GSBS Dissertations and Theses   Graduate School of Biomedical Sciences

2019-08-06

The TALE Factors and Nuclear Factor Y Cooperate to Drive Transcription at Zygotic Genome Activation

William J. Stanney III
University of Massachusetts Medical School

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

Part of the Biochemistry Commons, Developmental Biology Commons, and the Molecular Biology Commons

THE TALE FACTORS AND NUCLEAR FACTOR Y COOPERATE TO DRIVE TRANSCRIPTION AT ZYGOTIC GENOME ACTIVATION

A Dissertation Presented

By

WILLIAM J. STANNEY III

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 6, 2019

Program in Biochemistry and Molecular Pharmacology

THE TALE FACTORS AND NUCLEAR FACTOR Y COOPERATE TO DRIVE TRANSCRIPTION AT ZYGOTIC GENOME ACTIVATION

A Dissertation Presented

By

WILLIAM J. STANNEY III

This work was undertaken in the Graduate School of Biomedical Sciences
Program in Biochemistry and Molecular Pharmacology

Under the mentorship of

_____
Charles Sagerström, Ph.D., Thesis Advisor

_____
Thomas Fazzio, Ph.D., Member of Committee

_____
Sean Ryder, Ph.D., Member of Committee

_____
Anthony Imbalzano, Ph.D., Member of Committee

_____
Jesse Mager, Ph.D., External Member of Committee

_____
Scot Wolfe, Ph.D., Chair of Committee

_____
Mary Ellen Lane, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

August 6, 2019

DEDICATION

I dedicate this work to my wife, Samantha. Graduate school is a tremendous challenge, and without your unwavering support, constant encouragement, and unconditional love I would never have been able to achieve this.

I also dedicate this work to my parents, William and Barbara, who did more than I can say to provide me with every opportunity in life. You never hesitated to put my interests ahead of your own, and for that I am truly grateful.

ACKNOWLEDGEMENTS

I would first like to thank Charles for his guidance and mentorship throughout my research project. Graduate school teaches more skills than just science, and I was truly fortunate to be able to learn from his example. Charles fostered a productive and supportive lab environment in which I learned creating thinking, perseverance, and how to be a well-rounded scientist. I am grateful for his patience when facing the challenges that come with science, his insight when making sense of data, and his ability to remain calm and concise. His encouragement and support of everything from my work to my career goals has been indispensable.

I would also like to thank my fellow lab members. Without their support I could never have overcome the challenges of graduate school. I always had someone in lab I could reach out to, whether I needed to talk through a problem to make sense of it, insight into how to approach an issue, guidance on how to set up an experiment or use a new machine, or commiseration when I was frustrated. Franck, you are a spectacular role model, and I am truly grateful for your mentorship and patience. In addition to the techniques you taught me, I learned a great deal about working through problems from you. Jen, Priya, and Özge, I could never have handled the rigors of graduate school without you. From teaching me techniques, to helping me make sense of data, to lending an ear when I needed to vent, you were always there, and for that I am thankful.

To my TRAC members, Dr. Wolfe, Dr. Fazzio, Dr. Bach, and Dr. Ryder, I am extremely grateful for your guidance and feedback throughout my research project. Your insight was invaluable, and your suggestions helped me more times than I can say. I appreciate your willingness to take the time from your own busy schedules to discuss my project all these years. I am also grateful to Dr. Imbalzano for stepping in as a part of my DEC and Dr. Mager for taking the time to come all the way to Worcester as my external committee member.

I would also like to thank Dr. Craig Mello, who paid my tuition when I worked as a technician in his lab so that I could enroll in the Core Course at UMass and get a head start, as well as the Wolfpack, who helped me study, understand difficult concepts, and stay diligent during Core.

Finally, I would like to thank my family for their unwavering support and love. To my wife, Samantha, thank you for always being willing to help me in any way I may have needed; from problem solving to listening to encouragement, I could not have done this without you. To my parents, William and Barbara, thank you for the opportunity you have afforded me in life and support to make it through the tough times. To my sister, Christine, and in-laws, Terry and Robin, thank you for your support and encouragement. None of this would have been possible without the unwavering support and unconditional love from each of you.

