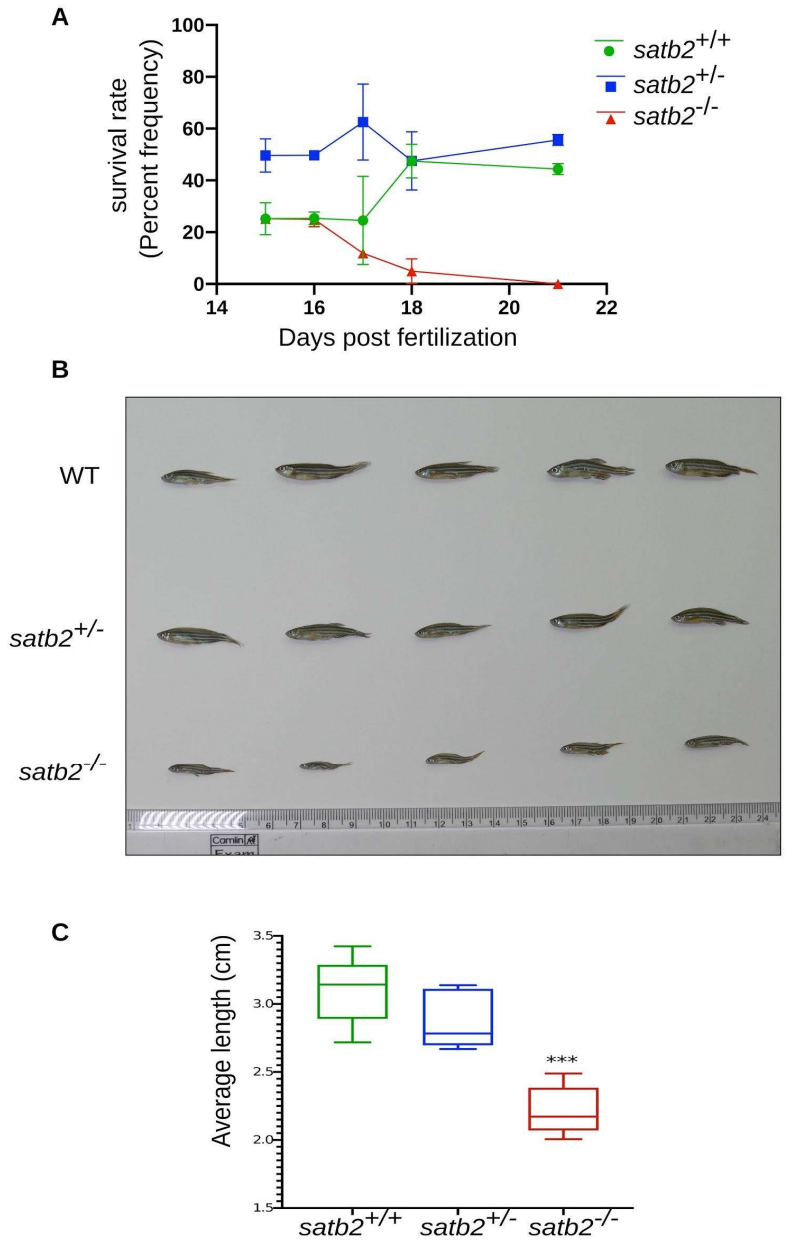
(A) Schematic of CRISPR-cas9 mediated mutant generation by introducing the STOP cassette (highlighted in red) in exon1 of the *satb2* gene resulting in loss of function mutant allele. **(B)** Identification of mutant allele by genotyping (60 bp insertion) and **(C)** confirmation of loss of Satb2 between siblings and homozygous larvae at 48 hpf by immunoblot. Tubulin was used as a loading control.

3.3.2 Phenotypic characterization of *satb2* mutants

Mutation in *satb2* led to premature termination, However, surprisingly, this did not result in an early developmental arrest as was in the case with the previously generated loss of function models in zebrafish (Ahn et al. 2010, Chapter 2). This could be presumably due to contribution from maternally deposited Satb2 or due to activation of compensatory mechanisms as previously described (Rossi et al., 2015). However, as mentioned above, this allowed us to study the zygotic function of Satb2 independent of maternal contribution.

Homozygous zygotic *satb2* mutants showed poor survival to adulthood with most of the *satb2*$^{-/-}$ fish being depleted from the pool after 16 days post fertilization (Figure 3.2A). However, we could occasionally recover homozygous survivors (~1%) from the total population (n=400). Such survivors displayed severe deformities and reduced body size compared to their heterozygous siblings (Figures 3.2B and 3.2C). The homozygous fish were also sterile.

Figure 3.2 *satb2*$^{-/-}$ mutants show poor survival, severe deformities and reduced body size.



(A) Lifespan analysis of zygotic *satb2* mutants depicted by a line graph. Wild-type (green), Heterozygous (blue) and Homozygous mutants (Red). Error bar represents +/- S.D. of two independent experiments with n=48. **(B)** Digital snapshots of 4 months old zebrafish Satb2 mutants along with wild-type and heterozygous siblings highlighting defects in body growth. **(C)** Box plot indicating the difference in body length between Satb2 mutants and siblings (n=5, ** p-value <0.05).

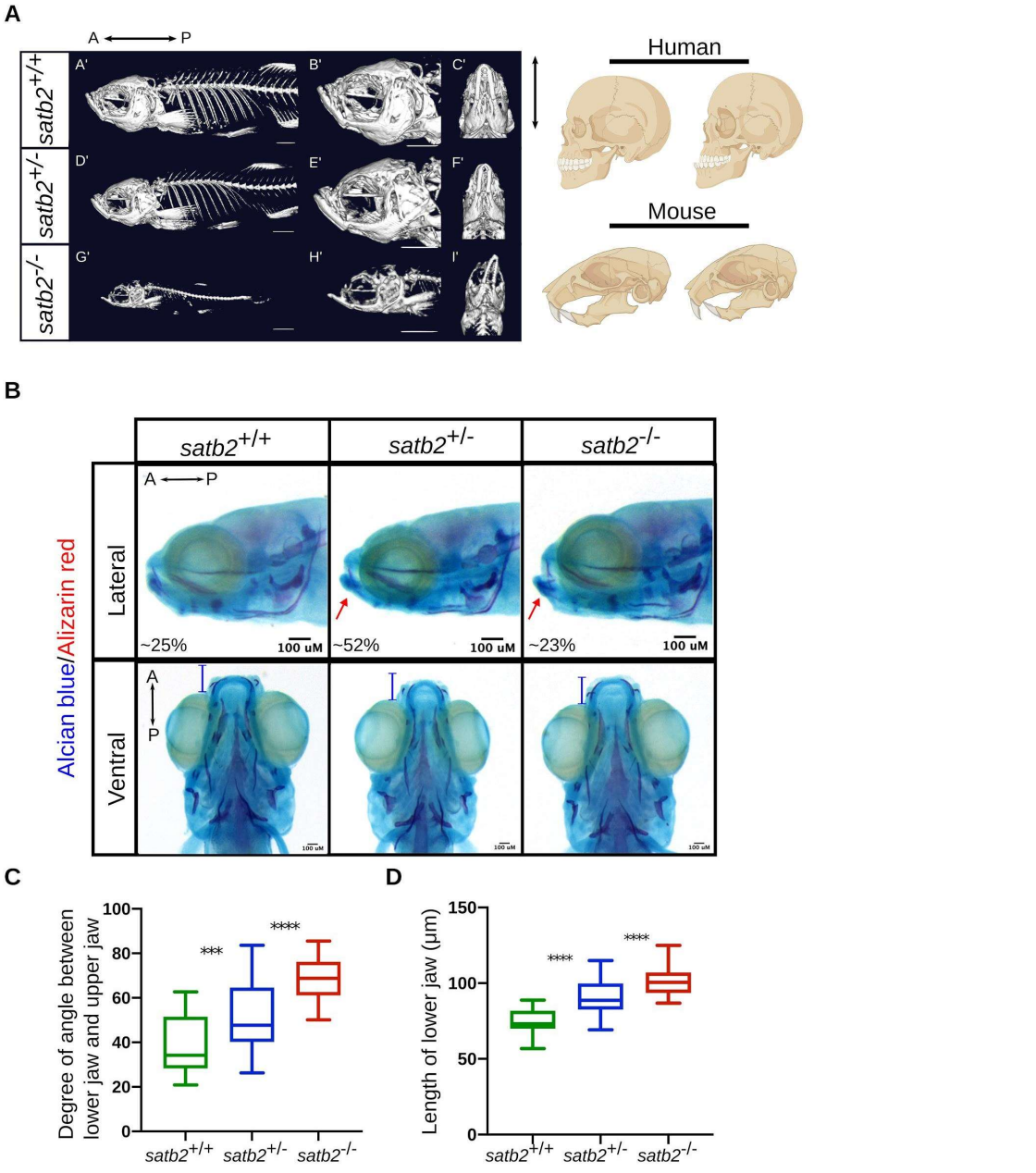
3.3.3. Disruption of Satb2 function leads to defective craniofacial patterning

Considering previously reported functions of Satb2 in craniofacial development and bone formation in mammalian systems, we recorded CT scans of homozygous survivors, heterozygous mutant siblings, and wild-type fish (n=5). Consistent with the effects of mutations in mammalian homologs, *satb2* mutant fish displayed deformed mandibles with a high degree of lower jaw protrusion, reduced maxilla with larger eye pockets and abnormal skull size and shape (Figure 3.3A). Supporting the haploinsufficient nature of Satb2 function, heterozygous fish also showed less severe yet detectable cranial structural abnormalities. Thus, zebrafish *satb2* mutant mimics the craniofacial defects observed in humans and mice (Figure 3.3A) and could serve as an excellent model to delineate underlying molecular mechanisms of craniofacial development (Leoyklang et al., 2007).

To trace back the onset of these defects during early development, we performed alcian blue: alizarin red staining on larvae obtained by mating *satb2* heterozygous adults. Larvae were collected 15 days post-fertilization, a day prior to depletion of zygotic *satb2*$^{-/-}$ fish from the pool. As in the case of adult survivors, we observed craniofacial defects in the homozygous larval samples (Figure 3.3B). We obtained 25% of wild-type, 52% of heterozygous and 23% of homozygous mutants in accordance with the mendelian ratio. We found an increase in the angle between mandibles and maxilla in heterozygous and homozygous as compared to wild-type siblings (Figure 3.3C). Further, Measurements of the total length of the lower jaw confirmed significantly higher protrusion in zygotic s*atb2* mutants (Figure 3.3D).

Collectively, these results suggest that the primary role of zebrafish zygotic Satb2 in cranial and skeletal morphogenesis is conserved across vertebrates and any perturbation in Satb2's function leads to severe developmental defects.

Figure 3.3 Loss of function of Satb2 in zebrafish leads to craniofacial defects and phenocopy function of mammalian homologues.



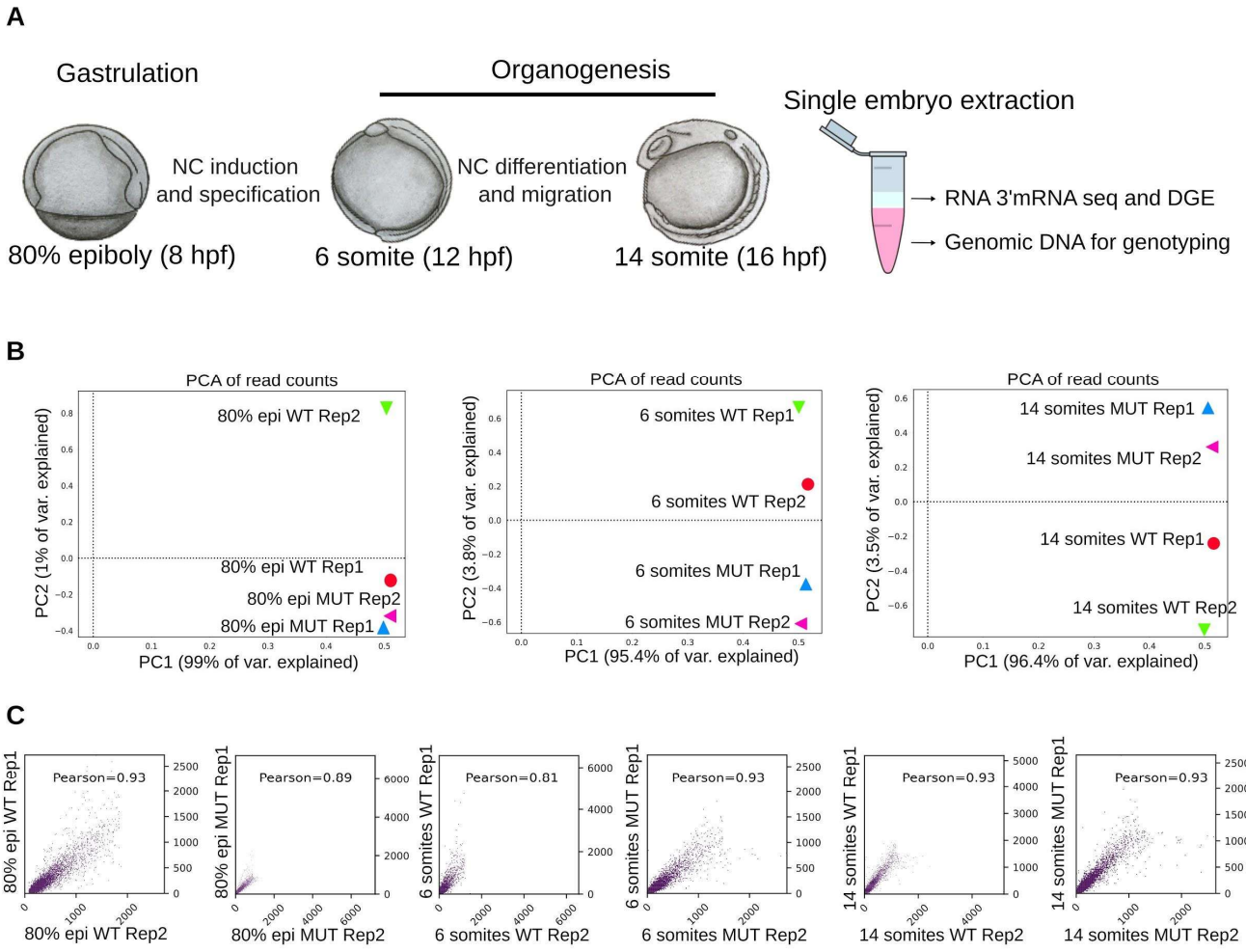
(A) Micro CT images of adult zebrafish (4 months old) to visualize defects in skeletal and craniofacial structures. Lateral view A', D', G', magnified lateral view B', E', H' and ventral view

C', F', I' of Wild-Type, Heterozygous and Mutant for Satb2 respectively are shown. Ventral view images are not scaled to attain a maximum field of view highlighting detailed structural deformities independent of the size of fish. Image is representative of N=5. Corresponding schematics depicting similarity with *SATB2* mutation reported in humans and mice. **(B)** Alcian blue/Alizarin red staining to visualize craniofacial defects at early larval stages, 15 dpf. Abnormal jaw protrusion in heterozygous and homozygous mutants is indicated by a red arrow. Numbers in respective boxes signifies the percentage of the larvae from the total population showing class of phenotype and genotype correlation. Degrees of the angle between the lower jaw and upper jaw **(C)** and a total length of protrusion measured from ventral view **(D)** as indicated in panel F, are plotted using a box plot. *** indicates p-value of significance 0.001 and **** indicates p-value of significance 0.0001 as determined by student's two-tailed t-test.

3.3.4 Satb2 potentiates early neurogenesis and neural crest specification

Aberrant neural crest specification and migration have been correlated with severe craniofacial defects (Cordero et al., 2011; Trainor, 2010). In the developing larvae, *satb2* is strongly expressed in the pharyngeal arches, a reservoir of NCCs (Sheehan-Rooney et al., 2010). This prompted us to investigate the possible involvement of Satb2 in NC development. We selected three developmental time points that correspond to specific stages of NC differentiation. To study NC induction, NC specification and differentiation we analysed gastrulating embryos at 80% epiboly, embryos at 5-6 somites stage (ss) and 14 ss respectively (Knight and Schilling). We analysed gastrulating embryos at 80% epiboly for understanding early specification events, 5-6 somites stage embryo at which neural crest cells get specified and to analyse the effect on neural crest cell differentiation we used 14 somites stage of embryo development (Figure 3.4A). PCA plots and Pearson correlation analysis for all stages showed significant reproducibility between the replicates (Figure 3.4B and 3.4C).

Figure 3.4 Schematic for 3’ mRNA seq analysis and corresponding correlation analysis.

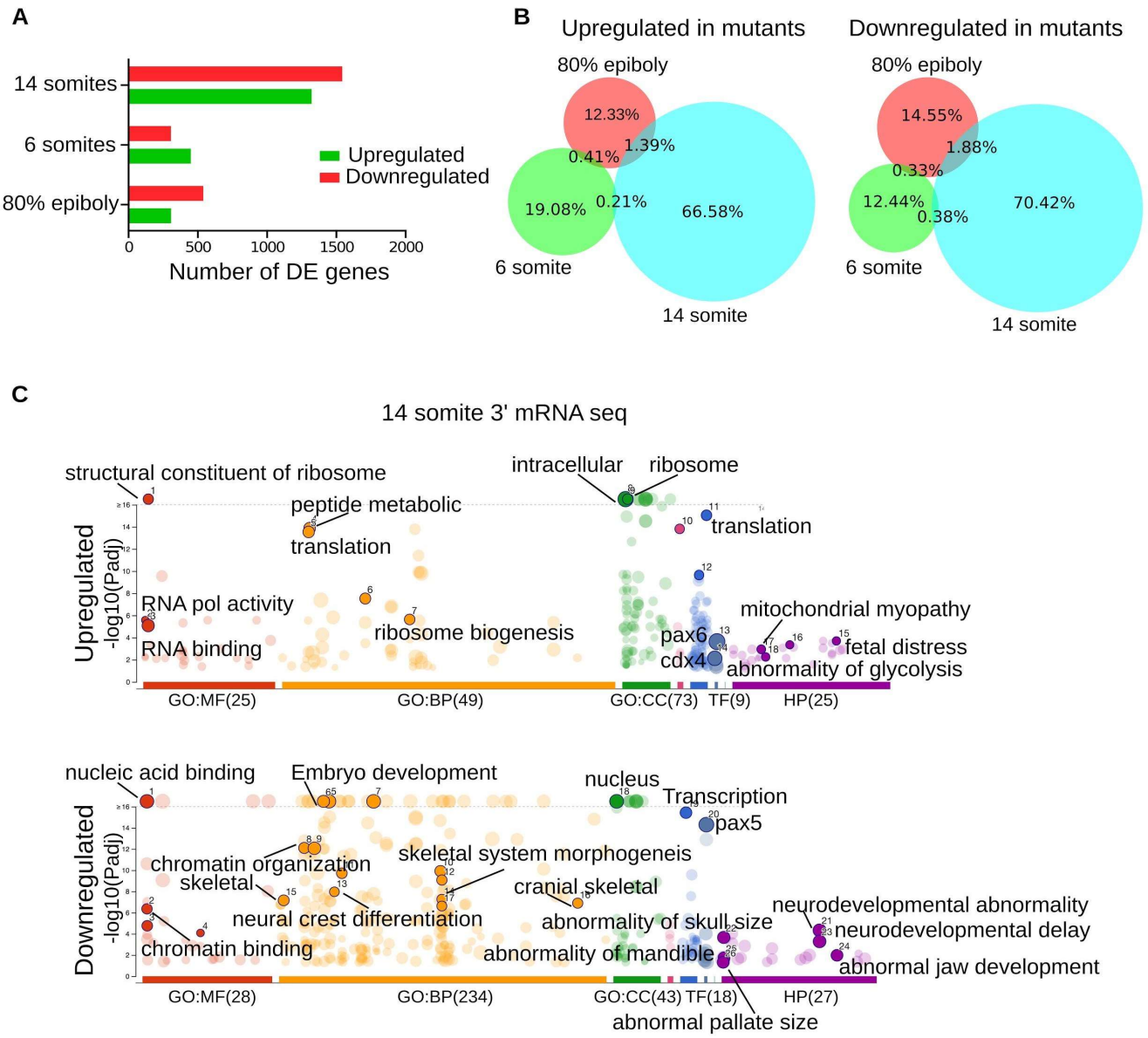


(A) Experimental strategy for gene expression analysis using 3’ mRNA sequencing at 80% epiboly, 6 ss and 14 ss. **(B)** PCA plots showing highlighting reproducibility between the replicates of 3’ mRNA seq at 80% epiboly, 6 ss and 14 ss respectively. **(C)** Pearson correlation analysis between replicates of transcriptome samples at the corresponding stages.

Transcriptome analysis revealed that loss of Satb2 function resulted in deregulation of numerous genes across all developmental stages (658 at 80% epi, 661 at 6 ss, 3015 at 14 ss). However, the maximum effect both at gene expression level and in terms of a number of deregulated genes was observed at 14 somites, suggesting Satb2 might play an essential role during migratory stages of neural crest differentiation (Figure 3.5A). Interestingly, we observed minimal overlap between differentially expressed genes at different stages emphasizing stage-specific functional outcome of Satb2 mutation (Figure 3.5B).

Gene ontology analysis of differentially regulated genes suggested, upregulated genes upon loss of function of Satb2, in other words, genes which are repressed by Satb2 in wild type condition belong to RNA polymerase activity, ribosome biogenesis and mRNA translation. Interestingly, these genes were associated with human pathological conditions such as mitochondrial myopathy and abnormality in glycolysis suggesting a novel functional role for Satb2. Importantly, downregulated genes show significant enrichment for skeletal morphogenesis, neural crest differentiation and early embryonic development. Predictably, these genes were associated with neurodevelopmental and jaw abnormality in human pathology, again providing sufficient evidence that zebrafish Satb2 shows conserved functional regulation (Figure 3.5C).

Figure 3.5 Deregulation of gene expression in *satb2* mutants at the onset of organogenesis leads to defective neurogenesis and craniofacial patterning.

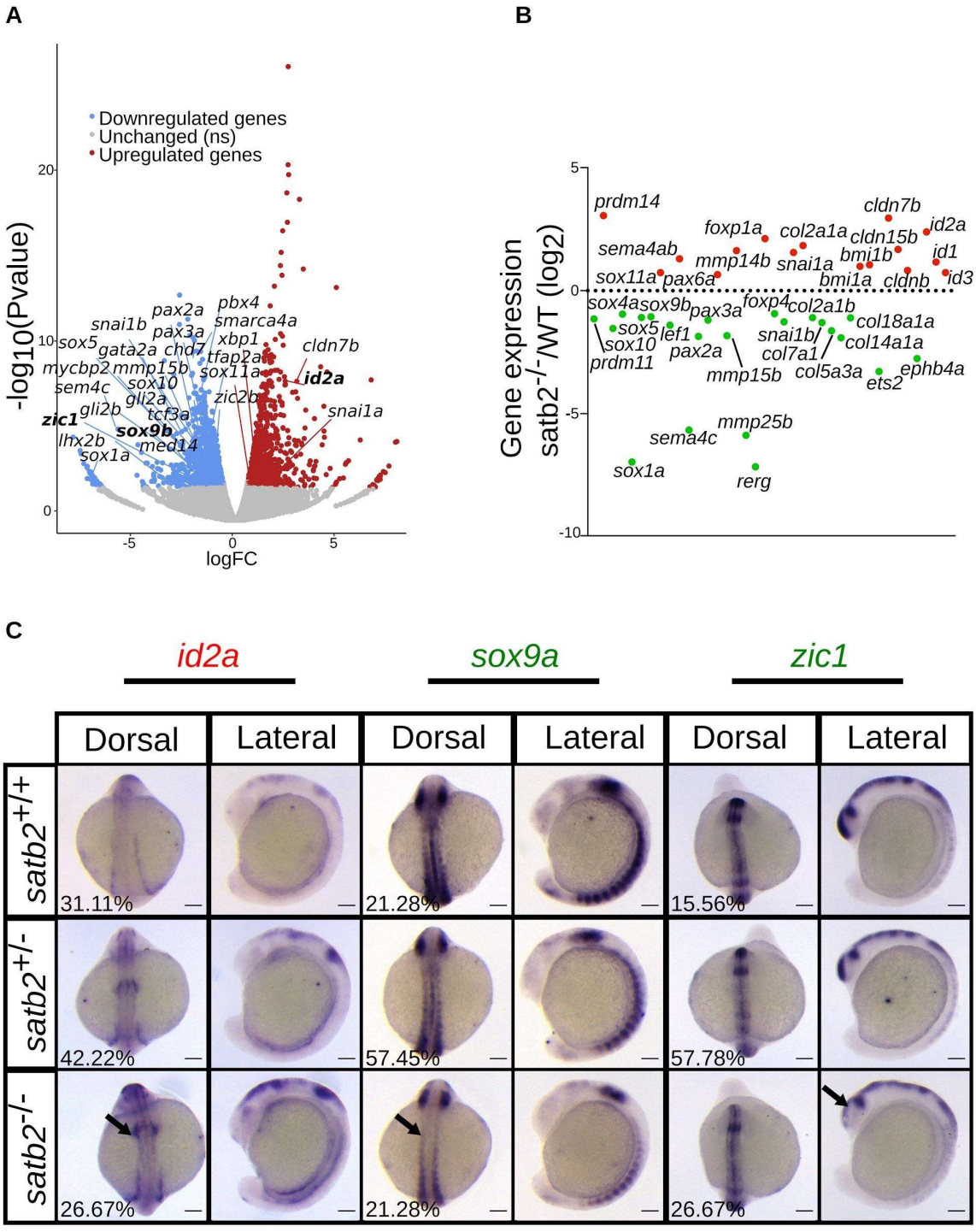


(A) Bar plot for the number of upregulated (red) and downregulated (green) genes at respective stages. **(B)** Venn diagram showing percentage overlap between differentially expressed genes in *satb2*$^{-/-}$ at different stages. **(C)** GO analysis of dysregulated genes at 14 ss highlighting significant classes under molecular function (MF), biological processes (BP), Cellular components (CC), Transcription factors (TF) and Human pathology (HP) categories.