ABSTRACT

The TALE factors, comprising the *pbx* and *prep*/*meis* gene families, are transcription factors (TFs) vital to the proper formation of anterior anatomical structures during embryonic development. Although best understood as essential cofactors for tissue-specific TFs such as the *hox* genes during segmentation, the TALE factors also form complexes with nuclear factor Y (NFY) in the early zygote. In zebrafish, Pbx4, Prep1, and NFY are maternally deposited and can access their DNA binding sites in compact chromatin. Our results suggest that TALE/NFY complexes have a unique role in early embryonic development which is distinct from each factor's independent functions at later stages.

To characterize these TALE/NFY complexes, we employed high-throughput transcriptomic and genomic techniques in zebrafish embryos. Using dominant negatives to disrupt the function of each factor, we find that they display similar, but not identical, loss-of-function phenotypes and co-regulate genes involved in transcription regulation and embryonic development. Independently, the TALE factors regulate homeobox genes and NFY governs cilia-related genes. ChIP-seq analysis at zygotic genome activation reveals that the TALE factors occupy DECA sites adjacent to CCAAT boxes near genes expressed early in development and involved with transcription regulation. Finally, DNA elements containing TALE and NFY binding sites drive reporter gene expression in transgenic zebrafish, and disruption of TALE/NFY binding via

mutation or dominant negatives eliminates this expression. Taken together, this data suggests that the TALE factors and NFY cooperate to regulate a set of development and transcription control genes in early zygotic development but also have independent roles after gastrulation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

COPYRIGHTED MATERIAL PRODUCED BY THE AUTHOR

CHAPTER I: INTRODUCTION

Transcription factors are a broad class of proteins that play one of the most fundamental roles in biology. They are essential for the regulation of gene expression and act as one of the major areas of control in nearly every cellular process. One such process is embryonic development. This extraordinary and complicated process, in which a complex organism arises from a single totipotent cell, requires fine-tuned management of timing and myriad factors that work together to ensure that every cell reaches its proper fate. Without the careful coordination of innumerable transcription factors, such a process would simply not be possible.

Maternally deposited factors

Upon fertilization, the zygotic genome is transcriptionally inert. Thus, factors deposited by the male and female gametes must control early zygotic development. The oocyte contributes most of this material, which includes mRNAs and proteins. These factors carry out basic cellular functions related to processes such as metabolism and the cell cycle while also specifying early patterning and cell fates. A key point in early development is the maternal to zygotic transition (MZT), where many maternal factors degrade and the zygotic genome undergoes zygotic genome activation (ZGA) to begin producing on its own gene products (reviewed in (Vastenhouw et al., 2019)).

Although the general principles remain the same across species, many parameters can vary greatly, such as the composition of the maternally deposited material relative to the size of the genome, the duration of its presence prior to degradation, and the amount degraded upon the ZGA. In *Caenorhabditis elegans*, maternally deposited material accounts for roughly 33% of the total genome (Baugh et al., 2003; Stoeckius et al., 2014). The MZT occurs just a few hours after fertilization, at which point about 33% of the maternally deposited material degrades (Baugh et al., 2003). Like in *C. elegans*, the MZT in *Drosophila melanogaster* takes place on the order of hours after fertilization; unlike *C. elegans*, however, maternally deposited material accounts for about 75% of the total genome and about 67% of this material degrades at the MZT (De Renzis et al., 2007; Lécuyer et al., 2007; Tadros et al., 2007). Maternally deposited material in *Danio rerio* (zebrafish), as in *Drosophila*, accounts for about 75% of the total genome (Aanes et al., 2011; Harvey et al., 2013). When MZT begins at roughly 4 hours post-fertilization (hpf), only about 25% of the maternally deposited material degrades (Aanes et al., 2011; Bazzini et al., 2012; Harvey et al., 2013; Mishima & Tomari, 2016). Finally, in *Mus musculus* (mice), the MZT does not take place until days after fertilization, with the maternally deposited material accounting for about 33% of the total genome and about 33% of that material degrading (Hamatani et al., 2004; Wang et al., 2004).