Most of the genes which showed significant deregulation were known molecular markers of neurogenesis and neural crest differentiation such as *sox10*, *sox9a*, *pax3a*, *zic1* to name a few (Figure 3.6A) (Carney et al., 2006; Dutton et al., 2001; Garnett et al., 2012; Sato, 2005; Van Otterloo et al., 2012; Yan et al., 2002). Invariably, Satb2 and Sox9 have been observed to work together through cis-regulation to control jaw development (Rainger et al., 2014).

Consistent with the selective nature of Satb2's function, many paralogs within the gene families such as *prdm (prdm11, prdm14)*, *foxp (foxp4*, *foxp1a)* and *snai (snai1 and snai1b)* were differentially regulated by Satb2 (Figure 3.6B). Whole-mount in situ analysis for *id2a*, *sox9a* and *zic1* confirmed readout from gene expression studies. Satb2$^{-/-}$ mutants showed severe loss of expression of *sox9a* in the neural tube and *zic1* in the developing head. However, *id2a* showed significant upregulation in expression as compared to wild type siblings (Figure 3.6C).

Figure 3.6 Satb2 positively regulates neural crest development programs.



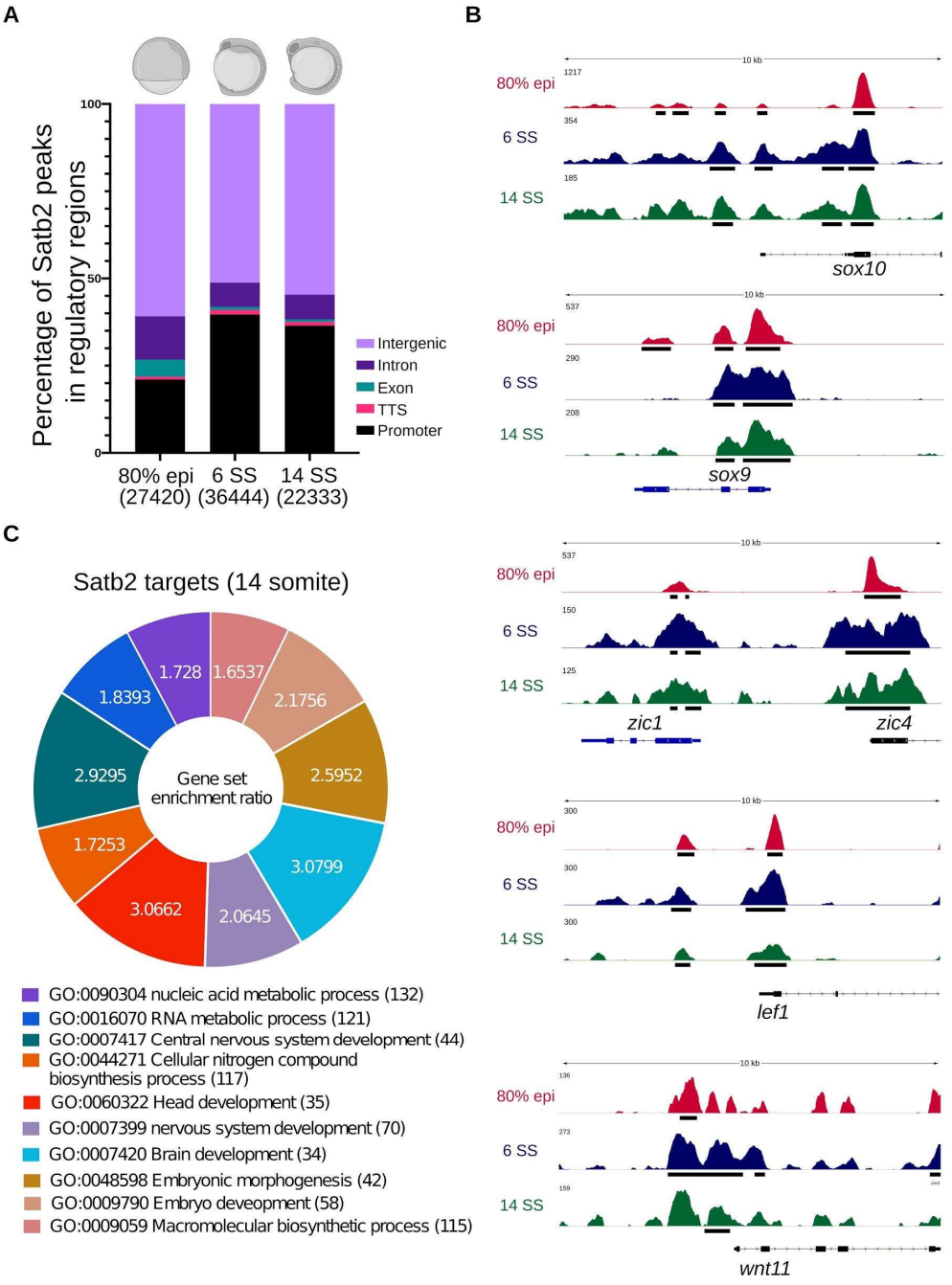
(A) Volcano plot of differentially expressed genes at 14 ss embryonic stage. DE genes with FDR < 0.1 and log2 Fold change > +/- 0.58 are coloured as red for upregulated, blue for

downregulated and grey for unchanged or non-significant genes. Few significantly differentially expressed genes representing GO categories neural crest differentiation, neurogenesis and cranial skeletal system development are highlighted. **(B)** Scatter plot for Log_2 Fold change in gene expression of various gene families belonging to the above-mentioned gene ontology categories. **(C)** Dorsal and lateral view of WISH for *Id2a* (upregulated), *sox9a* and *zic1* (downregulated). Arrows in the respective images mark the region of maximum differences in the expression pattern.

3.3.5 Genome-wide occupancy analysis confirms direct regulation of downstream target genes by Satb2

To investigate the molecular basis of gene regulation by Satb2, we performed chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) for the corresponding stages of embryonic development at which we had performed transcriptome analysis. We performed ChIP-sequencing using the anti-Satb2 antibody in duplicates and observed a very high correlation. To be able to distinguish functional Satb2 binding sites, we intersected consensus peaks with Ensembl of accessible regions of chromatin across various developmental stages (GSE106428, GSE130944, GSE101779). Satb2 ChIP-seq yielded 27420 binding sites at 80% epiboly, 36,444 sites at 6 somites and 22,333 sites at 14 somites (Figure 3.7A). Consistent with the studies concerning other factors involved in the NC specification program, we also observed a significant increase in the number of peaks at 6 somites, presumably due to the highly dynamic nature of chromatin at this stage (Lukoseviciute et al., 2018). Genomic distribution analysis for these peaks suggested Satb2 occupies both the promoter and non-promoter regions (intergenic + intronic + exonic + TTS). However, Satb2 occupies more of the intergenic regions at 80% epiboly than at 6 somites and 14 somites. The ratio between intergenic and promoter-bound peaks was similar for 6 somites and 14 somites (Figure 3.7A). Interestingly, we observed strong enrichment for Satb2 on the regulatory regions of target genes from 80% epiboly. Hence, although deregulation of genes related to neurogenesis and neural crest specification in *Satb2* mutants is highly evident at 14 somites, our ChIP-seq analysis suggested that Satb2 potentially book-marks target genes for regulation from as early as 80% epiboly. Examples of the binding profiles are represented for *sox10*, *sox9a*, *zic1* and *zic4*, *lef1* and *wnt11* (Figure 3.7B).

Figure 3.7 Genome-wide occupancy analysis confirms the direct regulation of downstream target genes by Satb2.



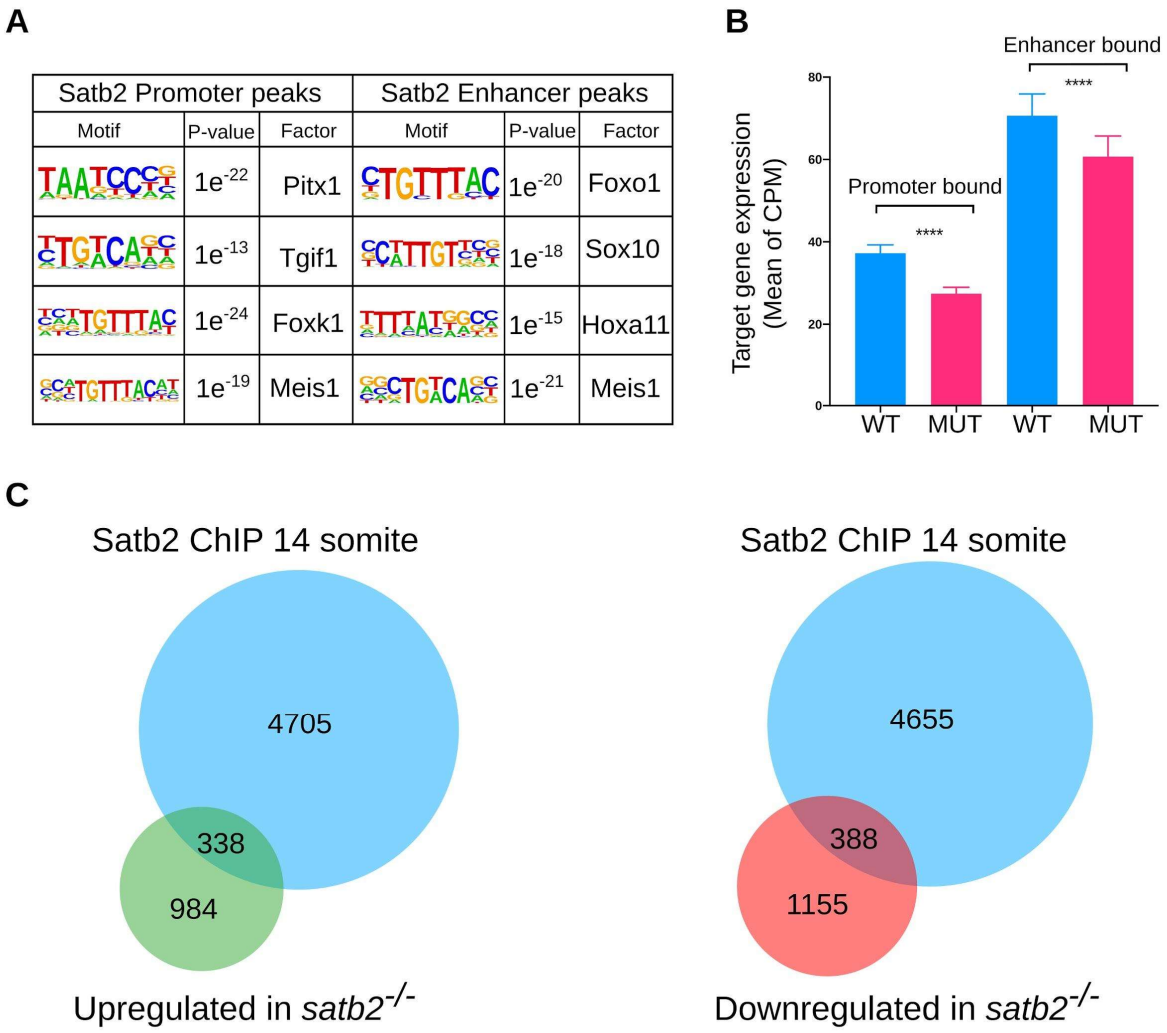
(A) Genome-wide distribution pattern of Satb2 occupancy in the regulatory regions. percentage of peaks overlapping with OCR at 80% epiboly, 6 ss and 14 ss. Numbers in the brackets indicate the total number of Satb2 peaks at the corresponding developmental stage. **(B)** Integrative genomics viewer (IGV) snapshot of Satb2 occupancy on genomic loci of *sox10*, *sox9a, zic1, zic4, lef1* and *wnt11*. Black coloured blocks represent enriched regions determined by the peak calling algorithms. **(C)** GO analysis for biological processes of Satb2 occupied genes at 14 ss. Numbers for each category of doughnut plot represents enrichment ratio, FDR < 0.05.

Further, we subjected Satb2 genomic targets to gene ontology analysis. Consistent with the transcriptome analysis, we observed enrichment for biological processes involved in nucleic acid metabolism, RNA metabolism, nitrogen biosynthesis processes (upregulated in mutants), central nervous system development, brain development and morphogenesis during early embryonic development (all downregulated in mutants) (Figure 3.7C).

To study Satb2 occupancy on putative enhancers of these genes, we intersected the Satb2 regulatory peak set with the previously published H3K4me1 (active enhancers) dataset across developmental stages (GSE32483, GSE74231). Satb2 peaks from the promoter and putative enhancer regions were further subjected to motif analysis. Interestingly, we observed a strong enrichment for factors including Pitx1, Tgif1, Foxk1 and Meis1 at promoters (Figure 3.8A). Mutations in these genes are known to affect craniofacial development in mice and humans (Amendt et al., 1998; Machon et al., 2015; Meeson et al., 2007; Melvin et al., 2013; Szeto et al., 1999; Taniguchi et al., 2012, 2017). In contrast, enhancer regions showed strong enrichment for the TFs Foxo1, Sox10, Hoxa11 which are known to regulate neural crest differentiation and skeletal development programs (Carney et al., 2006; Dutton et al., 2001; Gouti et al., 2011; Minoux and Rijli, 2010; Teixeira et al., 2010). Sox10 has been reported to regulate these processes through enhancer mediated activities (Betancur et al., 2010a, 2011; Prasad et al., 2011). To further understand the correlation between Satb2 binding sites and gene expression, we computed gene expression from wild-type and mutants. Both promoter-bound and enhancer-bound genes were significantly down-regulated in the mutants. However, in wild-type condition, enhancer-bound genes are expressed at a higher level implicating that these regions may potentially act as enhancers and Satb2 may regulate the expression of these genes through enhancer mediated regulation (Figure 3.8B). We next sought to identify genes that are directly

bound by Satb2 and are differentially expressed. We annotated Satb2 bound regions to the nearest genes and analyzed the overlap between the two. Such analysis demonstrated Satb2 occupied both upregulated and downregulated sets of genes, prompting us to hypothesize that Satb2 exhibits bimodal activity through direct chromatin binding required for the fine-tuning of gene regulatory networks during organogenesis. Taken together, the ChIP-seq analysis suggested, Satb2 directly activates neurogenesis programs at the expense of RNA metabolism pathways, which are mainly functional during the initial cleavage stages to support early development.

Figure 3.8 Characterization of functional genomic binding sites of Satb2.

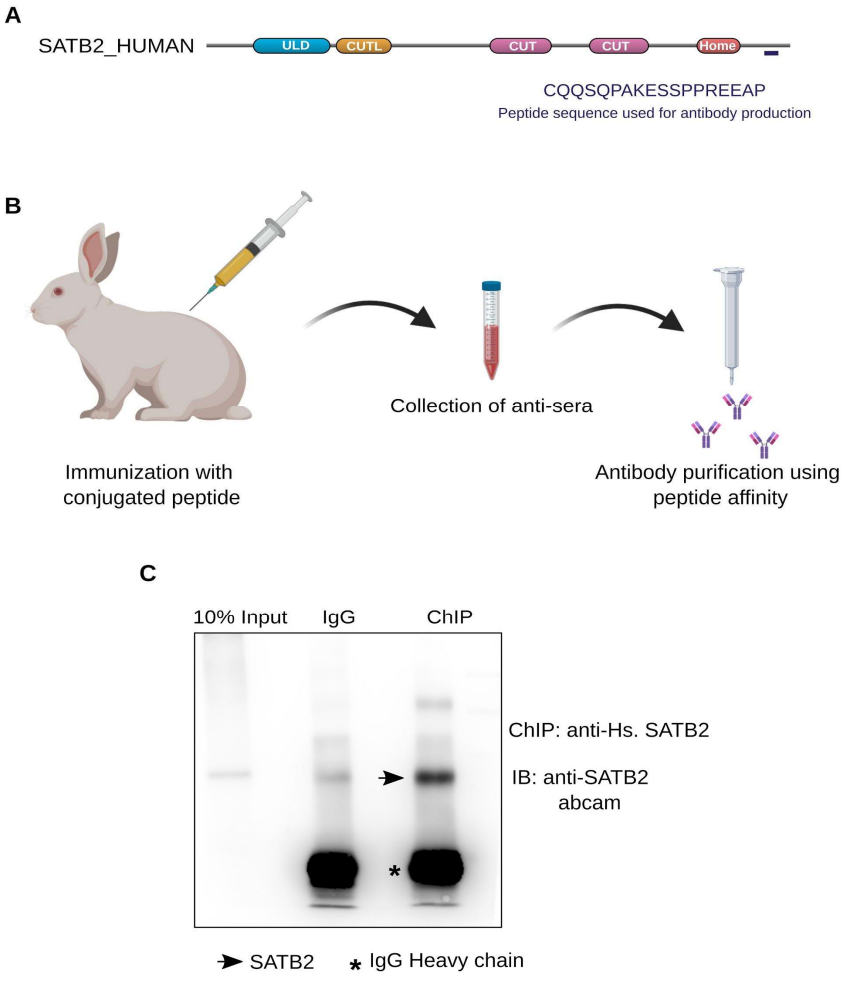


(A) Motif enrichment analysis for Satb2 binding sites at the promoter (+/- 2 Kb) and enhancer regions (overlapping with H3K4me1 binding sites) matched to known motif sets of danRer10 genome assembly. **(B)** Bar plots indicating normalized gene expression (Counts Per Million) of promoter- and enhancer- bound genes in wild type and Satb2 $^{-/-}$ mutants at 14 ss. **(C)** Venn analysis depicting the overlap between Satb2 bound genes and differentially regulated genes at 14 somites in the Satb2$^{-/-}$ mutant.

3.3.6 Direct binding to cis-regulatory elements is a conserved function of Satb2

To test whether the mechanism of gene regulation by zebrafish Satb2 is conserved in mammals, we performed ChIP-seq analysis on developing *Mus musculus* embryos using mammalian specific SATB2 antibody generated in-house (Figure 3.9A and 3.9B). Pull down the efficiency of anti-Human SATB2 antibody was validated using ChIP-western (Figure 3.9C). Immunoblotting was performed using a commercial antibody to confirm specificity.

Figure 3.9 Generation and characterization of anti-human Satb2 antibody.

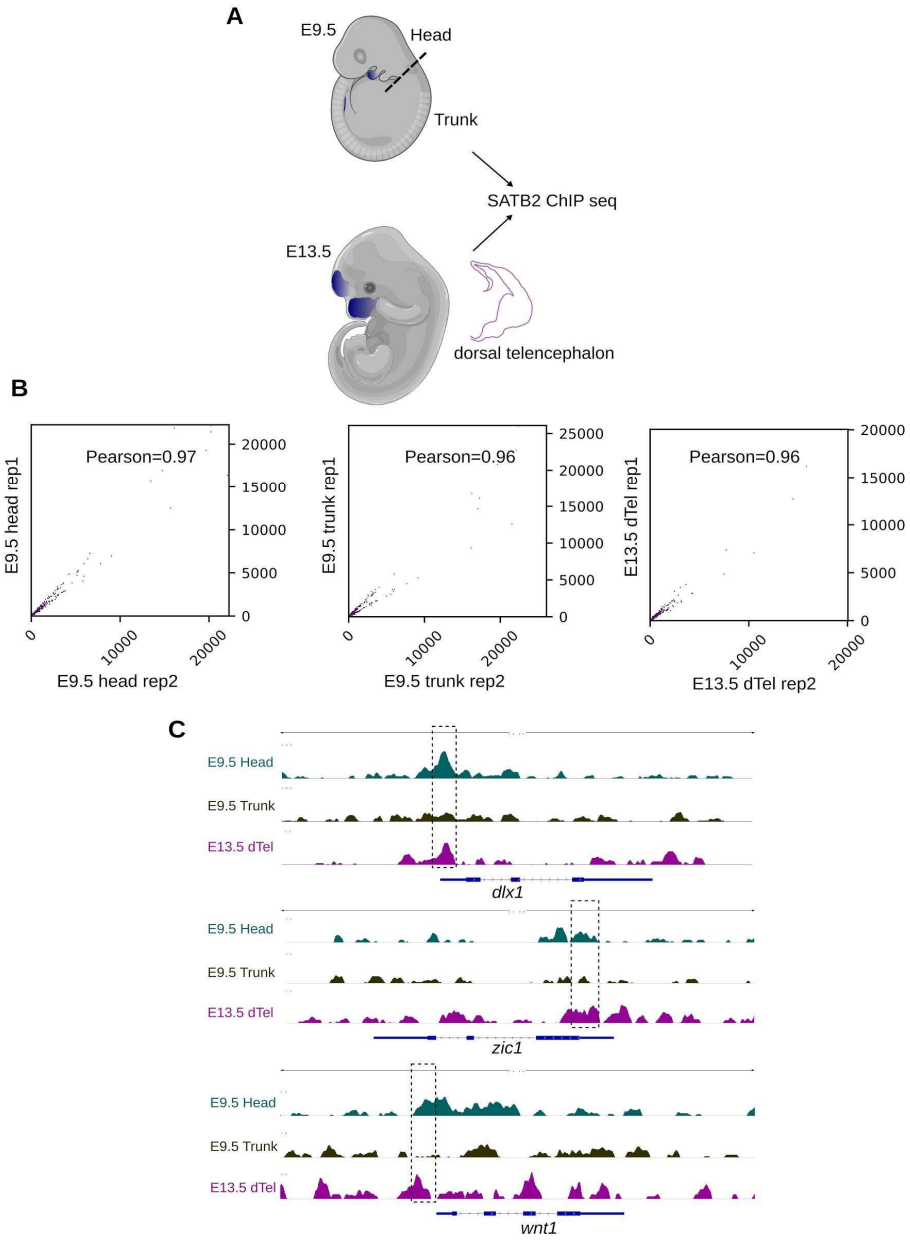


(A) Schematic of human SATB2 domain architecture highlighting the C-terminal region used for peptide synthesis (blue block). Peptide sequence is denoted in N-C terminal orientation. **(B)** Workflow of polyclonal antibody production followed by affinity purification. **(C)** ChIP western

analysis for validation of anti-Hs.-SATB2 antibody using mouse telencephalon tissues. Immunoblot was performed using an anti-SATB2 antibody from Abcam (ab34735).

Next, we chose two stages of embryonic development, E9.5 and E13.5 based on the known expression pattern of Satb2 and stages of NC development. At E9.5, Satb2 is expressed at the rhombomere of the hindgut, branchial arch (source of neural crest cells) and in the hindgut diverticulum. At E13.5, Satb2 is highly expressed in the neocortex, hence often used as a marker for layer 5, and at the sites of bone formation (Dobreva et al., 2006). The use of mouse developing embryos provide us with a unique opportunity to study spatial regulation of Satb2 during NC specification, which was limited with the zebrafish system. Towards this, we dissected the head (positive source for cranial neural crest cells) and trunk tissue (source of neural crest cells contributing to the peripheral nervous system) at E9.5. We isolated the dorsal telencephalon (dTel) at E13.5 to enrich the population expressing Satb2 in the developing cortex (Figure 3.10A). The ChIP-seq analysis in mice indeed showed significant enrichment for Satb2 on the regulatory regions of bonafide NCC markers such as *dlx1*, *zic1* and *wnt1.* However, Satb2 occupancy was not enriched in trunk tissues for these regulatory regions, suggesting gene-specific and spatially restricted regulation by mouse SATB2 (Figure 3.10B and 3.10C). Taken together, our data suggest that functional regulation by Satb2 through direct binding at cis-regulatory elements of target genes is a conserved mechanism across vertebrates.

Figure 3.10 Direct binding to cis-regulatory elements is a conserved function of Satb2.



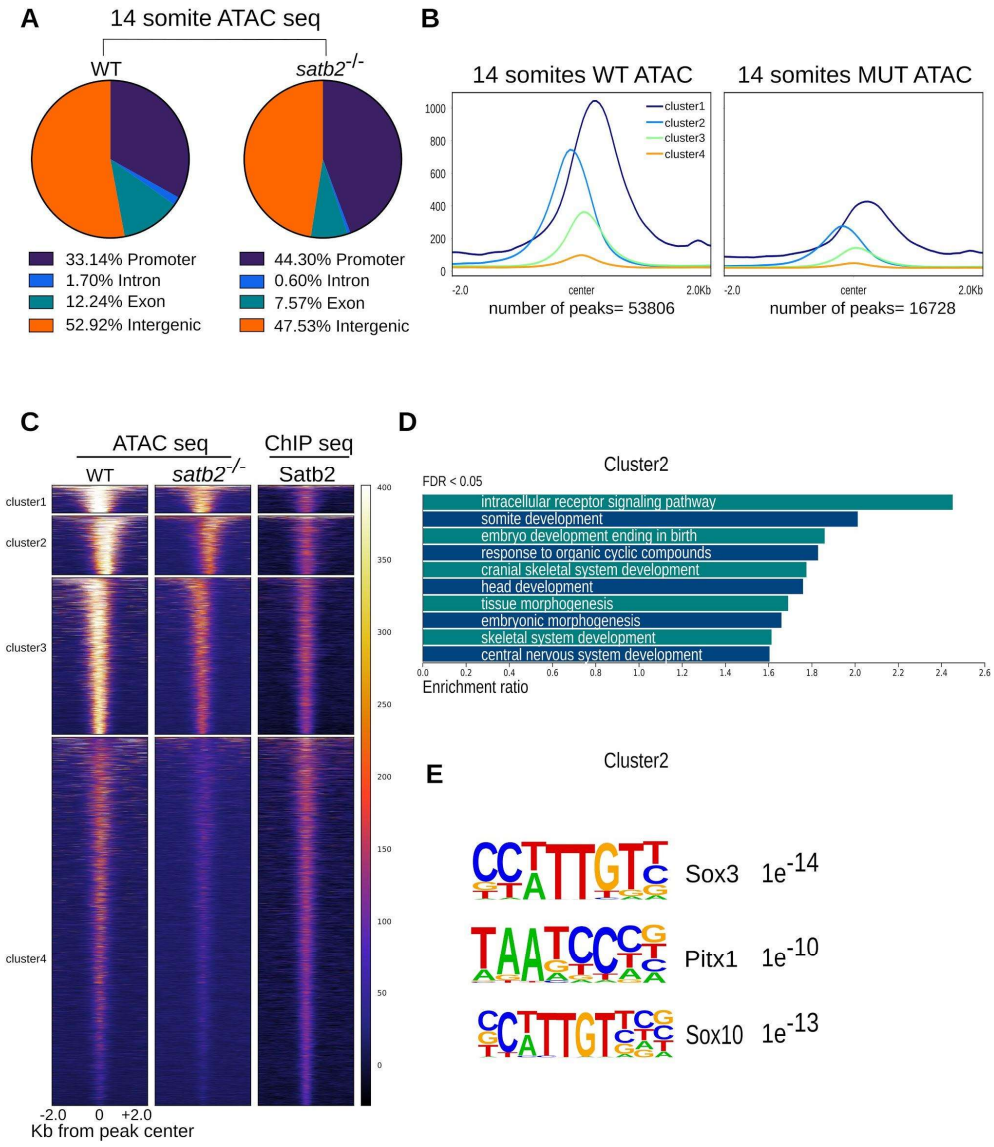
(A) Schematic of dissections performed for the ChIP-seq experiment in mouse embryos at E9.5 (head and Trunk) and E13.5 (dorsal telencephalon). Dotted lines represent the site of the incision. **(B)** Pearson correlation plots highlighting reproducibility between replicates of mousse ChIP-sequencing samples at corresponding stages and tissues. **(C)** IGV snapshot of mouse SATB2 occupancy profile on the genomic loci of *dlx1*, *zic1* and *wnt1.* Dashed line boxes highlight enriched regions in E9.5 head and E13.5 dTel which are absent in the E9.5 trunk region.

3.3.7 Loss of Satb2 results in reduced chromatin accessibility

After establishing the direct regulatory role of Satb2 during early neurogenesis and NCC differentiation, we were interested in understanding whether the observed regulation is through modulating local chromatin states. Towards this, we performed Assay for Transposase-Accessible Chromatin using high-throughput sequencing (ATAC-seq) in wild-type siblings and *satb2*$^{-/-}$ mutants at 14 somites. As discussed previously, we observed maximum change in gene expression through transcriptome analysis of corresponding stages. *satb2*$^{-/-}$ mutants yielded a significantly lower number of open chromatin regions. Interestingly, the percentage of open chromatin sites at intergenic regions dropped significantly as compared to that of the promoter regions (Figure 3.11A). Clustering (K-means=4) of ATAC-seq signals around the +/- 2 Kb centre of Satb2 regulatory peaks showed significant reduction across the genome in SATB2 mutants (Figure 3.11B and Figure 3.11C). The gene ontology analysis revealed cluster 2 to be enriched for biological processes such as somite formation, embryonic morphogenesis, head development and cranial and skeletal system development (Figure 3.11D). Motif enrichment analysis of cluster 2 yielded significant occurrences of known regulators of neurogenesis and NCC specification such as Sox3, Pitx1 and Sox10 (Figure 3.11E).

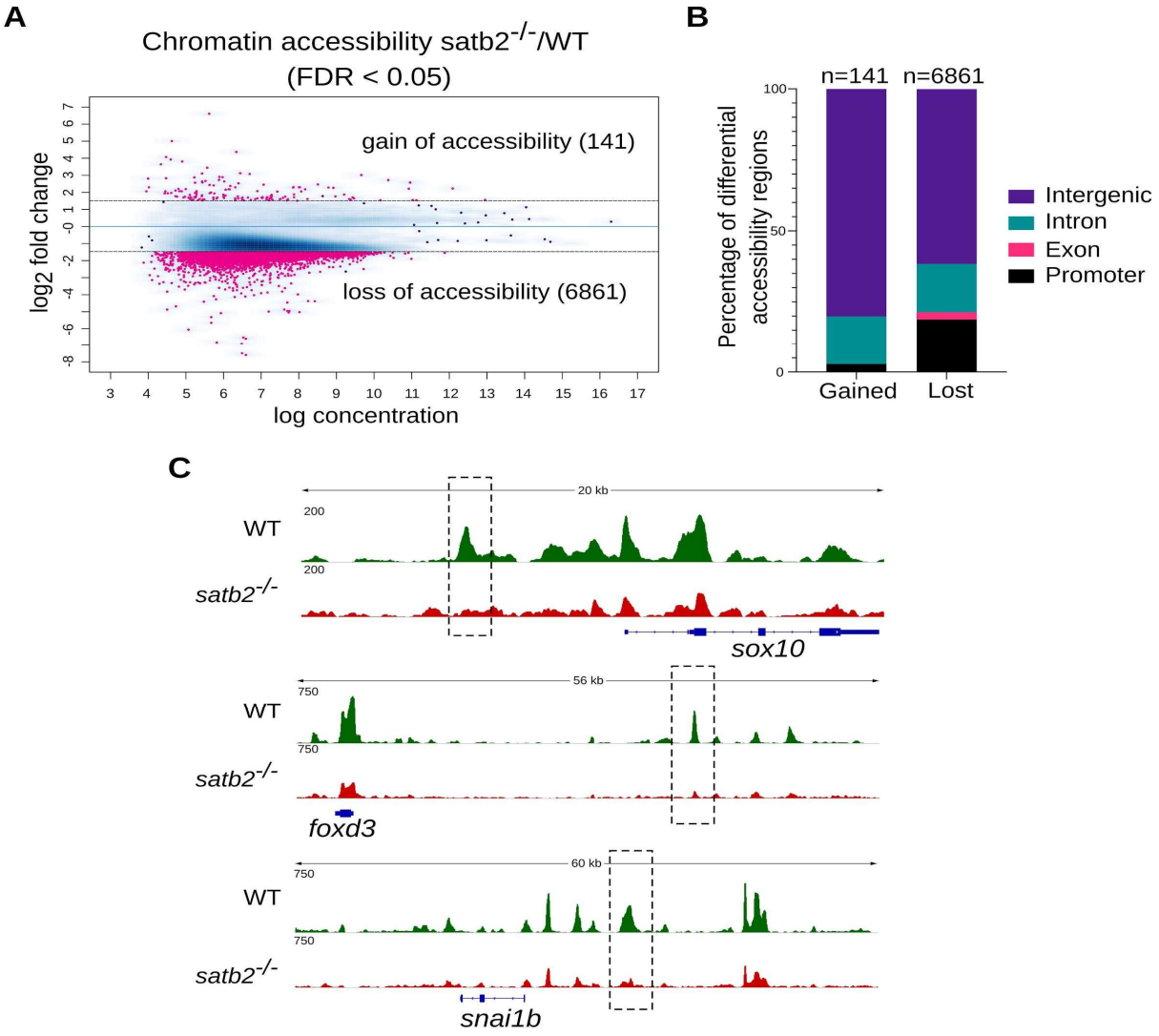
To identify significantly differential chromatin accessibility regions, we subjected the ATAC-seq data to DiffBind analysis. We found only 141 regions with a gain of accessibility and a strikingly high number of regions (6861) with loss of chromatin accessibility in mutants when compared to wild-type siblings (Figure 3.12A). Genomic distribution analysis reconfirmed the previous observation that modulation of chromatin accessibility is primarily within intergenic regions (Figure 3.12B) specifically for known regulators of neural crest cells such as *sox10*, *foxd3* and *snai1b* (Figure 3.12C).

Figure 3.11 Loss of Satb2 results in reduced chromatin accessibility.



(A) Genomic distribution of ATAC-seq peaks in wild type and Satb2$^{-/-}$ mutant at 14 ss as a percentage of the total. (B) Plot profile of chromatin accessibility profile across +/- 2 kb of the centre of ATAC peaks **(C)** Heatmaps for clustering (K-means 4) of chromatin accessibility around +/- 2 Kb of Satb2 ChIP-seq peak regions. Colour Bar represents the degree of chromatin accessibility from low (blue) to high (white). **(D)** GO analysis of cluster 2 highlighting biological processes involved in neural and craniofacial development. **(E)** Enriched TFBS are represented for cluster 2.

Figure 3.12 Differential accessibility analysis suggests Satb2 regulates gene expression through enhancer regions.

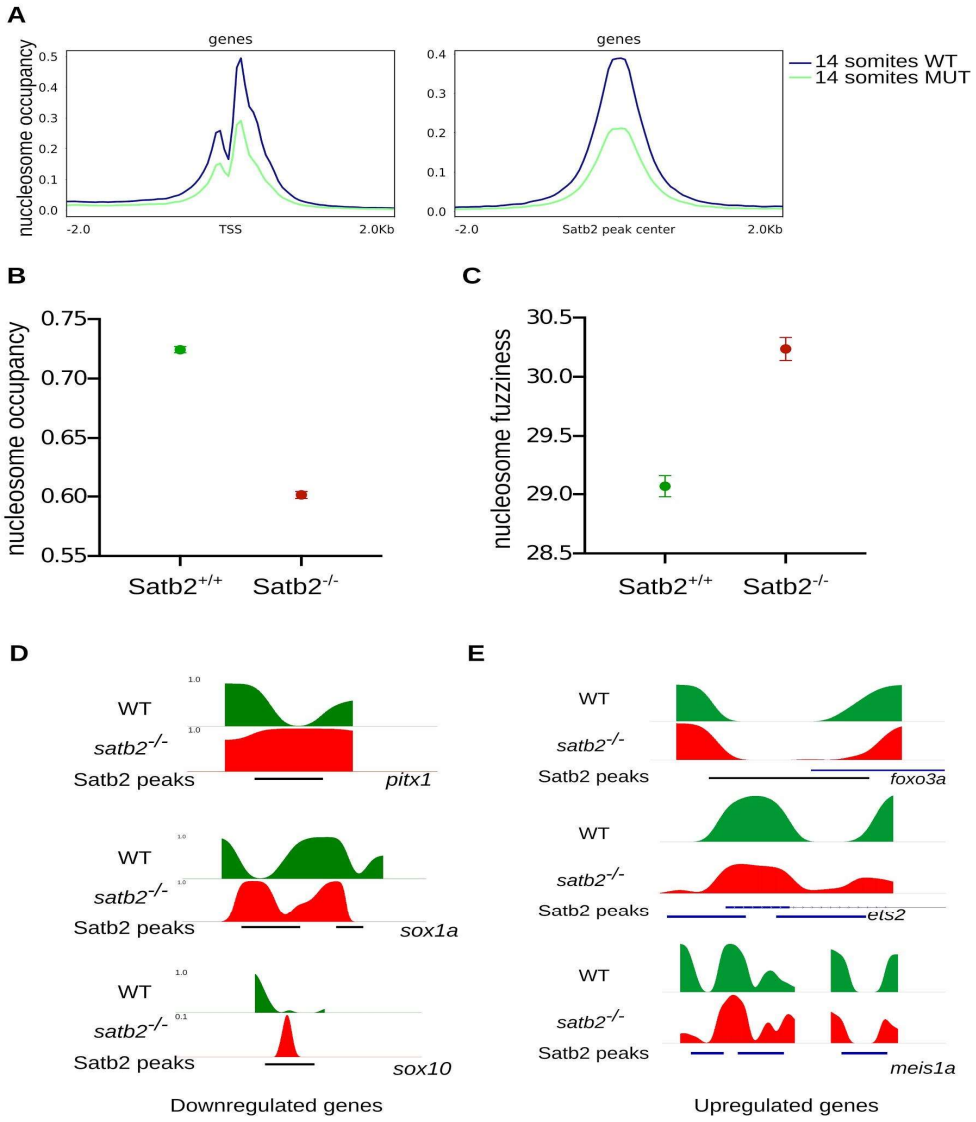


(A) Differential chromatin accessibility between wild type and Satb2$^{-/-}$ mutant embryos represented as binding affinity (FDR < 0.05). Y-axis represents Log2 fold change. Regions with Log2 fold (> +/- 0.58) were considered significant and marked by solid lines. **(B)** Genome-wide distribution of differentially gained or lost regions in Satb2$^{-/-}$ mutant embryos represented as a percentage. **(C)** IGV snapshot of ATAC-seq peaks over the genomic loci for enhancer regions of neural crest markers *sox10*, *foxd3* and *snai1b*. Dashed line boxes highlight regions that are identified as differentially expressed through Diffbind analysis.

3.3.8 Loss of Satb2 results in perturbed nucleosome positioning

Reduced chromatin accessibility is one of the mechanisms for gene regulation. However, genome organizer proteins may also directly affect nucleosome positioning to regulate gene expression (Jiang and Franklin Pugh, 2009). To test whether Satb2 has any impact on nucleosome positioning, we calculated nucleosome occupancy from ATAC-seq data using the nucleoATAC suite (Schep et al., 2015). Interestingly, we observed reduced occupancy signals in mutants compared to wild-type at the TSS as well as at the Satb2 binding regions (Figures 3.13A and 3.13B). This observation is counter-intuitive as current gene regulation models suggest that reduced occupancy leads to gene activation. To dissect this further, we monitored the nucleosome distribution at many of the target gene loci. We observed disruption in nucleosome phasing resulting in fuzzy nucleosome patterns, particularly at the Satb2 binding sites (Figures 3.13C and 3.13D). Interestingly, we did not observe such change in nucleosome positioning at the TSS for genes that are upregulated in $Satb2^{-/-}$ mutants (Figure 3.13E). These findings collectively suggest down-regulation of target genes could be resulting in a disruption in defined nucleosome phasing. In brief, the chromatin accessibility analysis directed us towards multiple mechanisms through which Satb2 actively ensures sustained gene expression to drive neurogenesis and neural crest differentiation programs.

Figure 3.13 Loss of Satb2 affects nucleosome positioning.



(A) K-means clustering of average profiles of chromatin accessibility for wild type and satb2$^{-/-}$ at 14 ss. **(B)** Dot plot depicting a decrease in average nucleosome occupancy in Satb2 mutants as compared to wild type (**** p-value < 0.0001). Error bar indicates +/- SEM. **(C)** Dot plot depicting an increase in average nucleosome fuzziness in Satb2 mutants as compared to wild type (**** p-value < 0.0001). Error bar indicates +/- SEM. **(D)** Nucleosome occupancy profile calculated using ATAC-seq showing changes in nucleosome phasing at Satb2 regulatory sites for downregulated genes *pitx1*, *sox1a* and *sox10*. **(E)** Integrative genome viewer snapshots of nucleosome occupancy tracks on the genomic loci of upregulated genes *foxo3a*, *ets2* and *meis1a* upon loss of Satb2.

3.4 Discussion

Phylogenetic analysis and domain architecture analysis as described previously (Chapter 2, Figure 2.2) highlights the conserved nature of Satb2 across vertebrates and argues for possible functional conservation of this ancient member of the family. Indeed, the zebrafish *satb2* mutants generated in this study recapitulated phenotypes seen in corresponding aberrations in mice and also resembles the human genetic conditions induced by the disruption of *SATB2* locus. Emboldened by this similarity we embarked on a detailed analysis of *satb2* mutant embryos using functional genomics. We simultaneously employed analogous samples using mouse tissues and human cell lines and, thus far, several observations of broad significance have emerged.

To uncover SATB2 function during early embryogenesis, we generated a loss of function mutation in the zebrafish *satb2* locus. The mutation faithfully mimics important features of SATB2-associated syndrome in humans including craniofacial abnormalities which often arise from aberrant neural crest (NC) specification and migration (Cordero et al., 2011; Trainor, 2010). The zebrafish Satb2 mutant generated in this study provides a unique opportunity for characterizing the function of Satb2 during very early specification events of neurogenesis. Indeed, gene expression analysis of these mutants resulted in the identification of a number of dysregulated genes led us to conclude that, Satb2 acts as a positive regulator of genes such as *sox10*, *sox9a*, *mycb*, *zic1*, *pax3a* which are known effector molecules of the neural crest specification process (Carney et al., 2006; Dutton et al., 2001; Gouti et al., 2011; Minoux and Rijli, 2010; Teixeira et al., 2010; Betancur et al., 2010a, 2011; Prasad et al., 2011). Interestingly, Satb2 negatively regulates ribosomal biosynthesis during early neurogenesis. This is indeed an unprecedented instance of Satb2 regulating metabolic pathways and would be of interest to study the further interplay between the two.

Gene expression is often controlled through coordinated interaction between proximal and distal regulatory elements resulting in the permissive or repressive state of chromatin. Chromatin accessibility analysis of Satb2 mutants during the early stages of neurogenesis showed a significant decrease and perturbed nucleosome positioning. Most of the statistically significant changes were seen around distal regulatory regions. These observations suggest that Satb2 modulates epigenetic machinery by bringing effective changes at the distal regulatory regions to

regulate target gene expression. Interestingly, differentially accessible regions showed high enrichment for the binding consensus for proteins such as Sox3, Pitx1 and Sox10. These factors have been shown to mediate long-range interactions during various developmental processes including NCC development (Murko and Bronner, 2017; Sarro et al., 2018). Our study further suggests that Satb2 may mediate long-range interaction by forming a complex with these proteins at distal regulatory regions.

Taken together, through comprehensive transcriptome, genome-wide occupancy and chromatin accessibility analysis we could identify several functional Satb2 targets. The genome-wide occupancy analysis during mouse embryonic stages further strengthens the notion of functional conservation of Satb2 during organogenesis.

3.5 References.

Ahn, H. J., Park, Y., Kim, S., Park, H. C., Seo, S. K., Yeo, S. Y., and Geum, D. (2010). The expression profile and function of Satb2 in zebrafish embryonic development. Mol. Cells. 30(4), 377-382.

Adameyko, I., and Fried, K. (2016). The Nervous System Orchestrates and Integrates Craniofacial Development: A Review. Front. Physiol. *7*, 49.

Alcamo, E.A., Chirivella, L., Dautzenberg, M., Dobreva, G., Fariñas, I., Grosschedl, R., and McConnell, S.K. (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. Neuron *57*, 364–377.

Amendt, B.A., Sutherland, L.B., Semina, E.V., and Russo, A.F. (1998). The molecular basis of Rieger syndrome. Analysis of Pitx2 homeodomain protein activities. J. Biol. Chem. *273*, 20066–20072.

Baranek, C., Dittrich, M., Parthasarathy, S., Bonnon, C.G., Britanova, O., Lanshakov, D., Boukhtouche, F., Sommer, J.E., Colmenares, C., Tarabykin, V., et al. (2012). Protooncogene Ski cooperates with the chromatin-remodeling factor Satb2 in specifying callosal neurons. Proc. Natl. Acad. Sci. U. S. A. *109*, 3546–3551.

Betancur, P., Bronner-Fraser, M., and Sauka-Spengler, T. (2010a). Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. Proc. Natl. Acad. Sci. U. S. A. *107*, 3570–3575.

Betancur, P., Bronner-Fraser, M., and Sauka-Spengler, T. (2010b). Assembling neural crest regulatory circuits into a gene regulatory network. Annu. Rev. Cell Dev. Biol. *26*, 581–603.

Betancur, P., Sauka-Spengler, T., and Bronner, M. (2011). A Sox10 enhancer element common

to the otic placode and neural crest is activated by tissue-specific paralogs. Development *138*, 3689–3698.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonilla-Claudio, M., Wang, J., Bai, Y., Klysik, E., Selever, J., and Martin, J.F. (2012). Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development. Development *139*, 709–719.

Britanova, O., Akopov, S., Lukyanov, S., Gruss, P., and Tarabykin, V. (2005). Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. Eur. J. Neurosci. *21*, 658–668.

Britanova, O., Depew, M.J., Schwark, M., Thomas, B.L., Miletich, I., Sharpe, P., and Tarabykin, V. (2006). Satb2 haploinsufficiency phenocopies 2q32-q33 deletions, whereas loss suggests a fundamental role in the coordination of jaw development. Am. J. Hum. Genet. *79*, 668–678.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., et al. (2008). Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron *57*, 378–392.

Bronner-Fraser, M. (2008). Ancient evolutionary origin of the neural crest gene regulatory network. Dev. Biol. *319*, 470.

Carney, T.J., Dutton, K.A., Greenhill, E., Delfino-Machín, M., Dufourcq, P., Blader, P., and Kelsh, R.N. (2006). A direct role for Sox10 in specification of neural crest-derived sensory neurons. Development *133*, 4619–4630.

Cordero, D.R., Brugmann, S., Chu, Y., Bajpai, R., Jame, M., and Helms, J.A. (2011). Cranial neural crest cells on the move: their roles in craniofacial development. Am. J. Med. Genet. A *155A*, 270–279.

Cornell, R.A., and Eisen, J.S. (2000). Delta signaling mediates segregation of neural crest and spinal sensory neurons from zebrafish lateral neural plate. Development *127*, 2873–2882.

Couly, G., Creuzet, S., Bennaceur, S., Vincent, C., and Le Douarin, N.M. (2002). Interactions between Hox-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebrate head. Development *129*, 1061–1073.

Cubbage, C.C., and Mabee, P.M. (1996). Development of the cranium and paired fins in the zebrafish Danio rerio (Ostariophysi, Cyprinidae). J. Morphol. *229*, 121–160.

Dobreva, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Fariñas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell *125*, 971–986.

Döcker, D., Schubach, M., Menzel, M., Munz, M., Spaich, C., Biskup, S., and Bartholdi, D. (2014). Further delineation of the SATB2 phenotype. Eur. J. Hum. Genet. *22*, 1034–1039.

Dupin, E., Creuzet, S., and Le Douarin, N.M. (2006). The contribution of the neural crest to the vertebrate body. Adv. Exp. Med. Biol. *589*, 96–119.

Dutton, K.A., Pauliny, A., Lopes, S.S., Elworthy, S., Carney, T.J., Rauch, J., Geisler, R., Haffter, P., and Kelsh, R.N. (2001). Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. Development *128*, 4113–4125.

FitzPatrick, D.R., Carr, I.M., McLaren, L., Leek, J.P., Wightman, P., Williamson, K., Gautier, P., McGill, N., Hayward, C., Firth, H., et al. (2003). Identification of SATB2 as the cleft palate gene on 2q32-q33. Hum. Mol. Genet. *12*, 2491–2501.

Frisdal, A., and Trainor, P.A. (2014). Development and evolution of the pharyngeal apparatus. Wiley Interdiscip. Rev. Dev. Biol. *3*, 403–418.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS One *9*, e98186.

Gans, C., and Northcutt, R.G. (1983). Neural crest and the origin of vertebrates: a new head. Science *220*, 268–273.

Garnett, A.T., Square, T.A., and Medeiros, D.M. (2012). BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. Development *139*, 4220–4231.

Gitton, Y., Heude, E., Vieux-Rochas, M., Benouaiche, L., Fontaine, A., Sato, T., Kurihara, Y., Kurihara, H., Couly, G., and Levi, G. (2010). Evolving maps in craniofacial development. Semin. Cell Dev. Biol. *21*, 301–308.

Gong, Y., Qian, Y., Yang, F., Wang, H., and Yu, Y. (2014). Lentiviral-mediated expression of SATB2 promotes osteogenic differentiation of bone marrow stromal cells in vitro and in vivo. Eur. J. Oral Sci. *122*, 190–197.

Gouti, M., Briscoe, J., and Gavalas, A. (2011). AnteriorHoxGenes Interact with Components of the Neural Crest Specification Network to Induce Neural Crest Fates. Stem Cells *29*, 858–870.

Gyorgy, A.B., Szemes, M., de Juan Romero, C., Tarabykin, V., and Agoston, D.V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. Eur. J. Neurosci. *27*, 865–873.

Hu, N., Strobl-Mazzulla, P.H., Simoes-Costa, M., Sánchez-Vásquez, E., and Bronner, M.E. (2014). DNA methyltransferase 3B regulates duration of neural crest production via repression of Sox10. Proc. Natl. Acad. Sci. U. S. A. *111*, 17911–17916.

Jiang, C., and Franklin Pugh, B. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature Reviews Genetics *10*, 161–172.

Kague, E., Gallagher, M., Burke, S., Parsons, M., Franz-Odendaal, T., and Fisher, S. (2012). Skeletogenic fate of zebrafish cranial and trunk neural crest. PLoS One *7*, e47394.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.

Knight, R.D., and Schilling, T.F. (2006) Cranial Neural Crest and Development of the Head Skeleton. ADV. EXP. MED. BIOL. 589,120-133.

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

Le Douarin, N.M., and Dupin, E. (2003). Multipotentiality of the neural crest. Curr. Opin. Genet. Dev. *13*, 529–536.

Le Douarin, N., LeDouarin, N.M., and Kalcheim, C. (1999). The Neural Crest (Cambridge University Press).

Leichsenring, M., Maes, J., Mössner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science *341*, 1005–1009.

Leone, D.P., Srinivasan, K., Chen, B., Alcamo, E., and McConnell, S.K. (2008). The determination of projection neuron identity in the developing cerebral cortex. Curr. Opin. Neurobiol. *18*, 28–35.

Leone, D.P., Heavner, W.E., Ferenczi, E.A., Dobreva, G., Huguenard, J.R., Grosschedl, R., and McConnell, S.K. (2015). Satb2 Regulates the Differentiation of Both Callosal and Subcerebral Projection Neurons in the Developing Cerebral Cortex. Cereb. Cortex *25*, 3406–3419.

Leoyklang, P., Suphapeetiporn, K., Siriwan, P., Desudchit, T., Chaowanapanja, P., Gahl, W.A., and Shotelersuk, V. (2007). Heterozygous nonsense mutation SATB2 associated with cleft palate, osteoporosis, and cognitive defects. Hum. Mutat. *28*, 732–738.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, M., Zhao, C., Wang, Y., Zhao, Z., and Meng, A. (2002). Zebrafish sox9b is an early neural crest marker. Dev. Genes Evol. *212*, 203–206.

Liem, K.F., Jr, Tremml, G., Roelink, H., and Jessell, T.M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell *82*, 969–979.

Lister, J.A., Cooper, C., Nguyen, K., Modrell, M., Grant, K., and Raible, D.W. (2006). Zebrafish Foxd3 is required for development of a subset of neural crest derivatives. Dev. Biol. *290*, 92–104.

Lowery, L.A., and Sive, H. (2004). Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation. Mech. Dev. *121*, 1189–1197.

Lukoseviciute, M., Gavriouchkina, D., Williams, R.M., Hochgreb-Hagele, T., Senanayake, U., Chong-Morrison, V., Thongjuea, S., Repapi, E., Mead, A., and Sauka-Spengler, T. (2018). From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. Dev. Cell *47*, 608–628.e6.

Machon, O., Masek, J., Machonova, O., Krauss, S., and Kozmik, Z. (2015). Meis2 is essential for cranial and cardiac neural crest development. BMC Dev. Biol. *15*, 40.