Maternal to zygotic transition

As embryonic development is a remarkably complex process requiring precise activity and timing, the stability and regulation of maternally deposited material is essential to successful progress. Since no new transcripts are being created prior to the ZGA, all regulation of maternally deposited factors must happen post-transcriptionally and post-translationally by other maternal factors. Regulation at the transcriptional level does not begin until after the ZGA, gradually becoming the predominant regulatory mechanism as transcription levels increase in the zygotic genome. One of the most significant mechanisms of regulating maternal material is its degradation. This process is tightly regulated to ensure that maternal materials are present when needed but removed once they have served their purpose (reviewed in (Vastenhouw et al., 2019)).

Maternal factors often must remain stable in oocytes for long periods of time, sometimes as long as years in some organisms. However, it also must be able to degrade rapidly at the appropriate time. Clearance of maternally deposited material rarely happens in one large wave; in most species, degradation of maternal material takes place in several discrete waves. In *Drosophila* and mammals, the first wave of maternal material degradation takes place upon egg activation prior to fertilization. Maternal factors carry out this clearing, as well as additional waves of degradation taking place before the ZGA. Following the ZGA, zygotic factors take over clearing maternal factors

(Bashirullah et al., 2001; Tadros et al., 2007). A diverse set of factors controls the clearance of maternal factors, including RNA-binding proteins (RBPs), RNA modifications, small non-coding RNAs, and codon optimality.

RBPs are one of the most versatile and widespread forms of mRNA stability regulation. In zebrafish, there are several examples of RBPs that control maternal mRNA stability and regulation. To prevent maternal mRNA expression in oocytes, the Y-box protein Ybx1 represses their translation (Sun et al., 2018). To destabilize maternal mRNAs and incite their degradation at the MZT, hnRNPA1 binds an AGGGA motif in the 3' UTR of many maternal transcripts (Despic et al., 2017). AU-rich element proteins play a similar role and work in a similar manner to hnRNPA1 across many species, including *C. elegans* (D'Agostino, Merritt, Chen, Seydoux, & Subramaniam, 2006; Gallo, Munro, Rasoloson, Merritt, & Seydoux, 2008), zebrafish (Rabani et al., 2017), *Xenopus laevis* (Graindorge, Thuret, Pollet, Osborne, & Audic, 2006; Voeltz & Steitz, 1998), and mice (Ramos et al., 2004). RNA modifications can also work with RBPs to destabilize maternal mRNAs. In zebrafish embryos, roughly 33% of maternal mRNAs contain an N^6-adenosine methylation modification, which causes a structural shift in the mRNA and exposes key RBP sites (Liu et al., 2015; Zhao et al., 2017). Additionally, clearance of nearly 50% of maternal RNAs early in the MZT and another 25% later requires the RBP YTHDF2, which binds N^6-adenosine methylated RNAs (Ivanova et al., 2017).

Micro RNAs (miRNAs) are small non-coding RNAs that can repress translation of specific mRNAs. Expressed during the early ZGA, the miR-430 family plays an essential role in clearing maternal transcripts from many species, including mammals. In zebrafish, miR-430 is responsible for degrading about 40% of the total maternal transcripts during the MZT via translation blocking and deadenylation (Bazzini et al., 2012). Indeed, poly-A tail (PAT) length is a major determinant of mRNA stability, making it a frequent target of factors responsible for the degradation of maternal transcripts. Transcripts with shorter PATs undergo less translation and degrade sooner. Terminal uridylation of many maternal transcripts increases during the MZT in zebrafish, *Xenopus*, and mice, which rapidly destabilizes mRNAs with short PATs (H. Chang et al., 2018). Finally, codon optimality is a strategy employed by some species such as zebrafish. Here, mRNAs with higher amounts of codons that have fewer conjugate tRNAs degrade preferentially (Mishima & Tomari, 2016).

Zygotic genome activation

Much like the timing of maternal factor clearance, timing of the ZGA varies widely between species. In some species, such as mice and *Homo sapiens* (humans), ZGA takes place before the first cell division (Hamatani et al., 2004; Yan et al., 2013). In other species, it takes place at later cell divisions; for example, in zebrafish it takes place at the sixth cell cycle (Mathavan et al., 2005).