Marchant, L., Linker, C., Ruiz, P., Guerrero, N., and Mayor, R. (1998). The inductive properties of mesoderm suggest that the neural crest cells are specified by a BMP gradient. Dev. Biol. *198*, 319–329.

Marcucio, R.S., Young, N.M., Hu, D., and Hallgrimsson, B. (2011). Mechanisms that underlie co-variation of the brain and face. Genesis *49*, 177–189.

Matsukawa, S., Miwata, K., Asashima, M., and Michiue, T. (2015). The requirement of histone modification by PRDM12 and Kdm4a for the development of pre-placodal ectoderm and neural crest in Xenopus. Dev. Biol. *399*, 164–176.

Mayor, R., Guerrero, N., and Martínez, C. (1997). Role of FGF andNogginin Neural Crest Induction. Developmental Biology *189*, 1–12.

Meeson, A.P., Shi, X., Alexander, M.S., Williams, R.S., Allen, R.E., Jiang, N., Adham, I.M., Goetsch, S.C., Hammer, R.E., and Garry, D.J. (2007). Sox15 and Fhl3 transcriptionally coactivate Foxk1 and regulate myogenic progenitor cells. EMBO J. *26*, 1902–1912.

Melvin, V.S., Feng, W., Hernandez-Lagunas, L., Artinger, K.B., and Williams, T. (2013). A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis. Dev. Dyn. *242*, 817–831.

Minoux, M., and Rijli, F.M. (2010). Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. Development *137*, 2605–2621.

Mitsiadis, T.A., Chéraud, Y., Sharpe, P., and Fontaine-Pérus, J. (2003). Development of teeth in chick embryos after mouse neural crest transplantations. Proc. Natl. Acad. Sci. U. S. A. *100*, 6541–6545.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. *42*, W401–W407.

Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods *12*, 982–988.

Mork, L., and Crump, G. (2015). Zebrafish Craniofacial Development: A Window into Early Patterning. Curr. Top. Dev. Biol. *115*, 235–269.

Murko, C., and Bronner, M.E. (2017). Tissue specific regulation of the chick Sox10E1 enhancer by different Sox family members. Dev. Biol. *422*, 47–57.

Piotrowski, T., Schilling, T.F., Brand, M., Jiang, Y.J., Heisenberg, C.P., Beuchle, D., Grandel, H., van Eeden, F.J., Furutani-Seiki, M., Granato, M., et al. (1996). Jaw and branchial arch mutants in zebrafish II: anterior arches and cartilage differentiation. Development *123*, 345–356.

Prasad, M.K., Reed, X., Gorkin, D.U., Cronin, J.C., McAdow, A.R., Chain, K., Hodonsky, C.J.,

Jones, E.A., Svaren, J., Antonellis, A., et al. (2011). SOX10 directly modulates ERBB3 transcription via an intronic neural crest enhancer. BMC Dev. Biol. *11*, 40.

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. Cell Stem Cell *11*, 633–648.

Rainger, J.K., Bhatia, S., Bengani, H., Gautier, P., Rainger, J., Pearson, M., Ansari, M., Crow, J., Mehendale, F., Palinkasova, B., et al. (2014). Disruption of SATB2 or its long-range cis-regulation by SOX9 causes a syndromic form of Pierre Robin sequence. Hum. Mol. Genet. *23*, 2569–2579.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Raterman, S.T., Metz, J.R., Wagener, F.A.D.T.G., and Von den Hoff, J.W. (2020). Zebrafish Models of Craniofacial Malformations: Interactions of Environmental Factors. Front. Cell Dev. Biol. *8*, 600926.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). g: Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. *35*, W193–W200.

Rinon, A., Lazar, S., Marshall, H., Büchmann-Møller, S., Neufeld, A., Elhanany-Tamir, H., Taketo, M.M., Sommer, L., Krumlauf, R., and Tzahor, E. (2007). Cranial neural crest cells regulate head muscle patterning and differentiation during vertebrate embryogenesis. Development *134*, 3065–3075.

Rocha, M., Singh, N., Ahsan, K., Beiriger, A., and Prince, V.E. (2020). Neural crest development: insights from the zebrafish. Dev. Dyn. *249*, 88–111.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y.R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature *524*, 230–233.

Sarro, R., Kocher, A.A., Emera, D., Uebbing, S., Dutrow, E.V., Weatherbee, S.D., Nottoli, T., and Noonan, J.P. (2018). Disrupting the three-dimensional regulatory topology of the locus results in overtly normal development. Development *145*.

Sato, T. (2005). Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm. Development *132*, 2355–2363.

Sauka-Spengler, T., Meulemans, D., Jones, M., and Bronner-Fraser, M. (2007). Ancient evolutionary origin of the neural crest gene regulatory network. Dev. Cell *13*, 405–420.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev. *23*, 2625–2638.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. *25*, 1757–1770.

Schilling, T.F., Piotrowski, T., Grandel, H., Brand, M., Heisenberg, C.P., Jiang, Y.J., Beuchle, D., Hammerschmidt, M., Kane, D.A., Mullins, M.C., et al. (1996). Jaw and branchial arch mutants in zebrafish I: branchial arches. Development *123*, 329–344.

Sheehan-Rooney, K., Pálinkášová, B., Eberhart, J.K., and Dixon, M.J. (2010). A cross-species analysis of Satb2 expression suggests deep conservation across vertebrate lineages. Dev. Dyn. *239*, 3481–3491.

Sheehan-Rooney, K., Swartz, M.E., Lovely, C.B., Dixon, M.J., and Eberhart, J.K. (2013). Bmp and Shh signaling mediate the expression of satb2 in the pharyngeal arches. PLoS One *8*, e59533.

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

Steventon, B., and Mayor, R. (2009). 04-P013 A temporal modulation of FGF, Wnt, BMP and Retinoic Acid signalling is required for neural crest induction. Mech. Dev. *126*, S110.

Stewart, R.A., Arduini, B.L., Berghmans, S., George, R.E., Kanki, J.P., Henion, P.D., and Look, A.T. (2006). Zebrafish foxd3 is selectively required for neural crest specification, migration and survival. Dev. Biol. *292*, 174–188.

Strobl-Mazzulla, P.H., and Bronner-Fraser, M. (2011). Snail2–PHD12 interaction recruits an epigenetic repressive complex that mediates neural crest epithelial–mesenchymal transition. Developmental Biology *356*, 260.

Strobl-Mazzulla, P.H., and Bronner, M.E. (2014). Epigenetic Regulation of Neural Crest Cells. Neural Crest Cells 89–100.

Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisúa-Belmonte, J.C., and Rosenfeld, M.G. (1999). Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. Genes Dev. *13*, 484–494.

Taniguchi, K., Anderson, A.E., Sutherland, A.E., and Wotton, D. (2012). Loss of Tgif function causes holoprosencephaly by disrupting the SHH signaling pathway. PLoS Genet. *8*, e1002524.

Taniguchi, K., Anderson, A.E., Melhuish, T.A., Carlton, A.L., Manukyan, A., Sutherland, A.E., and Wotton, D. (2017). Genetic and Molecular Analyses indicate independent effects of TGIFs on Nodal and Gli3 in neural tube patterning. Eur. J. Hum. Genet. *25*, 208–215.

Teixeira, C.C., Liu, Y., Thant, L.M., Pang, J., Palmer, G., and Alikhani, M. (2010). Foxo1, a novel regulator of osteoblast differentiation and skeletogenesis. J. Biol. Chem. *285*, 31055–31065.

Theveneau, E., and Mayor, R. (2012). Neural crest migration: interplay between chemorepellents, chemoattractants, contact inhibition, epithelial-mesenchymal transition, and

collective cell migration. Wiley Interdiscip. Rev. Dev. Biol. *1*, 435–445.

Thisse, C., Thisse, B., and Postlethwait, J.H. (1995). Expression ofsnail2,a Second Member of the Zebrafish Snail Family, in Cephalic Mesendoderm and Presumptive Neural Crest of Wild-Type andspadetailMutant Embryos. Dev. Biol. *172*, 86–99.

Thomason, A., Pankey, E., Nutt, B., Caffrey, A.R., and Zarate, Y.A. (2019). Speech, language, and feeding phenotypes of SATB2 associated syndrome. Clin. Genet. *96*, 485–492.

Tien, C.-L., Jones, A., Wang, H., Gerigk, M., Nozell, S., and Chang, C. (2015). Snail2/Slug cooperates with Polycomb repressive complex 2 (PRC2) to regulate neural crest development. Development *142*, 722–731.

Trainor, P.A. (2005). Specification of neural crest cell formation and migration in mouse embryos. Semin. Cell Dev. Biol. *16*, 683–693.

Trainor, P.A. (2010). Craniofacial birth defects: The role of neural crest cells in the etiology and pathogenesis of Treacher Collins syndrome and the potential for prevention. Am. J. Med. Genet. A *152A*, 2984–2994.

Trainor, P.A. (2013). Molecular Blueprint for Craniofacial Morphogenesis and Development. Stem Cells in Craniofacial Development and Regeneration 1–29.

Trainor, P., and Krumlauf, R. (2000). Plasticity in mouse neural crest cells reveals a new patterning role for cranial mesoderm. Nat. Cell Biol. *2*, 96–102.

Van Ho, A.T., Hayashi, S., Bröhl, D., Auradé, F., Rattenbach, R., and Relaix, F. (2011). Neural crest cell lineage restricts skeletal muscle progenitor cell differentiation through Neuregulin1-ErbB3 signaling. Dev. Cell *21*, 273–287.

Van Otterloo, E., Li, W., Garnett, A., Cattell, M., Medeiros, D.M., and Cornell, R.A. (2012). Novel Tfap2-mediated control of soxE expression facilitated the evolutionary emergence of the neural crest. Development *139*, 720–730.

Villanueva, S., Glavic, A., Ruiz, P., and Mayor, R. (2002). Posteriorization by FGF, Wnt, and retinoic acid is required for neural crest induction. Dev. Biol. *241*, 289–301.

Walker, M., and Kimmel, C. (2007). A two-color acid-free cartilage and bone stain for zebrafish larvae. Biotech. Histochem. *82*, 23–28.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene.

Westerfield, M. (2007) The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 5th Edition. University of Oregon Press, Eugene.

Yan, Y.-L., Miller, C.T., Nissen, R.M., Singer, A., Liu, D., Kirn, A., Draper, B., Willoughby, J., Morcos, P.A., Amsterdam, A., et al. (2002). A zebrafish sox9 gene required for cartilage morphogenesis. Development *129*, 5065–5079.

Yu, J.-K.S. (2010). The evolutionary origin of the vertebrate neural crest and its developmental gene regulatory network – insights from amphioxus. Zoology *113*, 1–9.

Zarate, Y.A., and Fish, J.L. (2017). SATB2 -associated syndrome: Mechanisms, phenotype, and practical recommendations. Am. J. Med. Genet. A. *173*, 327–337.

Zarate, Y.A., Kalsner, L., Basinger, A., Jones, J.R., Li, C., Szybowska, M., Xu, Z.L., Vergano, S., Caffrey, A.R., Gonzalez, C.V., et al. (2017). Genotype and phenotype in 12 additional individuals with SATB2 -associated syndrome. Clin. Genet. *92*, 423–429.

Chapter 4

Delineating the molecular functions of maternal Satb2 during the mid-blastula transition

4.1 Introduction

Zygotic genome activation (ZGA) initiates regionalized transcription responsible for the acquisition of distinct cellular identities in a developing embryo. Maternally deposited factors initially support the development of an organism and subsequently, the embryonic genome takes the responsibility to synthesize gene products required for pluripotency (Lee et al., 2014). Precise execution of ZGA is dependent upon dynamic chromatin architecture sculpted by conserved DNA-binding proteins. Moreover, ZGA is the culmination of transcriptional activation and transcriptional repression activities required for the fine-tuning of gene-regulatory networks.

In Chapter 2 and 3 we learned how the chromatin organizer Satb2 regulates early fate-specification during gastrulation as well as early events of neurogenesis that are responsible for skeletal morphogenesis. An earlier study has suggested that Satb2 is maternally contributed and widely expressed during the early cleavage stage of zebrafish embryogenesis (Ahn et al., 2010). The disparity between observed phenotypes using knockdown reagents and knock-out models of Satb2 raised several questions about the function of maternally deposited Satb2. Here, we, therefore, aimed to delineate the role of maternally deposited Satb2 using zygotic genome activation as a paradigm.

4.1.1 Zygotic genome activation (ZGA)

During cleavage stages of embryogenesis, maternally deposited factors contribute largely to the initial development of an organism. A fertilized embryo undergoes synchronous cell division cycles before undergoing coordinated morphogenetic movements during gastrulation to give rise to distinct germ layers (Foe and Alberts, 1983; Kane and Kimmel, 1993). This process is conserved across various phyla, however, the timing of onset of cell fate specification activity varies between species. For example, in zebrafish, the zygotic genome is activated within 3 hrs of initial development whereas the mammalian embryos take up to days to undergo ZGA/EGA (Oron and Ivanova, 2012). This discrepancy is potentially due to the time taken to confer transcriptional competence for successful gene activation using embryonic gene regulation systems instead of

maternally deposited resources. The maternal pool of activators and repressors is actively cleared from the embryonic system using various mechanisms. In *Drosophila melanogaster,* this is achieved by the active function of transcription factors Clamp and Zelda (Duan et al., 2020; Liang et al., 2008; Schulz et al., 2015). However, in zebrafish, microRNA miR430 performs this specialized function (Bazzini et al., 2012; Giraldez et al., 2006). Transcriptional inhibition using chemical inhibitors during stages of ZGA results in developmental defects in both Xenopus and zebrafish, highlighting the importance of active transcription for successful gastrulation and setting up new developmental programs (Kane et al., 1996; Newport and Kirschner, 1982).

Two classical models have been described for the successful onset of ZGA. First, discussing the active involvement of the cell cycle. A high nuclear/cytoplasmic (N/C) ratio during early mitotic rounds in cleavage stages is responsible for high transcriptional turnover (Newport and Kirschner, 1982). Second, an increase in translation potential and activity of maternally deposited factors are thought to be responsible for active transcription, independent of a maternal clock (Howe et al., 1995; Lee et al., 2013). Lately, this second model has gained popularity due to the characterization of maternally deposited factors using high throughput technologies such as ribosome-profiling and ChIPseq (Lee et al., 2013). However, these two models are not mutually exclusive and studies, particularly in fruit flies, suggest they are highly interlinked and work synergistically to ensure the successful onset of ZGA (Lu et al., 2009). Nonetheless, a detailed characterization of this process (ZGA) in vertebrate systems is warranted.

4.1.2 Role of chromatin remodelling during ZGA

Dynamic changes in chromatin architecture at regulatory regions such as promoters and enhancers are an important determinant of eukaryotic transcription regulation. The interplay between transcriptional activation and repression through active chromatin remodelling in the early embryo likely regulates the timing of ZGA (Chan et al., 2019; Pálfy et al., 2020; Prioleau et al., 1994). Initial studies by injecting plasmids into the early stages of embryos resulted in active transcription and production of mRNAs. This suggested that embryos are already equipped for active transcription before the major

wave of ZGA (Newport and Kirschner, 1982; Wiekowski et al., 1993). Such changes in chromatin architecture are mediated through DNA methylation and the local epigenetic landscape by acquiring post-translational modifications at histone tails.

4.1.3 Role of DNA methylation during ZGA

DNA methylation is one of the most well-characterized epigenetic modes of gene silencing. In zebrafish, the proteins associated with DNA methylation systems play a crucial role in chromatin regulation including MBD, Tets, and DNMTs (Jessop et al., 2018). However, a study by McGowan and Martin shows the absence of paternal imprinting due to a lack of DNMT3 ortholog in zebrafish (McGowan and Martin, 1997). Interestingly, zebrafish display 2-3% higher CpG methylation than other mammals (Potok et al., 2013). This study also revealed that the zebrafish genome undergoes a reduction in methylation during early cleavage cycles and gets re-methylated to display higher levels at the sphere stage. This also serves the purpose of matching methylation levels of the maternal genome to that of the sperm (Potok et al., 2013). This is in contradiction to the observations made in the mouse system where methylation reprogramming takes place after ZGA in short intervals (Borgel et al., 2010; Messerschmidt et al., 2014). These studies further showed that genes involved in transcription and germline specification such as *hox*, *ta*, *piwil1* and *vasa* are hypermethylated. In contrast, the genes involved in terminal muscle differentiation such as *myod1*, *pbx4* and *pbx2* are hypomethylated, suggesting that differential methylation status is required to exit pluripotency and initiate cell fate specification. Wang et al., performed loss of function of *dnmt1* which resulted in hypomethylation of lefty2, leading to defects in left-right determination in zebrafish embryos (Wang et al., 2017). Lefty is a known molecular inhibitor of Nodal signalling which is important for L-R patterning (Chen and Shen, 2004). Another mode of DNA methylation- 6mA, which has been extensively studied in prokaryotes, is abundant during the early cleavage cycles of zebrafish embryogenesis. However, DNA methylation- 6mA reduces and localises on repetitive genomic regions as embryos develop further (Liu et al., 2016). This makes further studies pertaining to DNA methylation status interesting upon loss of function of key cell fate determinant signalling such as Nodal, BMP and FGF during zebrafish development (Schier and Talbot, 2005; Zinski et al., 2018). To facilitate

the scoring of DNA methylation, Zhang et al. have developed elegant MBD gene reporter lines which can be used to study this Spatio-temporal dynamics of 5mC (Zhang et al., 2017).

4.1.4 Role of histones and histone modifications during ZGA

During the early stages of zebrafish embryogenesis, gamete-specific histone variants are replaced by their somatic versions. This process of histone exchange is believed to be essential for unpacking the genome before the onset of ZGA (Joseph et al., 2017). As embryos develop further, expression of macroH2A diminishes and the H2A.Z variant is highly translated. Joseph et al. further demonstrated how relative levels of core histones and transcription factors are important for active transcription of mesendoderm determinant *mxtx2* and subsequent gene regulatory cascade (Joseph et al., 2017). Mxtx2 is known to induce mesendoderm by elevating levels of nodal signalling in YSL (Hong et al., 2011). Interestingly, reduction in the concentration of non-DNA bound core histones leads to re-initiation of transcription by allowing accessibility of DNA to transcription factors. On the contrary, overexpression of core histones delayed gastrulation (Joseph et al., 2017).

Another level of regulation is through post-translational modification of histones. Developmentally significant regions often exhibit dynamic changes in histone modification status. H3K4me3 is often associated with active transcription whereas H3K9me3 is associated with inactive genes. H3K27me3 is an interesting modification as it is often associated with bivalent or transcriptionally poised regions. It is observed that during mid blastula transition (MBT), H3K4me3 is largely associated with constitutively expressed genes whereas H3K27me3 occupancy is observed at the developmentally regulated genes (Andersen et al., 2013). However post MBT, the levels of H3K4me3 significantly increases on developmental genes such as *sox19a* and *foxd5* that are involved in ectoderm specification resulting in active transcription of these genes (Leichsenring et al., 2013). Taken together, these observations suggest that epigenomic complexity increases during embryonic development, probably due to the initiation of various processes after the blastula stage to achieve differential cell segregation and lineage specification.

Recently, Chan et al. suggested that transcriptional competence is conferred by H3K27Ac as early as the 64 cell stage (first major wave) in developing embryos (Chan et al., 2019). Collectively, these studies highlight the importance of dynamic chromatin accessibility during ZGA. Moreover, as discussed in Chapter 2, long-range interactions mediating the effect of local epigenomic changes at the distal regions play a crucial role during ZGA.

4.1.5 Role of pioneer factors in ZGA

Individual transcription factors are essential for setting up ZGA in a species-dependent manner. For example, in fruit flies factors including the GAGA factor (GAF), CLAMP and Zelda have been shown to regulate ZGA (Duan et al., 2020; Gaskill et al., 2020; Schulz et al., 2015; Sun et al., 2015). Interestingly, Zelda is absent in vertebrates and its functional homologs such as Pou5f1, Sox2 and Nanog are highly translated and control the expression of the signalling molecules involved in the process of gastrulation (Lee et al., 2013). Moreover, OCT4 which is the mammalian homolog of Pou5f3 regulates ZGA in humans but not in mice (Gao et al., 03 22, 2018), whereas proteins such as DUX4 are important for successful ZGA in both humans and mice (De Iaco et al., 2017; Whiddon et al., 2017). The chromatin landscape undergoes massive remodelling during ZGA (Eckersley-Maslin et al., 2018) and importantly, these factors actively participate in these processes (Gao et al., 03 22, 2018; Li et al., 2018).

Three pioneer factors namely Pou5f3, Nanog and Sox19b have been studied extensively in the context of ZGA. These factors have also been shown to negatively regulate the expression of miR430, which targets maternal mRNAs to the degradation machinery (Bazzini et al., 2012). Moreover, the balance between Pou5f3 and Sox19b is an important regulator of active gene transcription during ZGA (Veil et al., 2019). Recent studies also implicated their role in establishing the permissive chromatin landscape necessary for ZGA through acetylation of histones at lysine 27 of H3 and lysine 8 of H4 (Miao et al., 2020; Pálfy et al., 2020). Collectively, these studies could identify a handful of factors that potentially regulate ZGA but highlight the importance of the identification and characterization of more such regulators of ZGA and subsequent regionalised expression.

4.2 Methods

4.2.1 Experimental models

Zebrafish (*Danio rerio*) were maintained as described previously (Westerfield, 2000). All the experimental procedures were carried out following the guidelines from the institutional animal ethics committees at IISER Pune and IST Austria. Embryos were raised at 23-31^{o}C in E3 medium and staged as described previously (Kimmel et al., 1995).

4.2.2 mRNA and morpholino injections

In-vitro mRNA transcription was performed using SP6 mMessage mMachine Kit as per the instructions from the manufacturer (Ambion). Synthesized mRNA was treated with Turbo-DNAse and precipitated using LiCl overnight at -20^{o}C. Synthesized mRNA was checked for quality and quantity by agarose gel electrophoresis and Nanodrop2000 respectively. Small single-use aliquots of mRNAs were kept frozen at -80^{o}C till further use. Glass capillaries (WPI) were pulled using a needle puller (P-97, Sutter Instruments) and mounted on a microinjection system (PV820, World Precision Instruments).

For ubiquitous overexpression and morpholino-mediated knockdown studies, embryos were arranged in agarose moulds and injections were performed at 1 cell stage as described previously (Westerfield, 2007). For overexpression studies, 200 pg of either pCS2-*satb2* or pCS2-3XFLAG-*satb2* were injected at 1 cell stage. For single-cell gene expression analysis, injections were performed with 200 pg of 3xFLAG-*satb2* at 16 cell stages targeting one cell per embryo to generate mosaicism. For rescue experiments, 5 pg of morpholino resistant Satb2 version was injected at the 1 cell stage followed by injections of 4 ng morpholinos.

The following morpholino sequences targeting translation initiation (MO1) and targeting exon splicing (MO2) were synthesized at Genetools LLC (Oregon, USA) and a 1:1 mixture was used for all knockdown experiments. Five base mismatch morpholino was used as

a control for all the experiments. Phenotypic changes were captured at indicated time points using a stereomicroscope (Olympus).

satb2 MO1 (targeting translation initiation) (5'→3')	GACTCTCTCCTCCACCACGCTCCAT
satb2 MO2 (targeting splicing) (5'→3')	TGTGAAGTGCCTGATGAGAAAAGAA
satb2 mismatch control MO (control MO) (5'→3')	GAGTGTCTCGTCCACGACCCTCCAT

4.2.3 Single-cell RNAseq analysis

To generate single-cell gene expression datasets, a mosaic overexpression system was generated by injecting one cell of 16 cell stage embryos with 200 pg of 3X-FLAG-*satb2*. Only one cell per embryo was injected randomly irrespective of its spatial arrangement. Embryos were raised at 28.5^{o}C, hand dechorionated and dissociated further in 1x DMEM-F12 + 125 mM EGTA by hand tapping. Cells were collected by centrifugation at 300x g for 2 minutes and resuspended in 1X DPBS + 0.1% BSA. The cell suspension was passed through a 70-µm Flowmi cell strainer (Sigma). Cell viability was estimated using Countess-II, an automated cell counter and only samples with more than 95% viability were processed further using the Chromium controller (10X Genomics, Pleasanton, CA, USA). Approximately 8000 cells were captured in GEMs. scRNA-seq libraries were prepared as described in 10X Genomics manuals (Single Cell 3′ Reagent Kits V3, User Guide PN-1000092). Immediately following GEM generation, reverse transcription reaction was carried out using a thermal cycler (Eppendorf MasterCycler Pro) by incubating at 53^{o}C for 45 minutes followed by denaturation at 85^{o}C for 5 minutes. Single-stranded cDNA was purified using DynaBeads MyOne Silane Beads (Thermo). cDNA amplification was performed for 11 cycles with an initial denaturation at 98^{o}C for 3 minutes followed by repeated cycles of 98^{o}C for 15 seconds, 63^{o}C for 20

seconds and 72ºC for 1 minute. The final extension was performed at 72ºC for 1 min. cDNA quality was determined using the Bioanalyzer 2100 High sensitivity assay kit (Agilent). cDNA was further fragmented, end-repaired and A-tailed according to the manufacturer's instructions. Before proceeding with adapter ligation, samples were purified using a double-side cleanup protocol with SPRI beads (Beckman Coulter). The adapter ligated sample was subjected to amplification for 11 cycles using indexing primer from Chromium i7 Multiplex Kit (PN-120262). Amplified libraries were again subjected to double-side purification using SPRI and quantified using Qubit fluorometer (Thermo Scientific) and library size was estimated using Bioanalyzer 2100 (Agilent).

1.5 pM of the denatured library was used as an input to obtain sequencing reads using 28 cycles for read1, 8 cycles for indexes and 101 cycles for read2 on Nextseq 550 (Illumina) at IISER Pune.

Sequencing data were further processed with default parameters using cell Ranger 3.0.2 (10X Genomics). Sequencing reads were aligned to the danRer10 genome using STAR aligner. Feature matrices generated using cell Ranger were utilized for further analysis using Seurat 3.0.