In such organisms, ZGA does not happen at the same moment across the entire embryo. Different species have various means of mitigating this, with zebrafish employing temporal averaging. Temporal averaging occurs during another key developmental state, the mid-blastula transition, in which cell cycles become longer and asynchronous. As ZGA initiates and transcription increases across the cells of the blastula, the longer cell cycles and accumulation of transcription events mitigate any complications stemming from the variable timing of ZGA (Stapel et al., 2017). This arrangement allows zebrafish embryos to develop rapidly while minimizing complications from asynchronous activation of the zygotic genome throughout the blastula.

ZGA is especially well-studied in zebrafish. The initial wave of ZGA in zebrafish sees roughly 25% of the entire genome transcribed, though only about 25% of them are not also present among the maternally deposited factors (Aanes et al., 2011; Harvey et al., 2013; Lee et al., 2013). In many species, one isoform of some genes is maternally deposited and a different isoform is expressed at the ZGA (Aanes et al., 2011). The initial wave of zygotic genes is enriched for transcription factors, and in particular master transcription factors, that activate large gene networks, as well as factors involved in degrading maternal contributions (Lee et al., 2013). Many early transcripts in zebrafish, as well as *Drosophila*, are short and lack introns due to interference from DNA replication as a consequence of rapid cell division; longer transcripts are coincident with the longer cell cycles of the mid-blastula transition (Dalle Nogare et al., 2009).
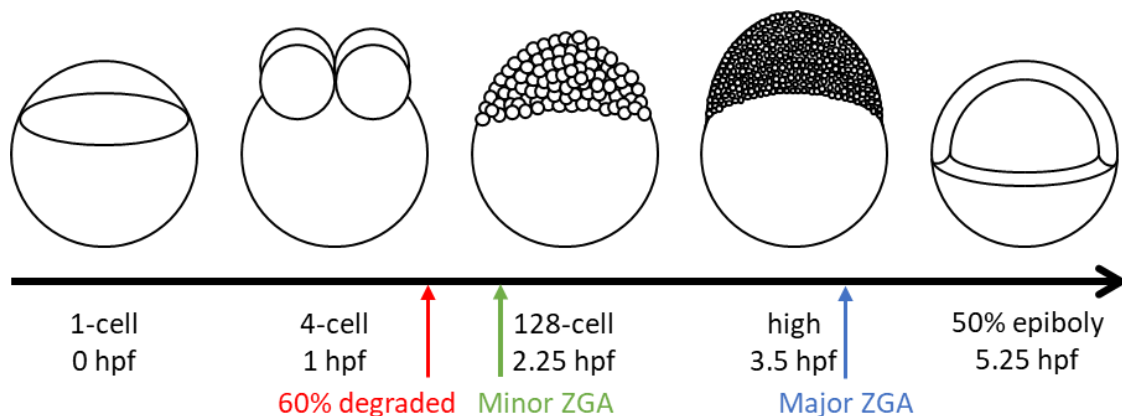
Figure 1.1: The zebrafish maternal to zygotic transition. In zebrafish, degradation of maternal material begins soon after fertilization. By the 16-cell stage at 1.5 hpf, roughly 60% of maternally deposited material is degraded (red arrow). Soon after, at about 2 hpf, a minor wave of zygotic transcription begins (green arrow), in which the zygotic genome begins to produce short, intronless transcripts in gradually increasing amounts. At around 3.7 hpf, coincident with the mid-blastula transition and its lengthened cell cycles, the major wave of zygotic transition begins (blue arrow). Degradation of the remaining 40% of the maternal material requires factors produced by the zygotic genome.

Chromatin

Timing of the ZGA depends on a multitude of circumstances. One component influencing ZGA timing is the access to the zygotic DNA. Access to DNA depends on the structure of its packaging, also known as the chromatin state. Histones are major players in this process. There are four core histones: H2A, H2B, H3, and H4. H2A dimerizes with H2B and H3 dimerizes with H4. Two H2A/H2B dimers combine with two H3/H4 dimers to form an octamer, around which 147 nucleotides of DNA wrap to form a nucleosome. Nucleosomes are the most fundamental unit of chromatin organization, efficiently packaging DNA within the nucleus of every cell (Kornberg, 1977; Tropberger & Schneider, 2013).

Modification of chromatin structure is one of the most commonly employed mechanisms of transcription regulation. This process involved arranging nucleosomes into higher-order structures, ranging from loosely packed open forms to tightly packed closed forms. In more open states, transcription factors can recognize and bind to their target sites on DNA, whereas in closed states those target sites are inaccessible. Thus, controlling the chromatin structure determines the level of access transcription machinery has to the DNA and, consequently, transcription activity. A widely used mechanism for controlling chromatin structure is post-translational modification of the core histones. Various molecular groups, such as acetyl and methyl groups, can be attached to certain amino acids of the core histone proteins. The nature of the moiety and location of its attachment affect the chromatin structure and, by extension, access to DNA binding sites. For example, acetylation of histone H3 lysine 27 (H3K27ac) is a permissive chromatin mark generally associated with active transcription. Trimethylation of the same residue (H3K27me3) is a repressive mark, generally indicating compacted chromatin not accessible for transcription. The same moiety, however, can have different effects when attached to different amino acids; H3K4me3, for example, marks actively transcribed promoters and H3K4me1 marks enhancers. Due to its ability to obfuscate the sequence motifs that many transcription factors rely on to know where to activate transcription, altering chromatin marks to affect chromatin structure is one of the most widely utilized methods of transcription regulation (reviewed in (Gardner et al., 2011)).

Dynamics of zygotic genome activation

Somewhat paradoxically, chromatin within the embryo generally exists in a loosely packed, open confirmation prior to ZGA (Bošković et al., 2014; Laue et al., 2019). While post-translational histone modifications play a more minor role in maintaining a transcriptionally inert zygotic genome, histones themselves are still a key factor in ensuring that ZGA does not begin prematurely. For example, zebrafish and *Xenopus* zygotes contain large numbers of histone proteins (Amodeo et al., 2015; Joseph et al., 2017). As histones can bind DNA with high affinity and little specificity, they easily outcompete transcription factors for DNA binding (Lorch et al., 1987). As the cells divide, DNA quantities grow exponentially, gathering up the excess histones (Jevtić & Levy, 2017; Prioleau et al., 1994). This causes the amount of excess histone proteins to drop. Meanwhile, the amount of transcription factor proteins translated from maternally deposited transcripts accumulate, finally reaching a critical threshold where they are able to compete with histones for DNA binding to initiate ZGA (Joseph et al., 2017; Pálfy at al., 2017; Shindo & Amodeo, 2019). At the same time, the larger number of cells causes a reduction in the number of proteins responsible for nuclear import, which alters the balance of histones in the cytoplasm versus the nucleus. Many histones end up trapped in the cytoplasm, shifting the DNA binding equilibrium even further in favor of the transcription factors within the nucleus to ensure ZGA (Kopito & Elbaum, 2009; Shindo & Amodeo, 2019).

In zebrafish, H3K27ac levels begin to rise preceding ZGA (Zhang et al., 2018). This increase is important for miR-430 expression, which is among the earliest genes expressed. In zebrafish, as well as *Drosophila*, mice, and humans, chromatin begins to compact across the genome following ZGA while local genome accessibility increases (Gao et al., 2018; Lu et al., 2016). In zebrafish, H3K4me3 begins to appear on promoters, indicating the beginning of transcription at those sites (Vastenhouw et al., 2010; Y. Zhang et al., 2014). Around this time, many master transcription and pluripotency factors begin working. One such class of transcription factors are the pluripotency factors, which can revert terminally differentiated cells to a more stem cell-like state. These factors play significant roles in ZGA (Soufi et al., 2012).

Pioneer transcription factors

The pluripotency factors are members of a unique classification of transcription factors known as pioneer transcription factors (PTF). As their name suggests, PTFs are among the first transcription factors to bind a given region of DNA. They can recognize their specific binding sequences in compact chromatin and initiate the transition to a more permissive chromatin environment. As the chromatin opens, binding sites for other transcription factors become available, stabilizing the region in a more open conformation (reviewed in (Mayran & Drouin, 2018; Zaret & Carroll, 2011)).