4.2.4 NT2/D1 cell culture

Human embryonic carcinoma cell line NT2/D1 (NTERA2-clone D1) was a kind gift from Dr Peter Andrews, University of Sheffield, UK. They were cultured in Dulbecco's Modified Eagle's Medium with sodium pyruvate, high glucose (DMEM, Sigma-Aldrich, St. Louis, Missouri, USA) supplemented with 10% foetal bovine serum (Invitrogen, Carlsbad, California, USA), 2 mM L-glutamine (Invitrogen, Carlsbad, California, USA) and penicillin-streptomycin (Invitrogen, Carlsbad, California, USA) and maintained at 37°C under 5% CO2 atmosphere. NT2/D1 cells were sub-cultured upon reaching 70% confluency by gentle scraping.

4.2.5 All-trans retinoic acid-mediated differentiation of NT2/D1

All-trans-retinoic acid (RA) (Sigma-Aldrich, St. Louis, Missouri, USA) was reconstituted at a concentration of 5 mg/ml in DMSO (Sigma-Aldrich, St. Louis, Missouri, USA) and stored at -80°C. For differentiation, NT2/D1 cells were harvested using 0.05% trypsin resuspended in fresh medium and seeded at a density of 1 X 10^6 cells in a 100 mm tissue culture dish (Corning, New York, USA). Cells were allowed to grow for 24 hours following which 1 X 10^7 cells were harvested for RNA and protein extractions as 0-day control. RA was added to the remaining plates at a concentration of 13.7 µM for the rest of 5 days. Each day cells were either replenished with fresh medium and RA or harvested for RNA.

4.2.6 qPCR analysis

Embryos were collected at the desired stage and were lysed in RNA iso plus (DSS-Takara, India) to extract total RNA. Each sample consisted of at least 10 embryos per experiment, and all experiments were repeated at least three times independently. cDNA was synthesized using High capacity cDNA synthesis kit (Applied Biosystems), following the manufacturer's instruction.

Quantitative real-time PCRs were performed using the SYBR green chemistry (KAPA biosystems) with gene-specific primers as listed below on the ViiA7 thermal cycler (Applied Biosystems). Changes in threshold cycles were calculated by subtracting the Ct values of the gene of interest from that of housekeeping control (for qRT-PCR) [Ct(target genes) – Ct(*ef1a* or *GAPDH*)]. ΔCt values of specific target genes from the experimental samples were then subtracted from their respective control samples to generate ΔΔCt values. The fold changes were calculated using the formula: 2^(-(ΔΔCt value).

Target site	Forward primer (5'→3')	Reverse primer (5'→3')
Danio_satb2	CAAGAGTTTGGTCGCTGGTA	CGCTGGGCTAATACACAGAA
Danio_ef1a	CTTCTCAGGCTGACTGTGC	ACGATCAGCTGTTTCACTCC
Human_satb2	CCAGAGCACATTAGCCAAAGA	TGTGCTATTTACAATGGATGAAATC
Human_pax6	GGCACACACACATTAACACACTT	GGTGTGTGAGAGCAATTCTCAG
Human_18s	CGCCGCTAGAGGTGAAATTCT	CGAACCTCCGACTTTCGTTCT

4.2.7 Molecular cloning of expression plasmids

For zebrafish Satb2 over-expression studies, full-length *satb2* was amplified from cDNA samples generated using High capacity cDNA synthesis kit (Thermo Scientific, USA) from 48 hpf wild type TU embryos. The resulting PCR product was purified using a Monarch PCR purification kit (NEB, USA) and was cloned into a pCS2+ vector using restriction enzymes BamH1 and SnaB1 based cloning to generate pCS2-*satb2*. To generate a 3xFLAG-tagged construct, an oligo containing 3x-FLAG sequence at the 5' end of the forward primer was used for PCR using pCS2-*satb2* plasmid as a template and a clone was obtained using the Gibson assembly method. To generate morpholino resistant clones, oligo containing 7 mismatches spanning the first 24 bases of the coding region were used for amplification using pCS2-*satb2* plasmid as a template.

Danio_*satb2* Forward (5'→3')	TCTTTTTGCAGGATCATGGAGCGTGGTGGAGGAG
Danio_*satb2* Reverse (5'→3')	CATGTCTGGATCTACTCTCTGATAGTCTGCTTGATCTCTG
3xFLAG-Danio_*satb2* Forward (5'→3')	TTGTTCTTTTTGCAGGACTACAAAGACCATGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGAGCGTGGTGGAGG
3xFLAG-Danio_*satb2* Reverse (5'→3')	CGAATCGATGGGATCTTATCTCTGATAGTCTGCTTGATCT
Mismatch-Danio_*satb2* Forward (5'→3') (mismatches are highlighted by red text)	TTGTTCTTTTTGCAGATGGA**A**CGTGG**C**GG**C**GGAGA**ATCA**CCTCGA

4.2.8 Protein extraction and immunoblotting

To confirm the overexpression of Satb2 in zebrafish embryos, control and 3x FLAG Satb2 mRNA injected embryos were manually dechorionated, deyolked and harvested at 4.5 hpf by boiling in 2X Laemmli buffer (0.25 M Tris-Cl pH 6.8, 1% SDS, 1% β-mercaptoethanol, 15% glycerol) at 98^{o}C for 5 minutes and electrophoresed on 7.5% of the SDS-PAGE gel. Proteins were transferred to the 0.45 µM PVDF membrane (Millipore) by wet transfer method at 0.6 A for 2.5 hrs at 4^{o}C. Non-specific sites on the membrane were blocked using 3% BSA and further incubated with anti-FLAG antibody (1:1000) in 0.3% BSA overnight at 4^{o}C with constant rocking. The next day, Membrane was washed with 1X TBST (50 mM Tris-Cl pH 7.4, 150 mM NaCl, 0.1% Tween-20) 4 times for 7 minutes each and incubated with anti-Mouse HRP conjugated antibody (Bio-Rad) in 1x TBST for 45 min at room temperature. After removing excess secondary antibody by

repeated washes with 1x TBST, the signal was developed with Clarity Western ECL substrate (Bio-Rad) and captured with LAS 4000 system (GE healthcare).

4.2.9 Synthesis of riboprobes for Whole-mount in situ hybridization

Synthetic oligonucleotides against target mRNAs (listed below) with T7 RNA polymerase promoter sequence on the 5' end of the reverse primers were obtained commercially (Sigma Aldrich, India). cDNA templates for corresponding stages were used for PCR amplification. PCR reactions were purified using Monarch DNA Gel Extraction Kit (New England Biolabs) or HighPrep PCR Clean-up System beads (MagBio, USA). Further, Riboprobes were synthesized using a DIG RNA labelling kit (Roche) at 37^{o}C for 3 hrs followed by DNAse treatment for 15 mins. Probes were further purified using Bio-Spin 30 Tris Columns (Biorad, Germany). The digoxigenin-labelled probes were 400-700 nucleotides in length and stored at -20 ^{o}C at 1 µg/ml.

For *chrd* WISH experiments amplified region was cloned into TOPO II dual promoter vector (Thermo Scientific, USA). To generate an antisense probe, the plasmid was linearized using BamHI and probe synthesis reaction was carried out using T7 RNA polymerase promoter as described above.

Target gene	Forward primer (5'→3')	Reverse primer (5'→3')	Product size
mych	CTCCGACATAGACACGCAGA	gagtaatacgactcactatagggGTCTGGCTTTCAGCTGTTCC	710 bp
vox	AGACGGAGAGCAGCAAAGAG	gagtaatacgactcactatagggGAGGATGAGGATGGTGAGGA	504 bp
foxd5	TCTCCAACCATGACCCTCTC	gagtaatacgactcactatagggACCTCTGGGTTTTGTGTTCG	763 bp
mxtx2	TCTGCAAGAGAGCTGCAAAA	gagtaatacgactcactatagggAGGCACAGATGGAGAGCAGT	770 bp
chrd	GTGGCCGCTTTTACTCTG	GTGAGGTTTCGGCACATTCT	596 bp
pou5f3	CGGAAAGATGACGGAGAGAG	gagtaatacgactcactatagggAGCTCTTTCGCAAACTGCTC	798 bp

4.2.10 Whole-mount In situ hybridization analysis

Whole-mount In situ hybridization (WISH) was performed as described previously (Thisse and Thisse, 2008). Fertilized embryos from satb2$^{+/-}$ incross were raised until the desired stage of development was reached. Embryos were fixed in 4% paraformaldehyde in 1x PBS overnight at 4^{o}C with gentle rocking. The next day, embryos were hand dechorionated with a pair of fine forceps and incubated in 100% MeOH overnight at -20^{o}C. Embryos were further downgraded in methanol: PBST solution (0.1% Tween-20 in 1x PBS pH 7.4). Following successive rehydration to PBST, embryos were treated with 5

µg/ul proteinase K solution in PBST for 2 min for 14ss stage embryos. The reaction was stopped by a subsequent 20 min incubation in 4% PFA. For the Dome stage, embryos were not treated with proteinase K. After washes in PBST, samples were pre hybridized for 2-5 hours and incubated overnight at 70^{o}C in approximately 500 ng of the respective probe. The next day, embryos were washed with SSC buffers and further blocked in blocking buffer (2% v/v normal goat serum, 2mg/ml BSA in PBST) for 3-4 hours at room temperature followed by overnight incubation at 4^{o}C in anti-DIG-AP Fab fragments (1:5000) (Roche 1093274). Embryos were washed with PBST followed by an alkaline Tris buffer. Staining was performed with BM Purple substrate (Roche). After staining, the colour reaction was stopped using a stop solution containing 20% Tween-20 and 0.5 M EDTA in 1x PBS (pH 5.5). Embryos were further washed with PBST and stored at 4^{o}C until imaging. Embryos for imaging were mounted on agarose moulds and imaged using Z-stacking mode on a Leica DFC450C microscope. Post imaging, embryos were genotyped using the standard protocol discussed above. Representative images were further processed using ImageJ for adjusting brightness and contrast.

4.2.11 mRNA sequencing and differential gene expression analysis

For Satb2 overexpression studies, 3x-Flag-Satb2 mRNA was injected at 1 cell stage and embryos were harvested at 4.5 hpf and were lysed in RNA iso plus (DSS-Takara, India) to extract total RNA. Each sample consisted of at least 10 embryos per experiment. For Satb2 knock-down studies, morpholino mix was injected at 1 cell stage and embryos were harvested at 4.5 hpf. Approximately 500 ng of Total RNA with RIN > 9 was used as a starting material to prepare mRNA sequencing libraries using the SENSE mRNA seq kit (Lexogen GMBH, Austria). All cDNA samples were amplified for 10-15 cycles depending on cycle number estimation by qPCR using the PCR add-on kit (Lexogen GMBH, Austria). Amplified libraries were purified using two rounds of 0.8x volume of Hi-prep PCR purification kit. The concentration of libraries was estimated using the Qubit 4 DNA HS system (Thermo Scientific). Finally, all the libraries were pooled and subjected to high throughput sequencing using 100 bp PE chemistry on Hiseq 2500 (Macrogen, Korea).

Sequencing reads were trimmed for quality using Trimmomatic and aligned to daRer10 genome assembly using HISAT2 aligner (Kim et al., 2019). Counts for each gene feature was estimated using the FeatureCounts package from Rsubread. Differential expression analysis was performed for replicates using EdgeR. A volcano plot for significantly differentially expressed genes was generated using the PlotVolcano tool from the Galaxy server. Gene ontology analysis for upregulated and downregulated genes were performed using a web-based tool, gProfiler (Reimand et al., 2007).

4.2.12 ChIP-sequencing for histone modification marks

ChIP setup

To profile the status of histone modifications upon overexpression of 3xFLAG tagged Satb2, 1000 embryos were harvested for each ChIP reaction at 4.5 hpf (Dome stage) in batches until a sufficient amount of input material is generated. Empty pCS2-GFP was used as a control for injections. Pulldowns were performed with 20 µg chromatin and 5 µg anti-H3K4me3 (ab8580 Abcam), 5 µg anti-H3K27Ac (ab4729 Abcam), 5 µg anti-H3K27me3 (ab6002 Abcam) and 5 µg of normal rabbit IgG overnight at 4^{o}C. Antibody-Protein complex was captured using 50 µl of Pre blocked Dynabeads A: G mixture (Thermofisher) for 3 hrs at 4^{o}C. Samples were processed further as described in the earlier section.

Library preparation, sequencing and data analysis

A total of 5 ng of purified ChIPped DNA was used for library preparation using Ultra II DNA kit (NEB) as described in the earlier section. Sequencing reads (100 bp PE) were obtained on the Hiseq X platform (Illumina) at Macrogen Inc, Korea. Sequencing reads were trimmed using TrimmomaticPE for Truseq2:PE adapters and read with quality greater than Phred 33 were retained (Bolger et al., 2014). The quality of sequencing reads was determined using fastQC. High-quality sequencing reads were aligned to zebrafish danRer10 genome version using default parameters of BWA (Li and Durbin, 2009). Aligned reads were subsampled to 40 million reads in each sample using BBmap (Bushnell, 2014). RPKM normalised Input signal subtracted BigWig tracks for

visualization were generated using bamCompare tool from deepTools 3.3.2 (Ramírez et al., 2016).

4.2.13 TF binding promoter scan

Consensus binding sites for respective TFs were extracted from the JASPAR database and MAST (MEME Suite) was used for scanning genomic regions extracted from the UCSC server using version danRer10.

4.2.14 ATAC-seq and data analysis

To determine the effect of Satb2 overexpression on chromatin accessibility, 200 pg 3xFLAG-*satb2* were injected at 1 cell stage. Embryos (a pool of 10 embryos for each replicate) were harvested at 4.5 hpf and processed for Omni-ATAC seq as described previously (Corces et al., 2017) with inputs from Dr Amanda Ackermann. Briefly, cells were washed with 1x DPBS and resuspended in cell lysis buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM KCl, 0.1 % NP 40, 0.1% Tween20 and 0.01% Digitonin) and incubated on ice for 3 minutes. Further, cells were washed with a wash buffer (10 mM Tris pH 7.5, 10 mM NaCl, 3 mM KCl and 0.1% Tween20) by centrifugation at 500 g for 10 mins 4^{o}C. The supernatant was discarded and the pellet was resuspended in 25 µl 2x Tagmentation buffer (Illumina, catalogue # 15027866), 16.5 µl DPBS, 0.5 µl 10% Tween 20, 0.5 µl 1% Digitonin, 5 µl nuclease and 2.5 µl Tn5 transposase enzyme (TDE1, Illumina, catalogue # 15027865) and incubated for 28 min at 37^{o}C. After the tagmentation reaction, DNA was isolated using the MinElute Reaction Cleanup Kit (Qiagen). Purified DNA was used as an input to generate a library by amplifying with 2x Q5 DNA polymerase mix (NEB) and indexing primers. Optimal cycles were determined using qPCR analysis. Amplified libraries were purified using Agencourt ampure XP beads to remove adapters and larger fragments.

Sequencing reads (41 bp PE) were obtained on NExtseq 550 at IISER Pune and trimmed for Nextera adapters using default parameters of Trimmomatic PE. Trimmed reads were aligned to danRer10 using default parameters of Bowtie2 (Langmead and Salzberg, 2012). Briefly, BAM files were subsampled to 55 million reads in each sample using

bbmap and sorted by name. paired-end bed files were obtained using bedtools bamtobed. Reads were displaced by +4 bp and -5 bp. BigWig files were generated using bamCoverage (deepTools). Peaks were annotated to the nearest genes using Homer and classified into promoter (+/- 2 Kb) and non-promoter regions. K-means clustering was performed around +/- 2 Kb of TSS of danRer10 genome using deepTools.

4.2.15 Statistical analysis

GraphPad Prism (GraphPad Software, USA) was used for statistical analysis. The type of analysis used in each experiment is mentioned in the respective figure legends.

4.3 Results

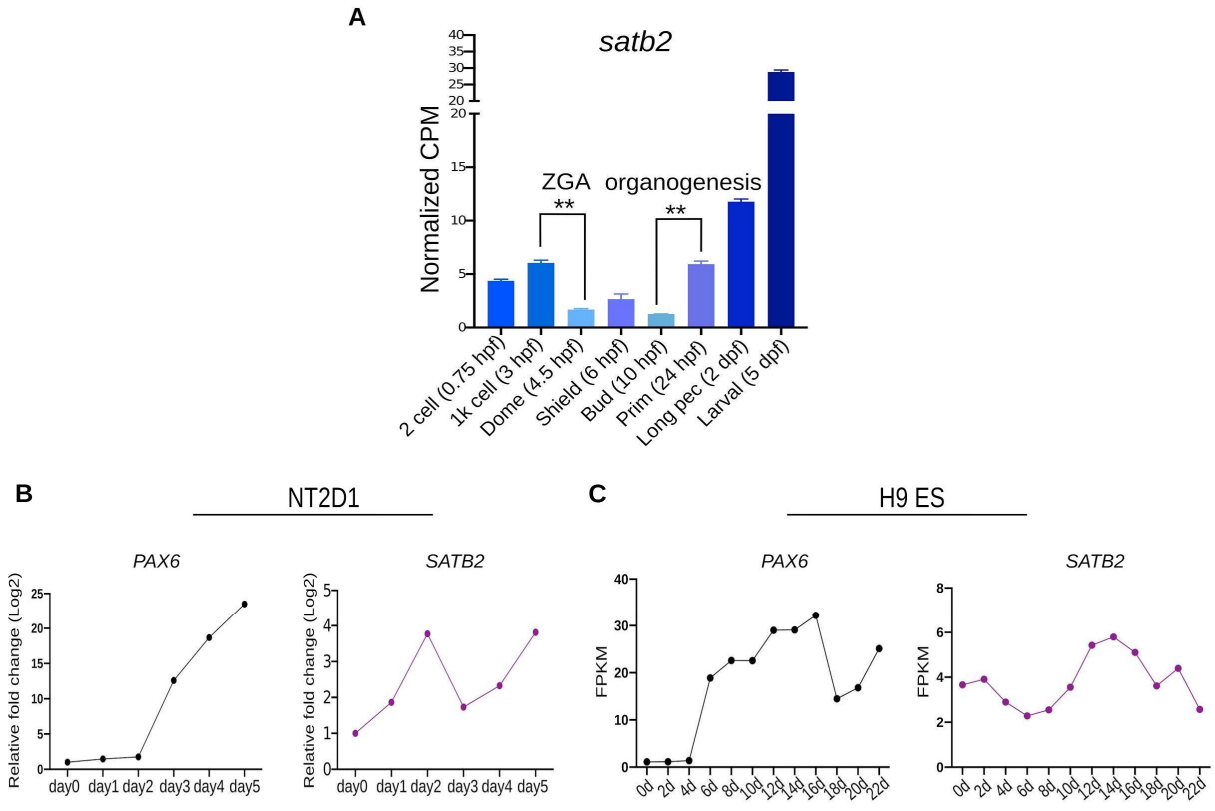
4.3.1 Satb2 exhibits bi-phasic expression patterns during zebrafish embryonic development and neuronal differentiation in stem cells.

Data presented in the previous chapter (Chapter 3) demonstrated a direct regulatory role of zygotic Satb2 during early neurogenesis and NCC development. Intriguingly, Satb2 is also maternally deposited and is thought to function primarily during epiboly (Ahn et al., 2010). To confirm the presence of maternal Satb2, we reanalyzed publicly available time-series transcriptome data during zebrafish embryogenesis (Pauli et al., 2012) and (https://www.ebi.ac.uk/gxa/experiments/E-ERAD-475). Interestingly, through temporal transcriptome analysis, we observed bi-phasic expression of Satb2 in which maternal Satb2 is significantly reduced at the onset of ZGA and maintained at a minimal level before zygotic Satb2 starts expressing during organogenesis (Figure 4.1A).

Further, we wondered if this characteristic pattern of Satb2 expression can be mimicked *in vitro* using pluripotent NT2D1 human embryonic cells. Upon retinoic acid stimulation, NT2D1 cells differentiate into neuronal progenitors (Coyle et al., 2011). We found that while *SATB2* is readily detectable in undifferentiated cells, its expression decreases during transition stages (day 3 post RA treatment). Interestingly, we observed a further

increase in the expression levels of *SATB2* as cells differentiated into neuronal progenitors, characterized by an increase in *PAX6* expression (Figure 4.1B). To further support this, we reanalyzed expression levels of *SATB2* and *PAX6* in a differentiation model established by Li *et al.* using H9 human embryonic stem cells (Li et al., 2017). This study sub-categorized the neuronal differentiation program into five gene expression modules and emphasized 'module 3' (day8-day10 post differentiation) to be largely responsible for neuronal fate commitment. Interestingly, Satb2 is expressed at the highest level in this gene module, thus highlighting its function as a driver for neuronal fate determination. Moreover, SATB2 exhibits a similar biphasic expression pattern as observed during the early embryonic development of zebrafish (Figure 4.1C).

Figure 4.1 Expression dynamics of Satb2 during zebrafish early embryogenesis and neural progenitor differentiation systems using human embryonic stem cells.



(A) Stage-specific gene expression analysis highlighting bi-phasic expression patterns of Satb2 during MZT and organogenesis. Normalized CPM values for each stage are used for the analysis. (** represents P-value of < 0.05 as measured by students two-tailed t-test). **(B)** Line graph depicting average mRNA expression levels for *PAX6* and *SATB2* in retinoic acid differentiation of NT2D1 cells (n=2). **(C)** Line graph representing mRNA expression analysis of *PAX6* and *SATB2* in neural differentiation model in H9 ES cells highlighting biphasic mode of expression.

Together, experiments using the NT2D1 cells and analysis of the H9 differentiation series suggested that *SATB2* is expressed during the pluripotency stages and its expression reduces during cell differentiation stages before it is restored during the stages of neural fate commitment. These results highlight the conserved nature of expression dynamics of Satb2 during early pluripotent stages to neurogenesis between zebrafish and humans. However, this observation is in contrast to earlier observations made in mouse embryonic stem cell differentiation models suggesting Satb2 expression decreases upon RA-induced differentiation (Savarese et al., 2009).

4.3.2 Satb2 is required for pattern specification during early embryogenesis

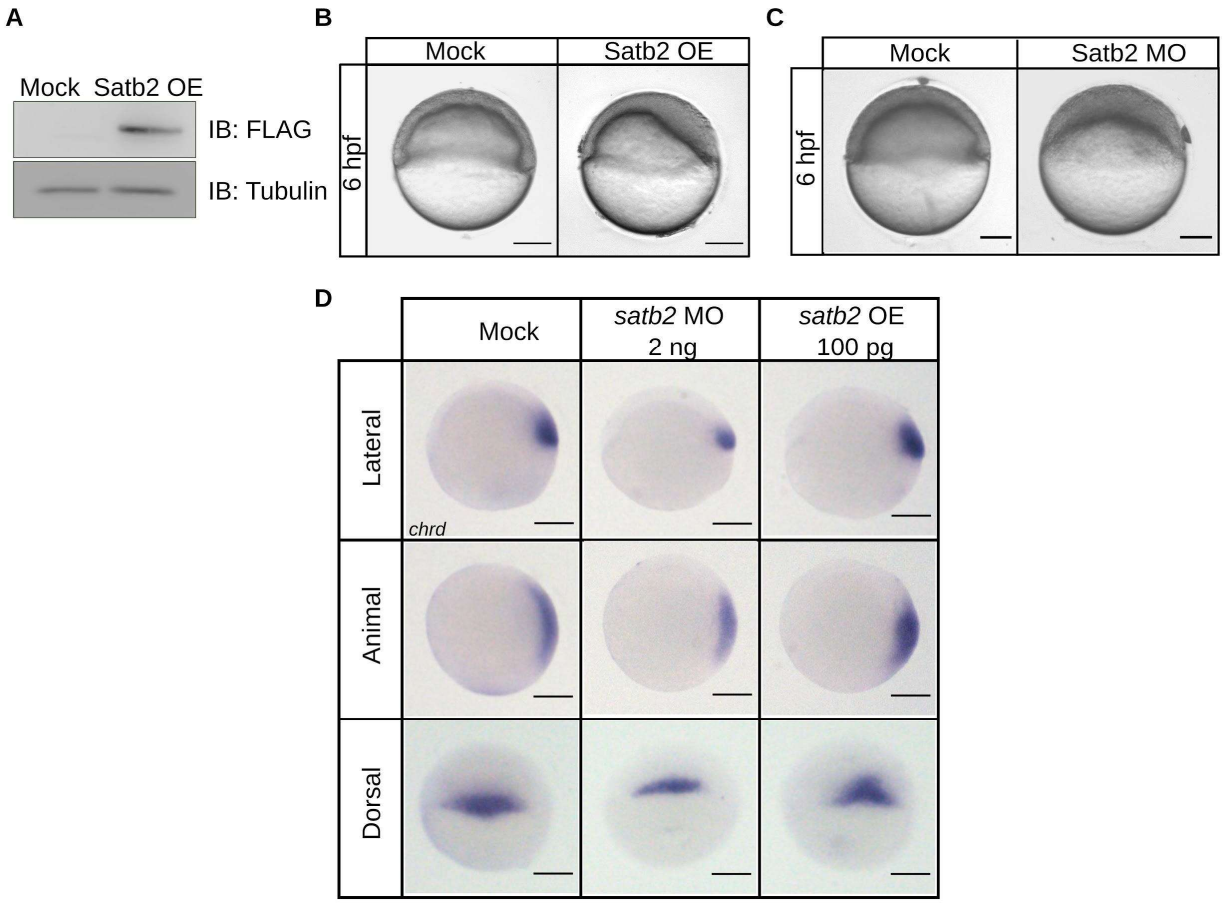
To assess the function of maternally deposited *satb2*, we attempted to generate maternal-zygotic (MZ) mutants for *satb2*. As in the case of mice, zygotic *satb2* mutants show poor survival and the rare survivors are infertile (Chapter 3, Figure 3.3) (Dobreva et al., 2006). To circumvent this difficulty, we attempted germ-line transplantations but the resulting fish were found to be infertile. Consequently, with the currently available methods, we could not obtain maternal zygotic mutants to directly assess the consequences of the loss of function of maternal Satb2.

We, therefore, turned our attention to the possible biological relevance of the disappearance of maternally deposited s*atb2* at the onset of a major wave of ZGA. To assess this we engineered sustained expression of Satb2 by using an overexpression

construct from the 1 cell stage (Figure 4.2A). Overexpression and knockdown of Satb2 resulted in significant developmental defects during epiboly (Figure 4.2B), corroborating the conclusion that temporally controlled disappearance of Satb2 is biologically relevant. Thickening of the dorsal organizer region was also apparent suggesting defects during early morphogenesis events. A previous study in *Xenopus* identified Satb2 as an uncharacterized factor enriched in the dorsal organizer, thus suggesting it to be involved in dorso-ventral patterning (Popov et al., 2017).