PTFs employ a variety of strategies for recognizing their binding sites within compacted chromatin. Most PTFs can bind nucleosomes, allowing them to scan for their DNA binding sequence. One such well-studied PTF is Forkhead box A (FOXA), which is involved in hepatocyte differentiation. FOXA contains a winged helix domain that mimics histone linker H1, helping it to dislodge nucleosomes (Cirillo & Zaret, 1999; Iwafuchi-Doi et al., 2016). The PTF PAX7 contains two DNA binding domains, a homeodomain and paired domain, and binds to a composite motif recognized by both of its DNA binding domains for a highly stable interaction (Budry et al., 2012). Once bound, both FOXA and PAX7 recruit a methyltransferase chromatin remodeling complex known as Trithorax (in *Drosophila*)/ COMPASS (in yeast)/ MLL (in mammals), which begins depositing permissive chromatin marks and opening the chromatin. These marks include H3K4me1 and H3K4me2 at target enhancers and H3K4me3 at target promoters (Jozwik et al., 2016; Mayran et al., 2018).

In mammals, the pluripotency factors are OCT4, SOX2, and NANOG; in zebrafish, the orthologous genes are *pou5f3*, *sox19b*, and *nanog*, respectively (Gao et al., 2018; Lee et al., 2013; Leichsenring et al., 2013). Inactivation of Pou5f3, Sox, and Nanog using antisense morpholino oligos (MOs) in zebrafish causes embryos to stall prior to gastrulation. This phenotype is similar to that caused by α-amanitin, which blocks RNA polymerase II activity and, consequently, transcription. Injection of *nanog* and *sox* mRNA can rescue this phenotype. Furthermore, the morphant embryos display widespread loss of

genes generally expressed in the first wave of zygotic transcription. Among these first-wave genes is miR-430, which clears maternally deposited transcripts in zebrafish and *Xenopus* (Lee et al., 2013).

In addition to the pluripotency factors, there are other factors which play a vital role in ZGA in different species. In *Drosophila*, a major player in ZGA is the zinc finger transcription factor Zelda. Similar to the pluripotency factors in vertebrates, Zelda binds specific regions of chromatin called TAGteam elements, maintaining the chromatin in an accessible state and allowing other transcription factors to bind their DNA motifs (Harrison et al., 2011; Liang et al., 2008; Nien et al., 2011). In mammals, the transcription factor DUX (DUX4 in humans) binds and activates many genes activated at ZGA; however, DUX is not maternally deposited. Activation of DUX requires the factors DPPA2 and DPPA4, which bind and activate DUX expression at the minor ZGA. Finally, nuclear factor Y (NFY) is required for ZGA in mice. (De Iaco et al., 2019; De Iaco et al., 2017; Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; Lu et al., 2016).

Enhancers

Precise timing and location of gene activation is critical to proper embryonic development. Enhancers are a key component in ensuring the correct initiation of transcription at target genes and they also play a significant role in increased organism complexity without a corresponding expansion in genome

size. Enhancers are cis-acting DNA elements that are often far from their target genes linearly, relying on the 3D chromosome structure to be near their target promoters. They serve as binding platforms for transcription factor complexes, which typically interact via the Mediator complex with transcription complexes at promoters. The DNA sequence of enhancers builds specific combinations of factors intended to initiate transcription of target genes under specific conditions, allowing for precise control of the timing and location of target gene expression. For example, an enhancer may assemble a transcription complex which awaits the arrival of a specific factor. This ensures that every necessary component is present to avoid delays in gathering the various required factors, but also avoids premature activation by controlling the timing of one factor to trigger the release of the transcription complex (reviewed in (Schoenfelder et al., 2010)).

The unique combination of factors assembled also contributes to increasing complexity in higher species. Despite the increase in complexity, higher organisms lack the expected corresponding increase in genome size. In vertebrates, this is partially explained by the presence of a diverse set of distal enhancers, which vastly outnumber protein coding genes in the same genome. Different enhancers contain different sequences, allowing them to assemble different complexes of proteins- some of which may have different functions based on the other factors present (see Pbx, below, in its interaction with Meis and HDACs). Thus, enhancers are a critical aspect of gene regulation, particularly in higher organisms (reviewed in (Schoenfelder et al., 2010)).