To confirm this directly, we inactivated the satb2 function using two different morpholinos targeting the translation start site and splice junction, thereby resulting in early developmental arrest (Figure 4.2C). Five base Mismatch morpholino was used as a control. Importantly, in contrast to *satb2* overexpression, knockdown embryos exhibit reduced expression of dorsal marker *chrd* (Figure 4.2D). Together, these results suggest that precise levels of Satb2 expression are required for dorso-ventral pattern specification in an embryo during gastrulation. Remarkably, both *satb2* overexpressing and *satb2* depleted embryos fail to survive beyond 8 hpf, which indicates possible global defects in cell-type specification and/or failed morphogenesis.

Figure 4.2 Manipulating levels of Satb2 during early stages of embryogenesis leads to defective pattern specification resulting in developmental arrest.

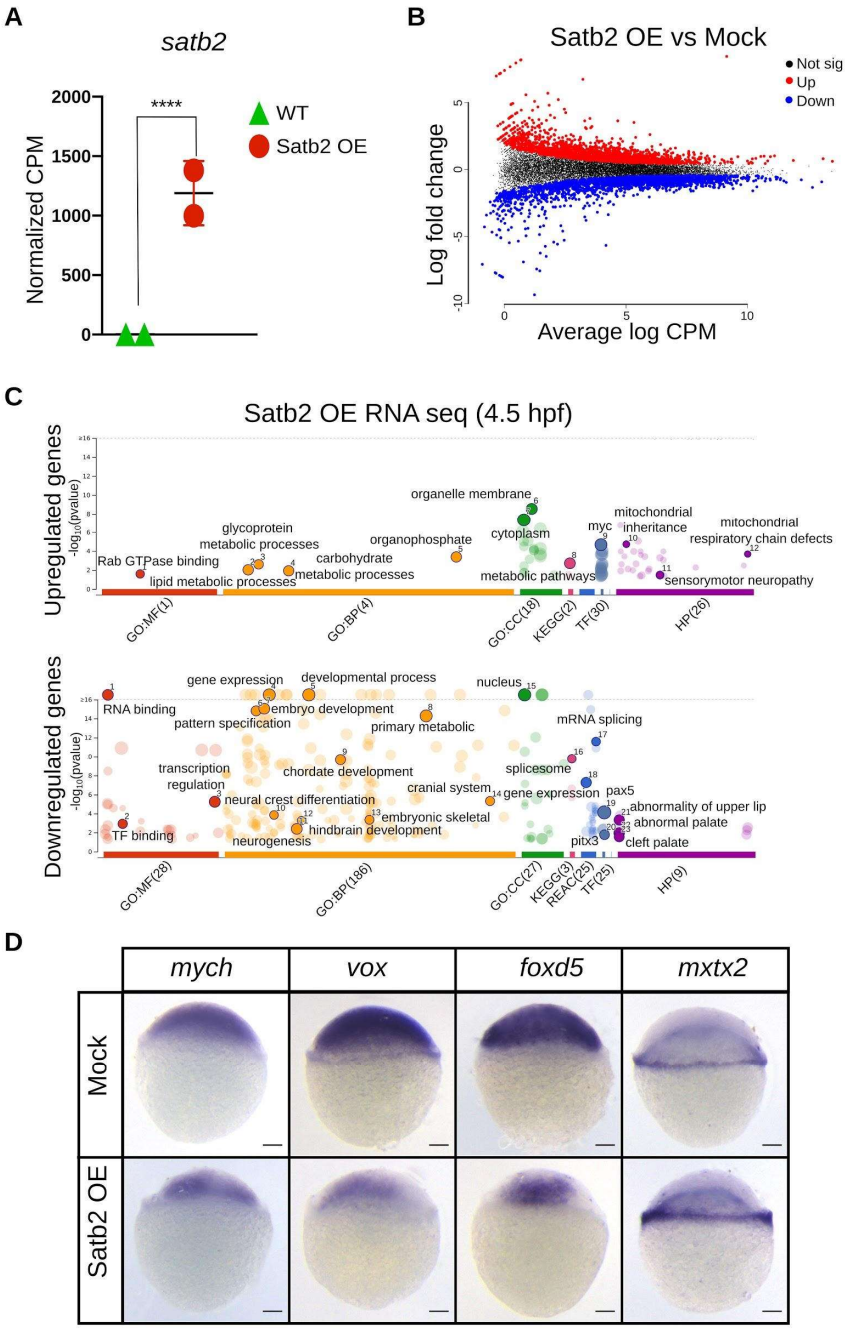


(A) Immunoblot analysis for validating overexpression of 3x-FLAG Satb2 at dome stage using anti-FLAG antibody. γ-Tubulin was used as a loading control. **(B)** Lateral view of zebrafish embryos at 6 hpf injected with 200 pg eGFP (control) and 3xFlag-Satb2 mRNA. **(C)** Lateral view of zebrafish embryos at 6 hpf injected with 4 ng of control and Satb2 targeting morpholinos. **(D)** Whole-mount in situ mRNA analysis for stage-matched embryos showing the effect of Satb2 knockdown and overexpression on the expression of dorsal marker *chrd*.

4.3.3 Maternal Satb2 acts as a transcriptional repressor to control activation of developmental pathways

To understand the molecular functions of maternally deposited Satb2, we performed transcriptome analysis in control mRNA and 3xFLAG-*satb2* coding mRNA injected embryos at 4.5 hpf (Figure 4.3A). Interestingly as in the case of zygotic *satb2* mutants, overexpression of *satb2* also resulted in significant changes in gene expression including both enhancement and downregulation of transcription of select groups of genes (Figure 4.3B). Interestingly, gene ontology analysis revealed that upregulated genes primarily constitute part of metabolic processes whereas downregulated genes were enriched in embryonic development and pattern specification (Figure 4.3C). We further validated our results from transcriptome analysis using whole-mount in situ hybridization analysis for selected genes. As evident from this analysis, overexpression of *satb2* indeed resulted in downregulation of *mych*, *vox* and *foxd5* and upregulation of *mxtx2* (Figure 4.3D). This data primarily suggests that, unlike zygotic *satb2*, maternal *satb2* potentially acts as a repressor of genes involved in developmental processes.

Figure 4.3 Transcriptome analysis to characterize the gain of function of maternal Satb2.



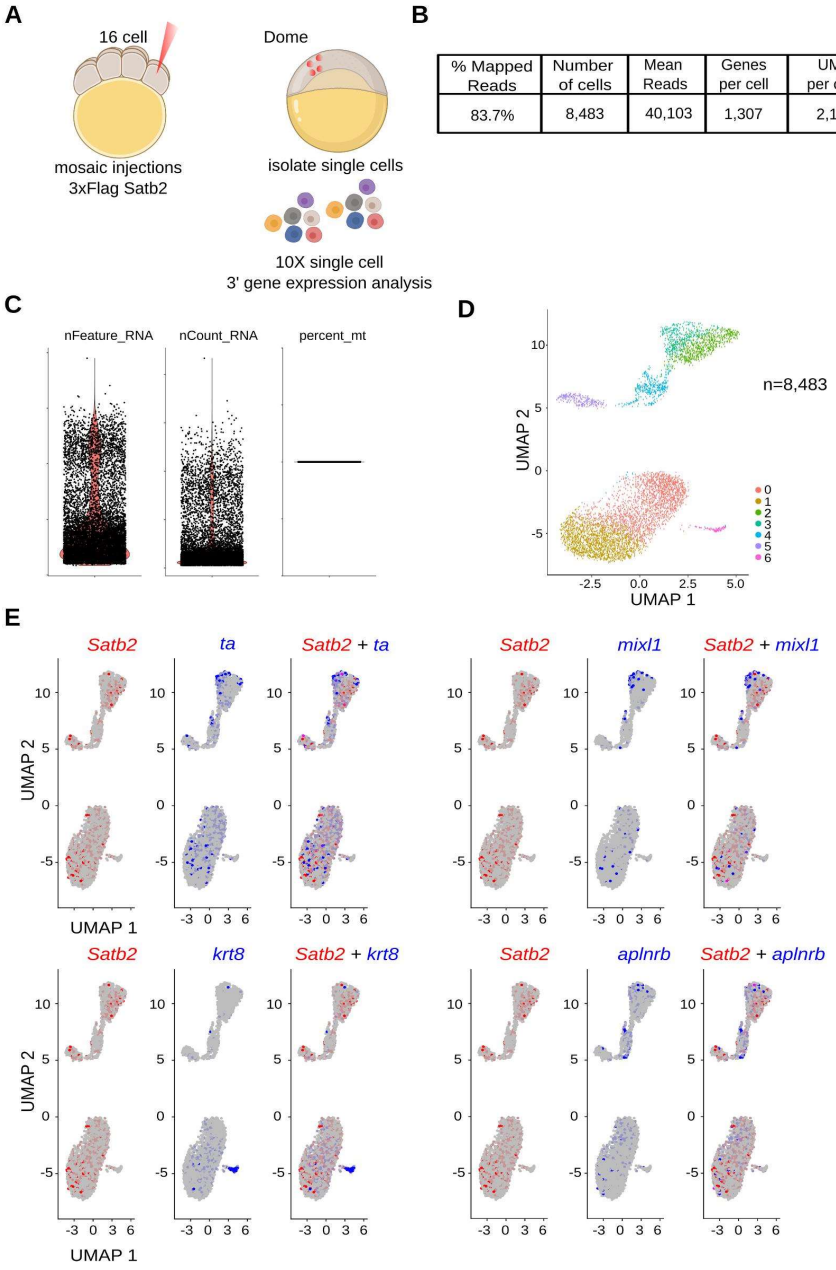
(A) Dot plot representing and confirming overexpression of Satb2 from RNAseq studies performed at dome stage (4.5 hpf). Normalized CPM counts were used for the analysis (**** p-value 0.0001). **(B)** Smear plot analysis for bulk mRNA seq analysis in Mock and

3xFLAG-Satb2 injected embryos at 4.5 hpf. Values in red and blue represent the number of genes differentially upregulated and downregulated respectively. **(C)** GO analysis for differentially expressed genes upon over-expression of Satb2 at 4.5 hpf. Significant ontologies under molecular function (MF), biological processes (BP), Cellular components (CC), Reactome (REAC), Transcription factors (TF) and Human pathology (HP) categories are highlighted. Numbers in brackets signify the total number of ontologies enriched under each category. **(D)** Whole-mount in situ mRNA analysis showing the effect of Satb2 overexpression on expression of *mych*, *vox, foxd5* (downregulated genes) *mxtx2* (upregulated upon overexpression).

4.3.4 Satb2 functions in a cell-autonomous manner

To assess the nature of Satb2's function during ZGA we employed single-cell gene expression analysis. We overexpressed *satb2* in a mosaic manner by injecting 3xFLAG-*satb2* in one of the cells at 16 cell stage embryos and harvested embryos at 4.5 hpf to prepare single-cell suspension (Figure 4.4A). This strategy also enabled us to obtain wild-type cells from the same source. We generated a high-quality dataset for 3' gene expression in 8483 single cells using scRNA-seq (Figures 4.4B and 4.4C). UMAP clustering of these cells identified 7 distinct clusters (cluster 0 to cluster 6) based on the correlation of gene expression patterns (Figure 4.4D). Next, we analyzed the expression of individual genes as specific features and generated spatial maps for *satb2* and developmental genes which are found to be negatively regulated by Satb2 through our bulk mRNA-seq analysis. To exemplify, we represent feature maps for *satb2*, *ta*, *mixl1*, *krt8* and *aplnrb* and highlight a negative correlation between the expression of *Satb2* and these developmental genes (Figure 4.4E). In conclusion, mosaic overexpression followed by scRNA-seq indicated that Satb2 functions in a cell-autonomous manner.

Figure 4.4 sc-RNAseq analysis reveals cell-autonomous nature of transcriptional regulation by maternal Satb2.



% Mapped Reads	Number of cells	Mean Reads	Genes per cell	UMI per cell
83.7%	8,483	40,103	1,307	2,168

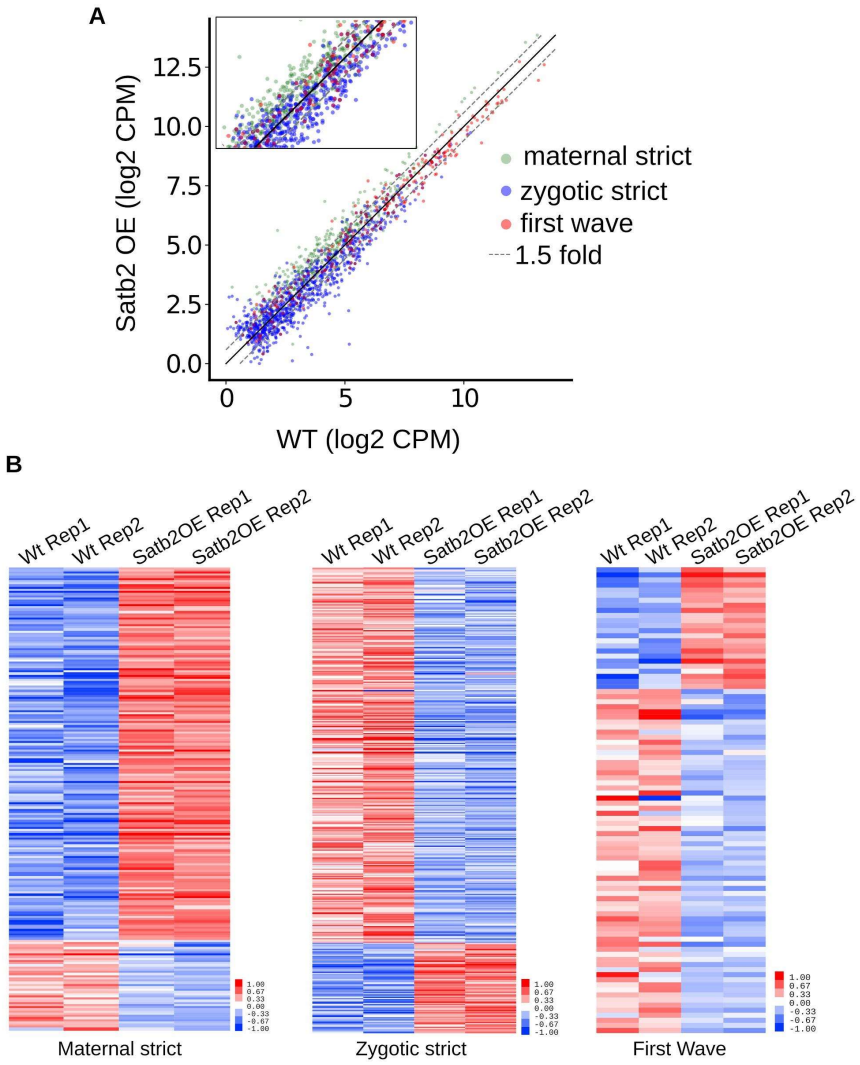
(A) Experimental strategy for single-cell analysis of embryos with mosaic overexpression of Satb2 at 4.5 hpf. **(B)** Table representing statistics associated with single-cell RNA seq experiment. **(C)** Violin plot for visualization of QC metrics of single-cell transcriptome

study. **(D)** UMAP clustering illustrates 8 different clusters based on gene expression patterns from 8,483 single cells. **(E)** FeaturePlots for exemplifying negative correlation between cells expressing *satb2* (blue) and zygotic factors *ta*, *mixl1*, *krt8* and *aplnrb* (red).

4.3.5 Sustained expression of maternal *satb2* perturbs successful onset of zygotic genome activation

To gain further insights into underlying mechanisms of the effect of *satb2* level manipulation on embryonic development, we decided to categorize the transcriptional targets of Satb2, based on the classification by Lee *et al*. to segregate genes into strictly maternal, strictly zygotic and first-wave zygotic genes (Lee et al., 2013). Interestingly, we observed a clear trend. The strictly zygotic, and first wave zygotic genes were significantly downregulated whereas maternal genes were upregulated (Figures 4.5A). K-means clustering analysis of the transcriptome dataset also confirmed negative regulation of zygotic genes by maternally deposited *satb2*. These observations were in accordance with the gene ontology analysis (Figure 4.3C). Since ribosomal biosynthesis and metabolism processes-related genes are deposited maternally to support the early development (Dworkin and Dworkin-Rastl, 1990) and are found to be upregulated upon sustained expression of *satb2.* These results shed insights into functions of Satb2 during early cleavage stages required for the precise onset of zygotic genome activation.

Figure 4.5 Sustained presence of maternal Satb2 leads to transcriptional repression of zygotic genes.

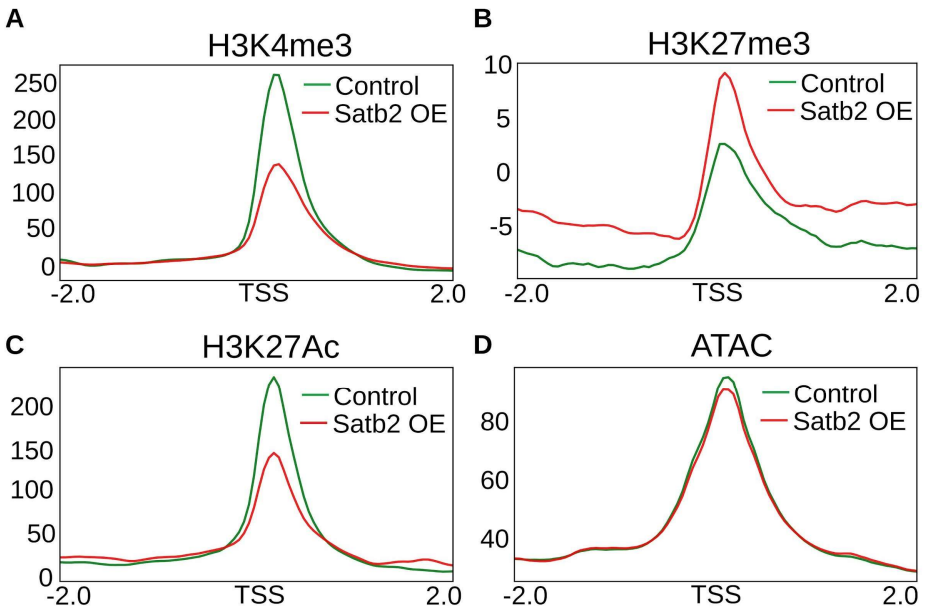


(A) Biplot depicting the effect of overexpression of Satb2 on maternal (green), zygotic (blue) and first wave zygotic genes (red). Log2 CPM values are used for the analysis. The dotted line marks genes that show more than 1.5 fold differences in Log2 CPM values. Inset is a zoomed version to highlight the differences. **(B)** Heatmaps representing differential expression patterns for strictly maternally expressed genes, strictly zygotically expressed genes and first wave zygotic genes. The individual samples are labelled on top.

4.3.6 Maternal Satb2 modulates local chromatin landscape

Considering the activity of Satb2 as a chromatin remodeler, we wondered if the repression of zygotic genes is correlated with corresponding epigenomic changes. For this, we profiled the occupancy of H3K4me3 (activation mark), H3K27me3 (repressive mark), H3K27Ac (activation mark) and analyzed the chromatin accessibility (ATAC-seq) focusing on the TSS of 'zygotic strict' genes. Curiously, we observed depletion and increase in the activation and repressive marks respectively, despite a very minimal effect on chromatin accessibility (Figure 4.6A-D). This could be presumably due to transient regulation of the epigenomic landscape or active compensatory mechanisms.

Figure 4.6 Transcriptional repression of zygotic genes by maternal Satb2 is mediated through changes in the local chromatin landscape.



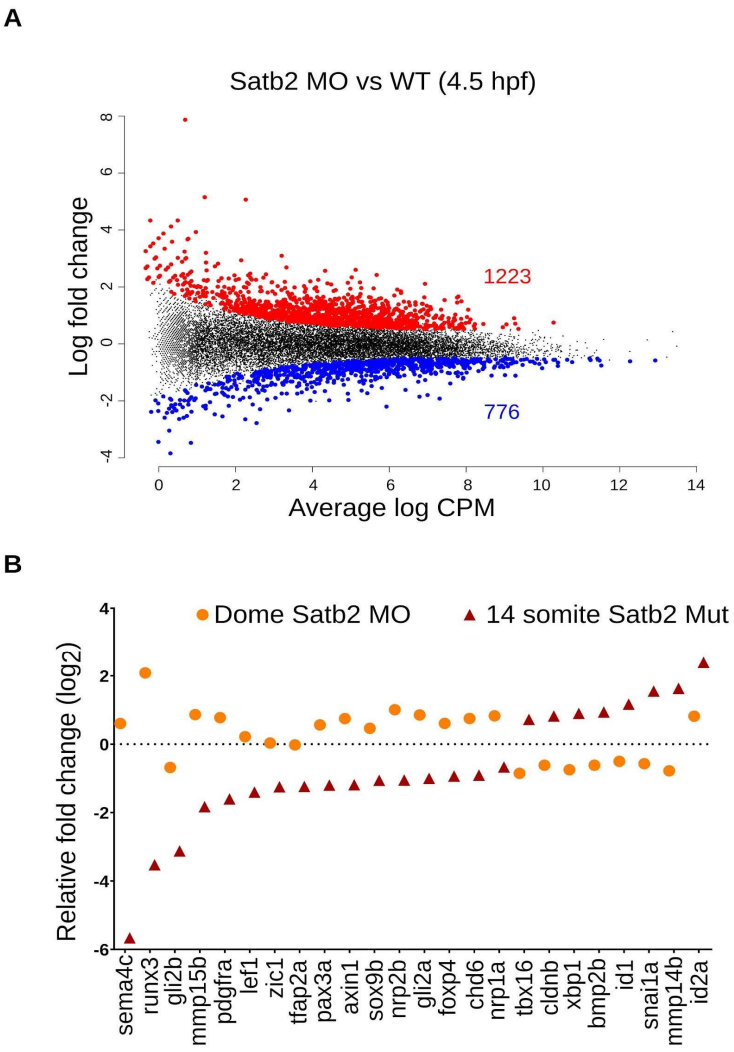
(A) Average mean density profiles around +/- 2 kb regions flanking the TSS of 'zygotic genes' for H3K4me3 (activation mark) **(B)** H3K27me3 (repressive mark) **(C)** H3K27Ac (Activation mark) and **(D)** Chromatin accessibility profiles around +/- 2 kb regions flanking the TSS (ATAC-seq analysis) upon Satb2 overexpression (red) compared to control (green).

4.3.7 Maternally deposited Satb2 is required to prevent premature activation of zygotic genes involved in organogenesis

In Chapter 3, we showed that Satb2 positively regulates the genes involved in neural crest cell specification and neurogenesis. However, we also observed that maternally deposited Satb2 acts as a transcriptional repressor of zygotically transcribed genes. This led us to ask, what would be the expression status of the genes involved in neural crest specification and neurogenesis if we deplete the maternal pool of Satb2. If maternal Satb2 acts primarily as a repressor as opposed to the function of zygotic Satb2 and thereby regulates the timing of ZGA, then upon depletion of maternal Satb2 we should observe premature upregulation of zygotic genes. As discussed in previous sections, since we could not generate maternal zygotic mutants for Satb2 with the help of available tools, we decided to deplete maternally-deposited Satb2 using antisense oligonucleotides, morpholinos.

Next, we performed bulk mRNA-seq and specifically analyzed the expression of genes involved in neurogenesis and neural crest differentiation, which are positively regulated by zygotic Satb2. We also analyzed genes that are negatively regulated as a control group. Indeed, opposite gene expression patterns were observed and many genes which serve as markers for neurogenesis and neural crest specification were prematurely upregulated (Figures 4.7A and 4.7B). Interestingly, as implicated in previous studies by Ahn *et al*, we did not observe enrichment for exocytosis and endocytosis pathways amongst the transcriptionally deregulated sets of genes. These results further strengthen our observations regarding the contrasting function of maternal Satb2 as a transcriptional co-repressor and suggest that downregulation of *satb2* is a requirement for precise mid-blastula transition.

Figure 4.7 Depletion of maternal Satb2 leads to premature activation of genes involved in neurogenesis.



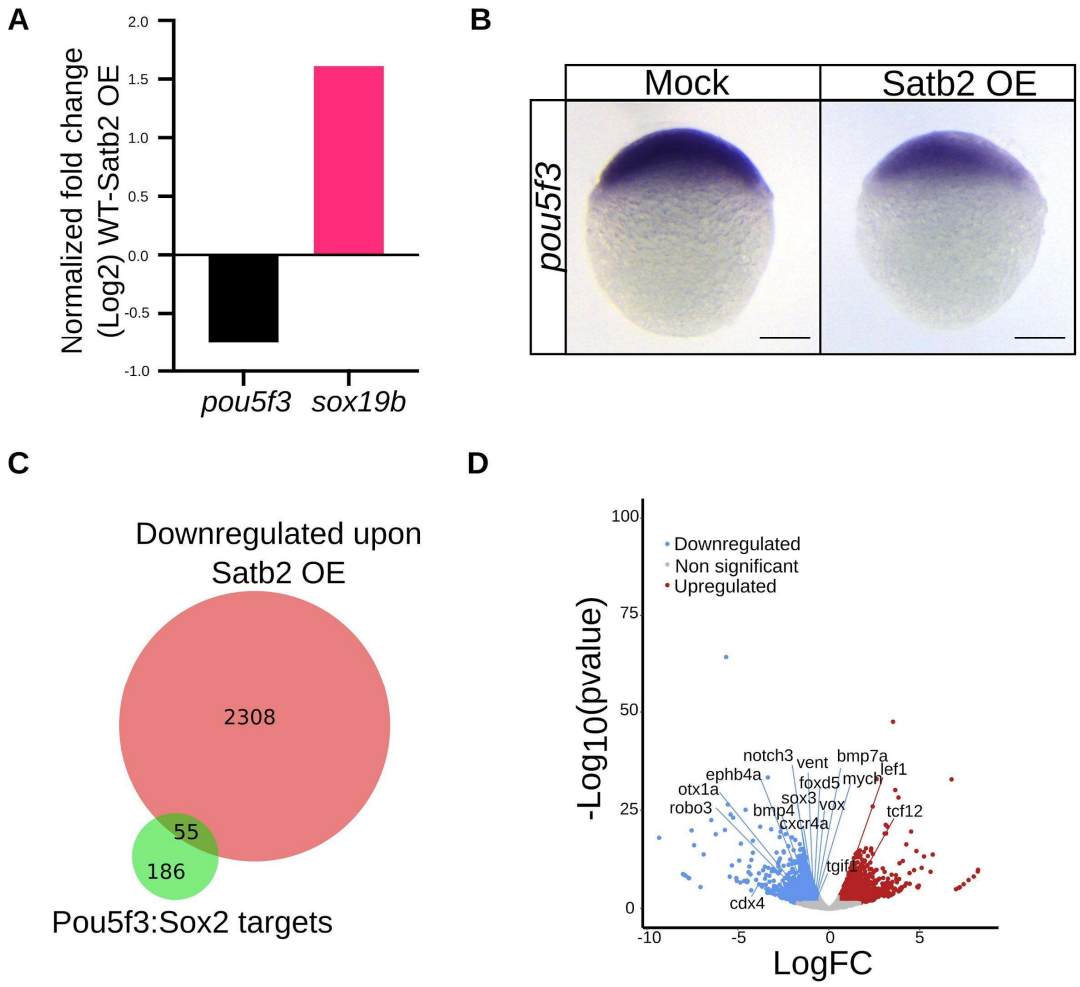
(A) Smear plot analysis for bulk mRNA seq analysis in Control and Satb2 morpholino injected embryos at 4.5 hpf. Values in red and blue represent the number of genes differentially upregulated and downregulated respectively. **(B)** Relative expression values for genes involved in neural crest differentiation upon morpholino mediated depletion of maternal Satb2 (yellow circles) analyzed at 4.5 hpf and in zygotic Satb2$^{-/-}$ mutants (red triangles) analyzed at 14 ss, highlighting the opposite effects on gene regulation by maternal and zygotic forms of Satb2 (FDR<0.1).

4.3.8 Interplay between Satb2 and pioneer factors during ZGA

Role of the pioneer factors such as Pou5f3, Nanog and SoxB1 have been well established in regulating ZGA (Lee et al., 2013; Leichsenring et al., 2013; Veil et al., 2019). Through our transcriptome analysis of Satb2 over-expressing embryos, we observed the downregulation of *pou5f3* (Figure 4.8A). We further validated this using the whole-mount in situ mRNA analysis (Figure 4.8B). Notably, *sox19b* was upregulated upon overexpression of Satb2, suggesting Satb2 differentially regulates pioneer factors presumably based on their functional association (Figure 4.8A).

Interestingly, many of the genomic targets of Pou5f3 and Sox2 showed significant overlap with the genes which are negatively regulated by Satb2 (Figure 4.8C). This gene set includes key important factors involved in fate specification such as *notch3*, *bmp4*, *cxcr4a*, *vox*, *vent*, *foxd5* and *mych* (Figure 4.8D).

Figure 4.8 Maternal Satb2 functions during ZGA by regulating pioneer factors.

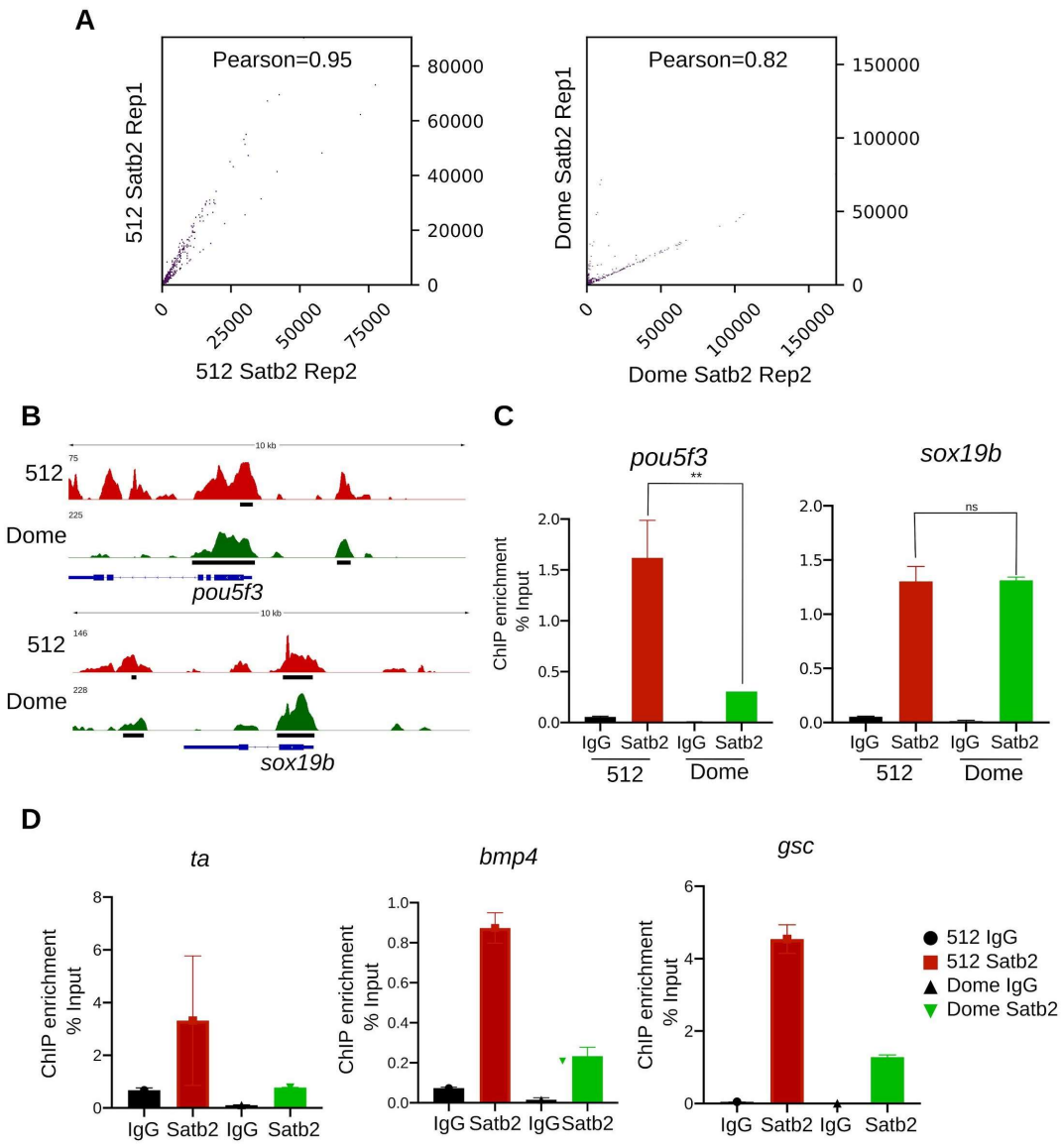


(A) Relative gene expression for *pou5f3* and *sox19b* as determined by RNA-seq analysis (FDR < 0.05). **(B)** Lateral view of zebrafish embryo at 4 hpf for whole-mount in situ hybridization using RNA probe against *pou5f3* in control and Satb2 overexpressing embryos (n=24). **(C)** Venn diagram analysis showing the overlap between negative targets of Satb2 and known genomic targets of Pou5f3 and Sox2. **(D)** Volcano plot of differentially expressed genes upon overexpression of Satb2 analyzed at 4.5 hpf. DE genes with FDR < 0.1 and log2 Fold change > +/- 0.58 are coloured as red for upregulated, blue for downregulated and grey for unchanged or non-significant genes. Key Pou5f3: Sox2 target genes are labelled.

4.3.9 Stage-specific dynamicity in occupancy of Satb2 during MZT ensures transcriptional program of first wave zygotic genes

To explore if the observed negative regulation between Satb2 and pioneer factors and/or first wave zygotic genes in zebrafish is through the direct binding activity of Satb2, we performed ChIPseq analysis at the pre-MBT stage (512 cells) and post MBT stage (dome) (Figure 4.9A). We observed a very strong binding of Satb2 at the TSS of both *pou5f3* and *sox19b* (Figure 4.9B). We further confirmed these binding sites using ChIP-qPCR which also facilitated quantitation of the relative enrichment between Pre-MBT and post-MBT stages. Interestingly, enrichment of Satb2 reduced significantly on the genomic locus of *pou5f3* from pre- to post- MBT but not on the promoter of *sox19b* (Figure 4.9C). Such dynamic occupancy pattern was observed for zygotic genes *ta*, *bmp4* and *gsc* which are negatively regulated by Satb2 (Figure 4.9D). Collectively, these results support that depletion in the expression levels of *satb2* is necessary for sustained expression of the pioneer factor *pou5f3* and in turn the establishment of ZGA.

Figure 4.9 Dynamic occupancy of Satb2 during MBT regulates expression of pioneer factors.



(A) Pearson correlation analysis of ChIP-sequencing datasets obtained using anti-Satb2 antibody at 512 cells stage and the dome stage showing a high level of correlation. **(B)** IGV snapshot of Satb2 ChIP-seq at 512 cells (red) and Dome stage (green) on genomic loci of *pou5f3 and sox19b*. **(C)** Quantitative relative enrichment (ChIP-qPCR) represented as per cent input using isotype-matched IgG and anti-Satb2 antibody for genomic locus

(TSS) of *pou5f3* and *sox19b*. ** indicates P-value < 0.001 as calculated by the Student's two-tailed t-test, N=2. **(D)** Quantitative relative enrichment (ChIP) represented as per cent input using isotype-matched IgG and anti-Satb2 antibody for genomic locus (TSS) of negatively regulated genes *ta*, *bmp4* and *gsc*. ** indicates P-value < 0.001 as calculated by the Student's two-tailed t-test, N=2.

4.3.10 Feedback regulation loop between Satb2 and pioneer factors facilitates the precise onset of ZGA

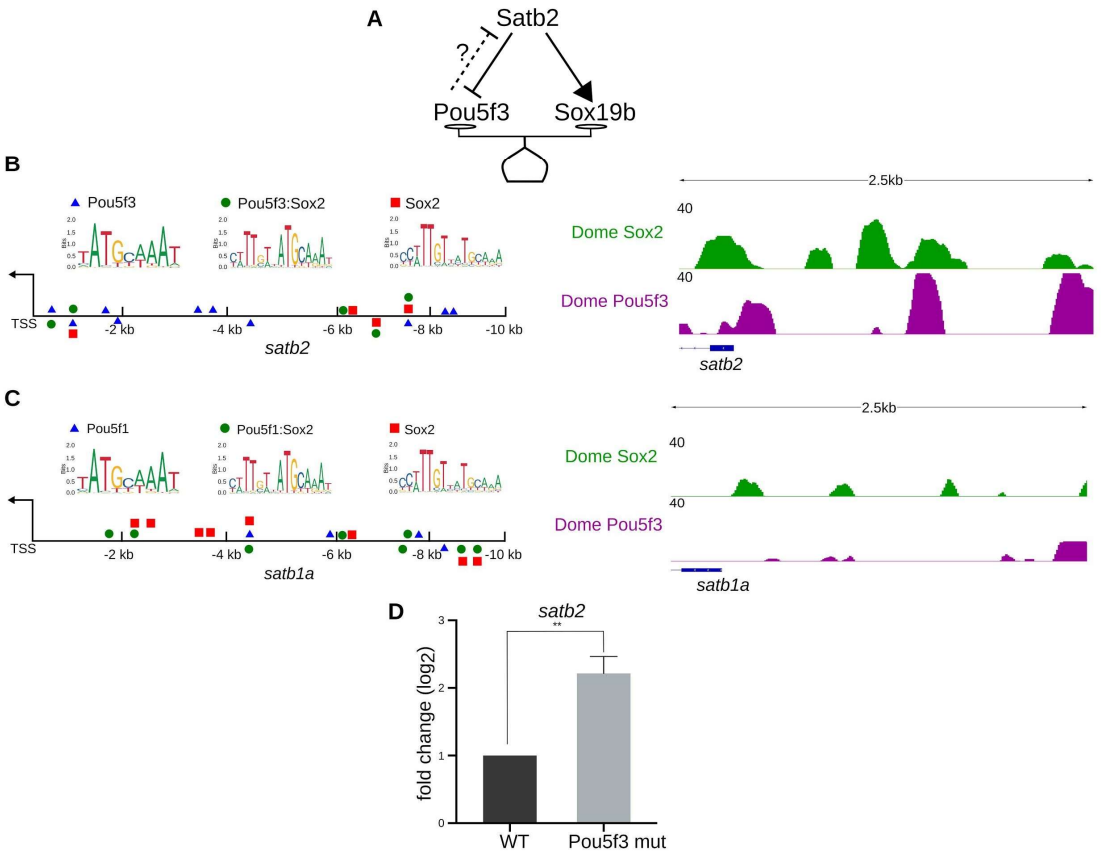
As indicated in the above sections, maternally deposited *satb2* potentially participates in the timely activation of the transcriptional program of zebrafish zygotes. However, the question of how Satb2 is itself regulated during early embryogenesis is intriguing and still unanswered. Characterization of molecular interplays is particularly important since Satb2 levels are downregulated, thus allowing a smooth transition of ZGA. Studies from Ronney et al. have suggested that Bmp and Shh signalling are important for the expression of zygotic Satb2 in the pharyngeal arch, failure of which leads to defects in neural crest differentiation (Sheehan-Rooney et al., 2013). However, it is equally important to identify regulators of maternal Satb2 and their interplay with Satb2 during ZGA. Since the Pou5f3, Nanog and SoxB1 family act as a pioneer factor for ZGA, we asked if there is a negative feedback regulation between these factors and Satb2. A recent study by Gao *et al*. proposed a model suggesting that a balance between Pou5f3 and Sox19b expression is essential during the early stages of embryonic development for the successful activation of ZGA (Gao et al., 2020). This study further suggests that Sox19b independently participates in the neurogenesis program during the late stages of embryogenesis. This also explains in part why Satb2 positively regulates *sox19b*, probably due to its function during neurogenesis. Based on data from the current study and published literature we hypothesized a model which suggests a negative feedback loop between Satb2 and Pou5f3 to maintain the gene expression homeostasis required for ZGA timing (Figure 4.10A). To test this, we first scanned the *satb2* promoter (up to -10 Kb) for binding motifs of Pou5f3:Sox2 complex, Pou5f3 alone and Sox2 alone. Pou5f3 and Sox2 are known to co-occupy target locus as a complex for a regulation. Here, we

used motifs reported in the JASPAR vertebrate system database than one reported in zebrafish literature, allowing us to scan the entire spectrum of binding probability. A number of binding sites were observed for these pioneer factors within a 2 Kb window upstream to TSS (Figure 4.10B). Interestingly, such motifs were absent on the core promoter region of a related protein *satb1a*, thereby highlighting gene-specific regulation by these factors (Figure 4.10C). Analysis of previously published ChIPseq data (Leichsenring et al., 2013) at post-MBT stages confirmed this observation (Figure 4.10B and 4.10C).

Further, to test our hypothesis we utilized previously characterized mutants for Pou5f3 (Burgess et al., 2002). We reasoned that if Pou5f3 indeed has the potential to repress Satb2 then we should observe the upregulation of Satb2 in these mutants. Indeed, the qRT-PCR analysis revealed significant upregulation of *satb2* in Pou5f3 mutants (Figure 4.10D).

Taken together, these results strongly support our model of negative feedback regulation between Satb2 and the pioneer factor Pou5f3 to regulate the timing of ZGA and in the process assign a novel molecular function to Satb2 during early cleavage cycles of embryo development.

Figure 4.10 Negative feedback regulation between Satb2 and Pioneer factors ensures precise onset of ZGA.



(A) The predictive regulatory model between maternal Satb2 and pioneer factors during the establishment of ZGA. **(B)** Schematic for *satb2* promoter displaying motif sites marked for Pouf53, Pou5f3: Sox2 and Sox2. Integrative genome viewer snapshots of Pou5f3 occupancy on genomic loci of zebrafish *satb2* highlighting selective regulation by Pou5f3. **(C)** Schematic for *satb1a* promoter displaying motif sites marked for Pouf53, Pou5f3: Sox2 and Sox2. Integrative genome viewer snapshots of Pou5f3 occupancy on genomic loci of zebrafish *satb1a* highlighting selective regulation by Pou5f3. **(D)** qPCR analysis for *Satb2* expression at 4.5 hpf in *pou5f3* mutant embryos. ** signifies P-value < 0.01 as calculated by the Student's two-tailed t-test, N=3.

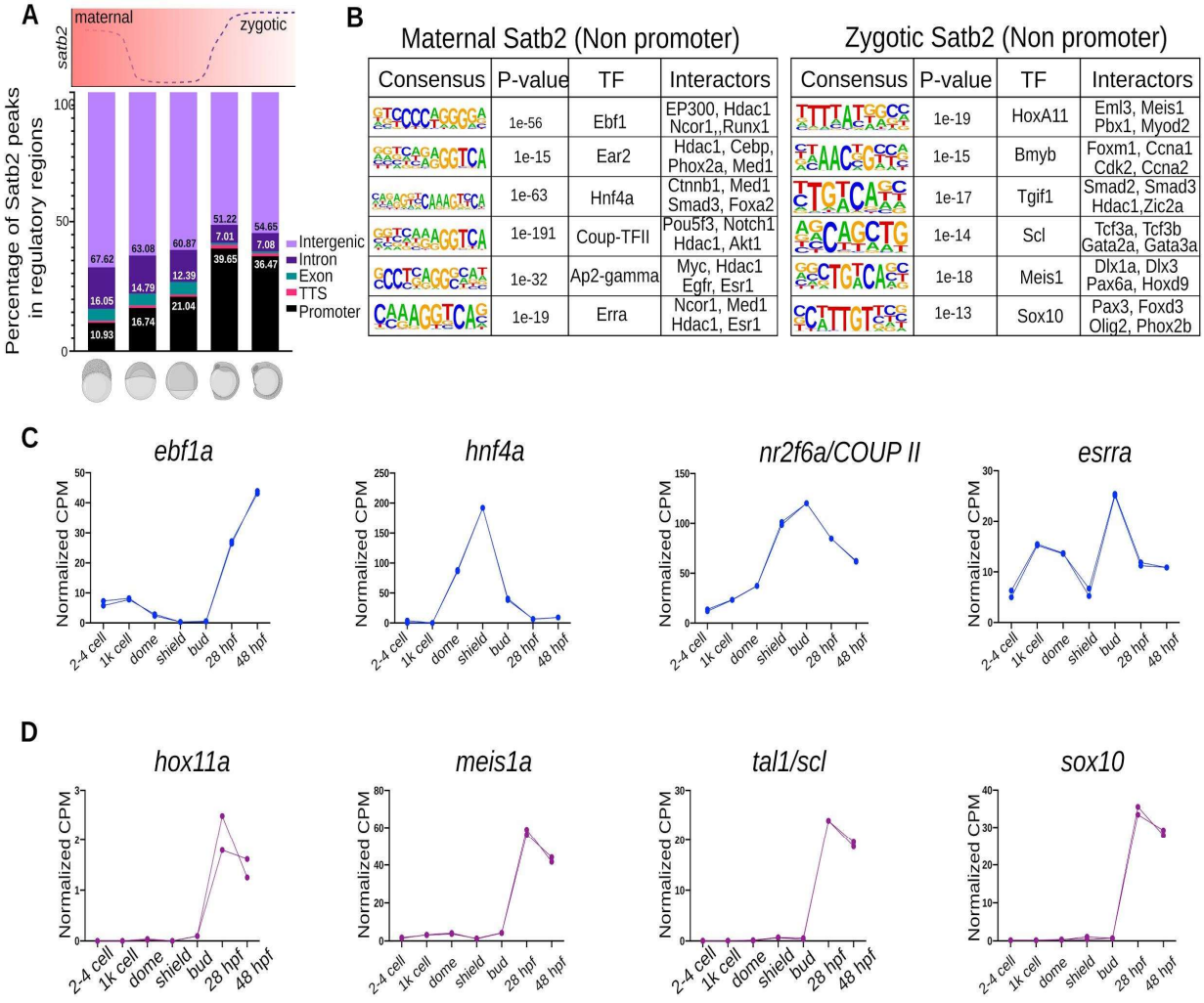
4.3.11 Stage-specific occupancy analysis of Satb2 reveals the underlying mechanisms for the differential activity of maternal and zygotic Satb2

In Chapter 3, we established that zygotic Satb2 promotes neurogenesis and neural crest specification, thus performing its conserved regulatory functions. Results presented in this chapter provide evidence for the uncharacterized function of maternal Satb2 during ZGA through its repressive activity. However, how these differential functions are achieved is unclear. Such disparate behaviour could be attributed to two potential mechanisms: (1) Differential occupancy at the target loci and/or (2) Stage-specific distinct interacting protein partners. To resolve this, we assessed the occupancy profiles of Satb2 throughout early embryogenesis at 512 cells (pre-MBT), dome (post-MBT), 80% epiboly (gastrulation), 6 somites (neural crest specification) and 14 somites (neural crest migrations). Interestingly, we observed a substantial shift in the ratio of non-promoter to promoter-bound regions by Satb2. Maternal Satb2 tends to occupy more of the non-promoter regions, about 67% intergenic and only about 10% promoter at the pre-MBT stage, whereas zygotic Satb2 shows enrichment on more than 36% of promoter regions and about 54% Intergenic regions at 6 somites and 14 somites stage (Figure 4.11A).

Next, we performed consensus binding analysis for non-promoter and promoter-bound regions at 512 cell stage and 14 somite stage to identify the potential interactors of Satb2. Interestingly, our analysis unravelled different sets of consensus binding sites at the intergenic regions (Figure 4.11B). Maternal Satb2 occupied regions were more enriched for binding sites of the transcription factors Ebf1, Ear2 (nr2f6), Hnf4a, Coup-TFII (Nr2f2), Ap2-gamma (Tfap2c) and Erra. Among these, Ebf1, Ear2, Coup-TFII proteins have been widely characterized for repressive activity during B cell and neural crest differentiation (Barske et al., 2018; Hermann-Kleiter et al., 2008; Hsu et al., 2017; Lin et al., 2010; Pereira et al., 1999). In contrast, zygotic Satb2 shows enrichment for the TFs Hoxa11, Bmyb, Tgif1, Scl, Sox10 and Meis1, which are essential factors in promoting neural crest differentiation and neurogenesis (Achim et al., 2013; Bouilloux et al.; Boulet and

Capecchi, 2004; Carney et al., 2006; Gongal and Waskiewicz, 2008; Liu and Greene, 2001; Liu et al., 2004). These findings suggest such differential occupancy and downstream effects could be an outcome of differential interacting partners.

Figure 4.11 Differential mechanisms employed by maternal and zygotic Satb2 to perform contrasting functions during embryogenesis.



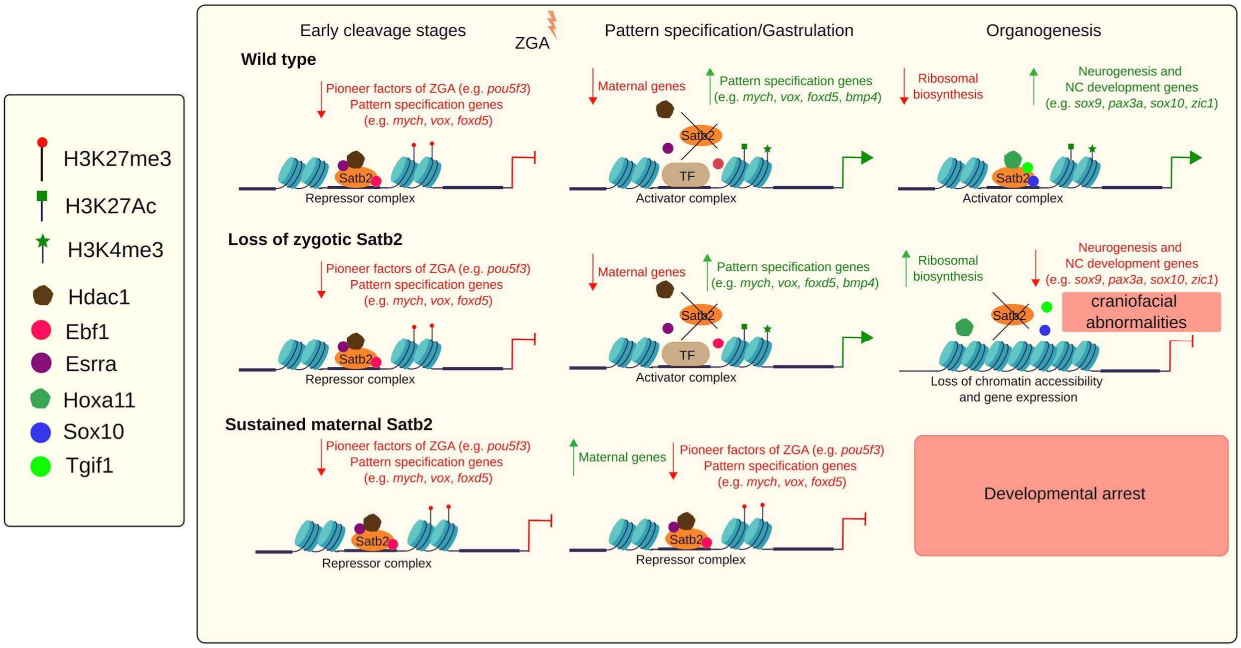
(A) Genomic distribution of Satb2-bound regions across developmental stages highlighting a shift in the ratio of non-promoter to promoter-bound peaks. The Schematic above the bar plot represents the expression pattern of *satb2* throughout early embryogenesis. **(B)** TFBS for transcription factors at maternal

(512 cell stage) and zygotic (14 ss) Satb2. Known Interactors for each transcription factor are enlisted. **(C)** Line graph depicting mRNA expression dynamics for putative interactors of maternal Satb2; *ebf1a*, *hnf4a*, *nr2f6a* and *esrra* across developmental stages as indicated on the x-axis. **(D)** Line graph depicting mRNA expression dynamics for putative interactors of zygotic Satb2 namely, *hox11a*, *meis1a*, *tal1*, and *sox10,* across developmental stages as indicated on the x-axis.

Expression dynamics analysis of these transcription factors highlights putative stage-dependent association with Satb2 (Figures 4.11C and 4.11D). Using the STRING database and published literature, we extracted putative interactors of these proteins. Interestingly, we observed many of the factors enriched in the maternal peak set to be associated with Hdac1, Myc and Med1 proteins, which are known co-repressors (Brenner et al., 2005; Laherty et al., 1997; Malik and Roeder, 2010). However, similar analysis at later stages revealed possible interactions with positive regulators, specifically during neurogenesis such as Pitx1, tcf3a, dlx3a, Tgif1, Olig2 and Zic2a (Beanan and Sargent, 2000; Borodovsky et al., 2009; Chang et al., 2001; Gongal and Waskiewicz, 2008; Kimelman and Griffin, 2000; Sanek et al., 2009). These proteins are not established interactors or targets of Hdac1 (exception of Tgif1). Association between SATB2 and HDAC1 has been well documented for the former's repressive activity in neuronal systems. These studies have indicated that loss of interaction between mouse SATB2 and HDAC1 results in higher expression of Ctip2 and turn the formation of a higher number of cortical connections (Baranek et al., 2012; Gyorgy et al., 2008). Thus, Satb2 may potentially act as a repressor prior to ZGA through interacting with repressor complexes. During later stages of development, Satb2 switches its molecular function to act as an activator of neurogenesis by selecting distinct binding partners.

Together, this analysis argues that the contrasting, stage-dependent functions of Satb2 are a consequence of qualitatively distinct binding profiles, resulting through interaction with differential binding partners (Figure 4.12). However, future experiments will be necessary to test this idea.

Figure 4.12 Schematic summary highlighting differential functions of Satb2 during various stages of zebrafish embryogenesis.



The figure summarizes the differential mechanisms employed by Satb2 to maintain gene expression homeostasis through the early development of zebrafish. During early cleavage stages, Satb2 acts as a repressor to prevent premature activation of the zygotic genome. However, during the early stages of ZGA, Satb2 may interact with multiple specific activators of neural crest specification and neurogenesis to activate these developmental programs. Moreover, changes in gene expression are correlated with changes in histone modifications suggesting active regulation in the presence of Satb2.

4.4 Discussion

Intrigued by the biphasic nature of *satb2* expression, we decided to analyze the function of maternally deposited Satb2 during stages of pluripotency. Analysis of the maternally deposited Satb2's function revealed its novel role as a

transcriptional repressor of zygotic genes. On the other hand, we observed significant upregulation of ribosomal biosynthesis pathways, suggesting Satb2 may regulate core developmental genes at the expense of genes involved in metabolic pathways. This again highlights a new function of Satb2 in regulating metabolic pathways. An amalgamation of perturbation in these multiple pathways resulted in delayed onset of ZGA. Consistent with our observations at 14 somites (Chapter 3), here, we found Satb2 regulates target gene expression through changes in the local epigenomic landscape. Although we observed major changes in histone modification profiles upon overexpression of Satb2 we failed to observe any changes in chromatin accessibility. We observed a significant reduction in the occupancy of H3K27Ac. A study from Chan et al. suggested that H3K27Ac activity is important for transcription competency during ZGA, thereby explaining in part the perturbation of ZGA upon overexpression of Satb2 (Chan et al., 2019).

A recent study indicates that reprogramming of histone modification patterns is essential for the successful onset of ZGA (Zhu et al., 2019). This study further confirms the notion that maternally-deposited H3K27me3 is depleted prior to the onset of ZGA with a concomitant increase in H3K4me3 occupancy (Zhu et al., 2019). Our data suggest that manipulating Satb2 levels prior to ZGA perturbs the ratio of histone modification levels, thereby affecting downstream gene expression networks. Together, our data indicate that during cleavage stages, Satb2 might be functioning more as a co-regulator to bring transient changes than generating long-lasting effects through active chromatin remodelling. Interestingly, the number of accessible regions was also higher in the Satb2 overexpression background probably due to additional sites of Satb2 binding post-MBT due to overexpression.

Pioneering factors such as Pou5f3, SoxB1 and Nanog have been implicated in the regulation of gene expression in a cell-autonomous manner (Gagnon et al., 2018; Messerschmidt et al., 2014). However, signalling molecules such as TGF beta, nodal and Notch signalling has been shown to regulate gene expression in both cell-autonomous as well as in a non-cell-autonomous manner. Our single-cell

analysis of embryos with mosaic overexpression of Satb2 pointed towards a cell-autonomous mode of regulation.

The bimodal activity of Satb2 during major gene regulatory transitions during development poses two major questions. What regulates the expression of Satb2 at different levels during embryonic development? Secondly, how does Satb2 confer differential regulatory activities? BMP and SHH signalling have been studied extensively and found to regulate the expression of *satb2 (Bonilla-Claudio et al., 2012; Sheehan-Rooney et al., 2013)*. However, these signalling pathways are less active during early cleavage cycles suggesting a need to find more upstream regulators of Satb2. Promoter scan of Satb2 revealed binding sites for the pioneer factors Pou5f3 and Sox2. We further confirmed the interplay between Satb2 and pioneer factors that regulate ZGA. A recent study by Gao *et al*. provides novel insights into how the balance between Pou5f3 and Sox19b is required for efficient ZGA (Gao et al., 2020). Here, we observed that Satb2 regulates Pou5f3 and Sox19b in contrasting ways to maintain gene expression homeostasis which is necessary for ZGA timing and we further propose a model for its regulation. Pou5f3 has been implicated in regulating ZGA in zebrafish and humans but surprisingly not in mice (Gao et al., 03 22, 2018; Lee et al., 2013, 2014; Leichsenring et al., 2013). Taken together, these studies suggest that the higher-order GRNs are potentially more conserved between humans and zebrafish than mice. Moreover, our results further indicate that Satb2 may potentially act as a determinant molecular node of species-specific GRNs by regulating one or more of the Yamanaka factors during the process of MBT.

In the quest to answer the second question, stage-dependent ChIP-seq analysis helped us to identify potential mechanisms by which Satb2 confers differential activity. Firstly, the dynamic shift in non-promoter to promoter ratio caught our attention and underlying motif analysis confirmed that Satb2 indeed has the potential to form a complex with multiple partners and regulate gene expression. Interestingly, the choice of partner could be based on the expression dynamics of the respective proteins, since the majority of predicted interactors during

organogenesis express after ZGA. Similarly, Satb2 tends to form a complex with the generic transcriptional repressors prior to ZGA. Satb family proteins are known to employ multiple mechanisms for switching their interacting partners to regulate its targets (Kumar et al., 2006, Baranek et al., 2012). Interestingly, Satb1interacts with either an activator complex or a repressor complex based on its post-translational modifications such as phosphorylation and acetylation. Phosphorylated Satb1 recruits HDAC complex leading to transcriptional repression of target genes. Whereas, acetylated form of Satb1 can no longer bind to regulatory elements and carry out repressive activities (Kumar et al., 2006). Characterizing post-translational modifications of zebrafish Satb2 and its effect on target gene expression during development would be instrumental in advancing the current understanding of the bimodal activity of Satb2. Such bimodal activity has been reported for Foxd3 which is a known determinant of the neural crest differentiation programs. Foxd3 acts as an activator during NC specification events (5-6 somites stage), whereas switches its regulatory role to a repressor by selecting differential interacting partners during NC differentiation and migratory stages (Lukoseviciute et al., 2018).

Taken together, our current study highlights the importance of characterizing the function of genome organizer proteins during multiple developmental processes although they are expressed at low levels as compared to classical transcription factors. Through this study, we have attempted to characterize one such organizer protein and our data collectively assigns Satb2 a novel role as a gate-keeper during major gene-regulatory transitions throughout embryogenesis.

4.5 References

Achim, K., Peltopuro, P., Lahti, L., Tsai, H.-H., Zachariah, A., Astrand, M., Salminen, M., Rowitch, D., and Partanen, J. (2013). The role of Tal2 and Tal1 in the differentiation of midbrain GABAergic neuron precursors. Biol. Open *2*, 990–997.

Ahn, H.-J., Park, Y., Kim, S., Park, H.-C., Seo, S.-K., Yeo, S.-Y., and Geum, D. (2010). The expression profile and function of Satb2 in zebrafish embryonic development. Mol. Cells *30*, 377–382.

Andersen, I.S., Lindeman, L.C., Reiner, A.H., Østrup, O., Aanes, H., Aleström, P., and Collas, P. (2013). Epigenetic marking of the zebrafish developmental program. Curr. Top. Dev. Biol. *104*, 85–112.

Baranek, C., Dittrich, M., Parthasarathy, S., Bonnon, C.G., Britanova, O., Lanshakov, D., Boukhtouche, F., Sommer, J.E., Colmenares, C., Tarabykin, V., et al. (2012). Protooncogene Ski cooperates with the chromatin-remodeling factor Satb2 in specifying callosal neurons. Proc. Natl. Acad. Sci. U. S. A. *109*, 3546–3551.

Barske, L., Rataud, P., Behizad, K., Rio, L.D., Cox, S.G., and Crump, J.G. (2018). Essential role of Nr2f nuclear receptors in patterning the vertebrate upper jaw. Dev. Cell *44*, 337–347.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233–237.

Beanan, M.J., and Sargent, T.D. (2000). Regulation and function of Dlx3 in vertebrate development. Dev. Dyn. *218*, 545–553.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonilla-Claudio, M., Wang, J., Bai, Y., Klysik, E., Selever, J., and Martin, J.F. (2012). Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development. Development *139*, 709–719.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. *42*, 1093–1100.

Borodovsky, N., Ponomaryov, T., Frenkel, S., and Levkowitz, G. (2009). Neural protein Olig2 acts upstream of the transcriptional regulator sim1 to specify diencephalic dopaminergic neurons. Dev. Dyn. *238*, 826–834.

Bouilloux, F., Thireau, J., Venteo, S., Farah, C., Karam, S., Dauvilliers, Y., Valmier, J., Copeland, N.G., Jenkins, N.A., Richard, S. and Marmigere, F. (2016). Loss of the transcription factor Meis1 prevents sympathetic neurons target-field innervation and increases susceptibility to sudden cardiac death. *Elife*, *5*,e11627.

Boulet, A.M., and Capecchi, M.R. (2004). Multiple roles of Hoxa11 and Hoxd11 in the formation of the mammalian forelimb zeugopod. Development *131*, 299–309.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., et al. (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J. *24*, 336–346.

Burgess, S., Reim, G., Chen, W., Hopkins, N., and Brand, M. (2002). The zebrafish spiel-ohne-grenzen (spg) gene encodes the POU domain protein Pou2 related to mammalian Oct4 and is essential for formation of the midbrain and hindbrain, and for pre-gastrula morphogenesis. Development *129*, 905–916.

Carney, T.J., Dutton, K.A., Greenhill, E., Delfino-Machín, M., Dufourcq, P., Blader, P., and Kelsh, R.N. (2006). A direct role for Sox10 in specification of neural crest-derived sensory neurons. Development *133*, 4619–4630.

Chan, S.H., Tang, Y., Miao, L., Darwich-Codore, H., Vejnar, C.E., Beaudoin, J.-D., Musaev, D., Fernandez, J.P., Benitez, M.D.J., Bazzini, A.A., et al. (2019). Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. Dev. Cell *49*, 867–881.e8.

Chang, W., KhosrowShahian, F., Chang, R., and Crawford, M.J. (2001). xPitx1 plays a role in specifying cement gland and head during early Xenopus development. Genesis *29*, 78–90.

Chen, C., and Shen, M.M. (2004). Two Modes by which Lefty Proteins Inhibit Nodal Signaling. Curr. Biol. *14*, 618–624.

Corces, M.R., Ryan Corces, M., Ryan Corces, M., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., et al. (2017). Omni-ATAC-seq: Improved ATAC-seq protocol. Protocol Exchange.

Coyle, D.E., Li, J., and Baccei, M. (2011). Regional Differentiation of Retinoic Acid-Induced Human Pluripotent Embryonic Carcinoma Stem Cell Neurons. PLoS One *6*, e16174.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. *49*, 941–945.

Dobreva, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Fariñas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell *125*, 971–986.

Duan, J.E., Rieder, L.E., Huang, A., and Jordan, W.T. (2020). CLAMP and Zelda function together as pioneer transcription factors to promote Drosophila zygotic genome activation. bioRxiv.

Dworkin, M.B., and Dworkin-Rastl, E. (1990). Functions of maternal mRNA in early development. Mol. Reprod. Dev. *26*, 261–297.

Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat. Rev. Mol. Cell Biol. *19*, 436–450.

Foe, V.E., and Alberts, B.M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J. Cell Sci. *61*, 31–70.

Gagnon, J.A., Obbad, K., and Schier, A.F. (2018). The primary role of zebrafish nanog is in extra-embryonic tissue. Development *145*.

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., et al. (03 22, 2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. Cell *173*, 248–259.

Gao, M., Veil, M., Rosenblatt, M., Gebhard, A., and Hass, H. (2020). Pou5f3 and Sox19b select gene expression repertoire at Zygotic Genome Activation. bioRxiv.

Gaskill, M., Gibson, T.J., Larson, E.D., and Harrison, M. (2020). The pioneer factor GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. bioRxiv.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Gongal, P.A., and Waskiewicz, A.J. (2008). Zebrafish model of holoprosencephaly demonstrates a key role for TGIF in regulating retinoic acid metabolism. Hum. Mol. Genet. *17*, 525–538.

Gyorgy, A.B., Szemes, M., de Juan Romero, C., Tarabykin, V., and Agoston, D.V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. Eur. J. Neurosci. *27*, 865–873.

Hermann-Kleiter, N., Gruber, T., Lutz-Nicoladoni, C., Thuille, N., Fresser, F., Labi, V., Schiefermeier, N., Warnecke, M., Huber, L., Villunger, A., et al. (2008). The Nuclear Orphan Receptor NR2F6 Suppresses Lymphocyte Activation and T Helper 17-Dependent Autoimmunity. Immunity *29*, 205–216.

Hong, S.-K., Jang, M.K., Brown, J.L., McBride, A.A., and Feldman, B. (2011). Embryonic mesoderm and endoderm induction requires the actions of non-embryonic Nodal-related ligands and Mxtx2. Development *138*, 787–795.

Howe, J.A., Howell, M., Hunt, T., and Newport, J.W. (1995). Identification of a developmental timer regulating the stability of embryonic cyclin A and a new somatic A-type cyclin at gastrulation. Genes Dev. *9*, 1164–1176.

Hsu, W.-H., Chen, C.-M., and You, L.-R. (2017). COUP-TFII is required for morphogenesis of the neural crest-derived tympanic ring. Sci. Rep. *7*, 12386.

Jessop, P., Ruzov, A., and Gering, M. (2018). Developmental Functions of the Dynamic DNA Methylome and Hydroxymethylome in the Mouse and Zebrafish:

Similarities and Differences. Front. Cell Dev. Biol. *6*, 27.

Joseph, S.R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A. and Vastenhouw, N.L.(2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *Elife*, *6*, p.e23326.

Kane, D.A., and Kimmel, C.B. (1993). The zebrafish midblastula transition. Development *119*, 447–456.

Kane, D.A., Hammerschmidt, M., Mullins, M.C., Maischein, H.M., Brand, M., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Haffter, P., Heisenberg, C.P., et al. (1996). The zebrafish epiboly mutants. Development *123*, 47–55.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.

Kimelman, D., and Griffin, K.J. (2000). Vertebrate mesendoderm induction and patterning. Curr. Opin. Genet. Dev. *10*, 350–356.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. *203*, 253–310.

Pavan Kumar, P., Purbey, P.K., Sinha, C.K., Notani, D., Limaye, A., Jayani, R.S., and Galande, S. (2006). Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. Mol. Cell *22*, 231–243.

Laherty, C.D., Yang, W.M., Sun, J.M., Davie, J.R., Seto, E., and Eisenman, R.N. (1997). Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. Cell *89*, 349–356.

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

Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature *503*, 360–364.

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

Leichsenring, M., Maes, J., Mossner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 Transcription Factor Controls Zygotic Gene Activation In Vertebrates. Science *341*, 1005–1009.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with

Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, Y., Wang, R., Qiao, N., Peng, G., Zhang, K., Tang, K., Han, J.-D.J., and Jing, N. (2017). Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. J. Biol. Chem. *292*, 19590–19604.

Li, Y., Wu, B., Liu, H., Gao, Y., Yang, C., Chen, X., Zhang, J., Chen, Y., Gu, Y., Li, J., et al. (2018). Structural basis for multiple gene regulation by human DUX4. Biochem. Biophys. Res. Commun. *505*, 1161–1167.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400–403.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J., et al. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat. Immunol. *11*, 635–643.

Liu, D.X., and Greene, L.A. (2001). Regulation of neuronal survival and death by E2F-dependent gene repression and derepression. Neuron *32*, 425–438.

Liu, D.X., Biswas, S.C., and Greene, L.A. (2004). B-myb and C-myb play required roles in neuronal apoptosis evoked by nerve growth factor deprivation and DNA damage. J. Neurosci. *24*, 8720–8725.

Liu, J., Zhu, Y., Luo, G.-Z., Wang, X., Yue, Y., Wang, X., Zong, X., Chen, K., Yin, H., Fu, Y., et al. (2016). Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat. Commun. *7*, 13052.

Lu, X., Li, J.M., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2009). Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development *136*, 2101–2110.

Lukoseviciute, M., Gavriouchkina, D., Williams, R.M., Hochgreb-Hagele, T., Senanayake, U., Chong-Morrison, V., Thongjuea, S., Repapi, E., Mead, A., and Sauka-Spengler, T. (2018). From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. Dev. Cell *47*, 608–628.e6.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat. Rev. Genet. *11*, 761–772.

McGowan, R.A., and Martin, C.C. (1997). DNA methylation and genome imprinting in the zebrafish, Danio rerio: some evolutionary ramifications. Biochem. Cell Biol. *75*, 499–506.

Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. *28*, 812–828.

Miao, L., Tang, Y., Bonneau, A.R., Chan, S.H., and Kojima, M.L. (2020). Synergistic activity of Nanog, Pou5f3, and Sox19b establishes chromatin accessibility and developmental competence in a context-dependent manner. bioRxiv.

Newport, J., and Kirschner, M. (1982). A major developmental transition in early xenopus embryos: II. control of the onset of transcription. Cell *30*, 687–696.

Oron, E., and Ivanova, N. (2012). Cell fate regulation in early mammalian development. Phys. Biol. *9*, 045002.

Pálfy, M., Schulze, G., Valen, E., and Vastenhouw, N.L. (2020). Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. PLoS Genet. *16*, e1008546.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. *22*, 577–591.

Pereira, F.A., Qiu, Y., Zhou, G., Tsai, M.-J., and Tsai, S.Y. (1999). The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. Genes Dev. *13*, 1037–1049.

Popov, I.K., Kwon, T., Crossman, D.K., Crowley, M.R., Wallingford, J.B., and Chang, C. (2017). Identification of new regulators of embryonic patterning and morphogenesis in Xenopus gastrulae by RNA sequencing. Dev. Biol. *426*, 429–441.

Potok, M.E., Nix, D.A., Parnell, T.J., and Cairns, B.R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. Cell *153*, 759–772.

Prioleau, M.N., Huet, J., Sentenac, A., and Méchali, M. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell *77*, 439–449.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007-7). g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. *35*, W193–W200.

Sanek, N.A., Taylor, A.A., Nyholm, M.K., and Grinblat, Y. (2009). Zebrafish zic2a patterns the forebrain through modulation of Hedgehog-activated gene expression. Development *136*, 3791–3800.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. Genes Dev. *23*, 2625–2638.

Schier, A.F., and Talbot, W.S. (2005). Molecular genetics of axis formation in zebrafish. Annu. Rev. Genet. *39*, 561–613.

Schulz, K.N., Bondra, E.R., Moshe, A., Villalta, J.E., Lieb, J.D., Kaplan, T., McKay, D.J., and Harrison, M.M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. Genome Res. *25*, 1715–1726.

Sheehan-Rooney, K., Swartz, M.E., Lovely, C.B., Dixon, M.J. and Eberhart, J.K.(2013). Bmp and Shh signaling mediate the expression of satb2 in the pharyngeal arches. *PloS one*, *8*(3), p.e59533.

Sheehan-Rooney, K., Swartz, M.E., Lovely, C.B., Dixon, M.J., and Eberhart, J.K. (2013). Bmp and Shh Signaling Mediate the Expression of satb2 in the Pharyngeal Arches. PLoS One *8*.

Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J., and Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res. *25*, 1703–1714.

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat. Protoc. *3*, 59–69.

Veil, M., Yampolsky, L.Y., Grüning, B. and Onichtchouk, D.(2019). Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. *Genome Res. 29*(3), 383-395.

Wang, L., Liu, Z., Lin, H., Ma, D., Tao, Q., and Liu, F. (2017). Epigenetic regulation of left-right asymmetry by DNA methylation. EMBO J. *36*, 2987–2997.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). 4th ed., Univ. of Oregon Press, Eugene.

Westerfield, M. (2007) The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 5th ed. Univ. of Oregon Press, Eugene.

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

49, 935–940.

Wiekowski, M., Miranda, M., and DePamphilis, M.L. (1993). Requirements for promoter activity in mouse oocytes and embryos distinguish paternal pronuclei from maternal and zygotic nuclei. Dev. Biol. *159*, 366–378.

Zhang, R., Liu, L., Yao, Y., Fei, F., Wang, F., Yang, Q., Gui, Y., and Wang, X. (2017). High Resolution Imaging of DNA Methylation Dynamics using a Zebrafish Reporter. Sci. Rep. *7*, 5430.

Zhu, W., Xu, X., Wang, X., and Liu, J. (2019). Reprogramming histone modification patterns to coordinate gene expression in early zebrafish embryos. BMC Genomics *20*, 248.

Zinski, J., Tajer, B. and Mullins, M.C. (2018). TGF-β family signaling in early vertebrate development. Cold Spring Harb. Perspect. Biol. 10(6), a033274.

List of Publications

Published Manuscripts

1. Saurabh J. Pradhan, Reddy Puli Chandramouli, Michael Smutny, Ankita Sharma, Keisuke Sako, Meghana S. Oak, Rini Shah, Mrinmoy Pal, Ojas Deshpande, Yin Tang, Rakesh Mishra, Girish Deshpande, Antonio J. Giraldez, Mahendra Sonawane, Carl-Philipp Heisenberg, and Sanjeev Galande. 2020. "Satb2 Acts as a Gatekeeper for Major Developmental Transitions during Early Vertebrate Embryogenesis." *Cold Spring Harbor Laboratory*. *BioRxiv.* https://doi.org/10.1101/2020.11.23.394171

2. Reddy, Puli Chandramouli, Saurabh J. Pradhan, Krishanpal Karmodiya, and Sanjeev Galande. 2020. "Origin of RNA Polymerase II Pause in Eumetazoans: Insights from Hydra." *Journal of Biosciences* 45.

3. Reddy, Puli Chandramouli, Akhila Gungi, Suyog Ubhe, Saurabh J. Pradhan, Amol Kolte, and Sanjeev Galande. 2019. "Molecular Signature of an Ancient Organizer Regulated by Wnt/β-Catenin Signalling during Primary Body Axis Patterning in Hydra." *Communications Biology* 2 (November): 434.

4. Balyan, Renu, Rupali Gund, Amanpreet Singh Chawla, Satyajeet P. Khare, Saurabh J. Pradhan, Sanket Rane, Sanjeev Galande, Jeannine Marie Durdik, Anna George, Vineeta Bal and Satyajit Rath. 2019. "Correlation of Cell-Surface CD8 Levels with Function, Phenotype and Transcriptome of Naive CD8 T Cells." *Immunology* 156 (4): 384–401.

5. Mir, Rafeeq, Ankita Sharma, Saurabh J. Pradhan, and Sanjeev Galande. 2018. "Regulation of Transcription Factor SP1 by the β-Catenin Destruction Complex Modulates Wnt Response." *Molecular and Cellular Biology* 38

(22). https://doi.org/10.1128/MCB.00188-18.

6. Muralidharan, Bhavana, Marc Keruzore, Saurabh J. Pradhan, Basabdatta Roy, Ashwin S. Shetty, Veena Kinare, Leora D'Souza, Upasana Maheshwari, Krishanpal Karmodiya, Agasthya Suresh, Sanjeev Galande, Eric J Bellefroid and Shubha Tole. 2017. "Dmrt5, a Novel Neurogenic Factor, Reciprocally Regulates Lhx2 to Control the Neuron-Glia Cell-Fate Switch in the Developing Hippocampus." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37 (46): 11245–54.

7. Barone, Vanessa, Moritz Lang, S. F. Gabriel Krens, Saurabh J. Pradhan, Shayan Shamipour, Keisuke Sako, Mateusz Sikora, Călin C. Guet, and Carl-Philipp Heisenberg. 2017. "An Effective Feedback Loop between Cell-Cell Contact Duration and Morphogen Signaling Determines Cell Fate." *Developmental Cell* 43 (2): 198–211.e12.

8. Muralidharan, Bhavana, Zeba Khatri, Upasana Maheshwari, Ritika Gupta, Basabdatta Roy, Saurabh J. Pradhan, Krishanpal Karmodiya, Hari Padmanabhan, Ashwin S Shetty, Chinthapalli Balaji, Ullas Kolthur-Seetharam, Jeffrey D Macklis, Sanjeev Galande and Shubha Tole. 2017. "LHX2 Interacts with the NuRD Complex and Regulates Cortical Neuron Subtype Determinants Fezf2 and Sox11." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37 (1): 194–203.

9. Sako, Keisuke, Saurabh J. Pradhan, Vanessa Barone, Álvaro Inglés-Prieto, Patrick Müller, Verena Ruprecht, Daniel Čapek, Sanjeev Galande, Harald Janovjak, and Carl-Philipp Heisenberg. 2016. "Optogenetic Control of Nodal Signaling Reveals a Temporal Pattern of Nodal Signaling Regulating Cell Fate Specification during Gastrulation." *Cell Reports* 16 (3): 866–77.

Manuscript under preparation

1. Saurabh J. Pradhan, Ankita Sharma, Vanessa Barone, Greg D'silva, Reddy Puli Chandramouli, Keisuke Sako, Satyajeet Khare, Daniel Capek, Sanjeev Galande and Carl-Philipp Heisenberg. "Dynamic chromatin remodelling determines mesendoderm segregation during vertebrate gastrulation."