The TALE factors

A key class of transcription factors often found at enhancers in embryonic development is the *t*hree *a*mino acid *l*oop *e*xtension, or TALE, factors. Two families of genes comprise the TALE factors: the *pbx* family, and the *prep*/*meis* family. The TALE factors bind DNA with homeodomains which contain an extra set of three amino acids in the linker between α-helix 1 and α-helix 2, increasing the flexibility of the linker loop and distinguishing them from other homeodomain-containing proteins (Bertolino et al., 1995). This extension forms a hydrophobic pocket essential for interactions with Hox proteins, which are the most widely studied binding partner of the TALE factors (Passner et al., 1999; Piper et al., 1999). Indeed, the TALE factors are vital Hox cofactors, recruiting many Hox proteins to their target enhancers (reviewed in (Ladam & Sagerström, 2014; Selleri et al., 2019).

The Pbx family

The *pbx* gene family is broadly conserved, with orthologs in *C. elegans*, *Drosophila*, zebrafish, and mammals. Vertebrates have four *pbx* genes, while *Drosophila* and *C. elegans* have only one, known as *extradenticle* (*exd*) and *ceh-20*, respectively (Bürglin & Ruvkun, 1992; Bürglin, 1997; Van Auken et al., 2002). The mutation that identified *exd* in *Drosophila* caused similar phenotypes as the

hox genes but did not disrupt Hox expression. The discovery of *pbx* in humans happened independently of *exd* as part of study into a chromosomal translocation implicated in pre-B cell leukemia (Korsmeyer, 1992). Further study of Pbx revealed its essential role in Hox function in several systems and its orthologous nature to *exd* (Piper et al., 1999; Saadaoui et al., 2011; Shen et al., 1997). Initially, *pbx* is maternally deposited and ubiquitously expressed, generally becoming tissue restricted later in development. For example, Pbx4 expression in mice is restricted to the testes, and in zebrafish each of the four *pbx* genes has a distinct region of expression in the body. Functionally, the Pbx proteins are very similar, and in zebrafish can all rescue one another's function. (Capellini et al., 2006, Capellini et al., 2010; Pöpperl et al., 2000; Wagner et al., 2001).

Pbx plays a critical role in early embryogenesis. Loss of Pbx1 function in mice results in death *in utero* as well as gross abnormalities in organ development (Brendolan et al., 2005; Capellini et al., 2006, 2011, 2010, 2008; DiMartino et al., 2001; Ferretti et al., 2011; Grebbin et al., 2016; Hurtado et al., 2015; Kim et al., 2002; Koss et al., 2012; Losa et al., 2018; Manley et al., 2004; McCulley et al., 2018; Selleri et al., 2001; Stankunas et al., 2008; Villaescusa et al., 2016; Vitobello et al., 2011; Welsh et al., 2018). Pbx3-/- mutant mice die shortly after birth due to respiratory failure (Rhee et al., 2004), while Pbx2-/- mice show no obvious defects (Selleri et al., 2004); however, doubling either Pbx2 or Pbx3 mutants with a Pbx1-/- mutant results in earlier death and distinct abnormalities. These include skeletal, craniofacial, and distal limb defects which

are not present in Pbx1-/- mutants alone (Capellini et al., 2006, 2011, 2010, 2008; Ferretti et al., 2011; Golonzhka et al., 2015; Hanley et al., 2016; Koss et al., 2012; Vitobello et al., 2011). This suggests overlapping function of the Pbx family members during embryonic development.

In addition to Hox proteins, Pbx interacts with many other factors. One such class of proteins are histone acetyltransferases (HAT) and histone deacetylases (HDAC), which add and remove acetyl groups from histones, respectively (Choe et al., 2009; Saleh et al., 2000). As acetylated histones are generally associated with open, active chromatin, this suggests that Pbx can participate in the opening of chromatin through the recruitment of HATs such as CBP as well as condensation through the recruitment of HDACs. Pbx also binds N-CoR/SMRT, a transcriptional corepressor that facilitates the recruitment of additional HDACs (Asahara et al., 1999). Pbx is also a cofactor for many other proteins besides Hox, including Engrailed, Pdx1, Emx2, Smads, and MyoD (Bailey et al., 2004; Capellini et al., 2010; Knoepfler et al., 1999; Kobayashi et al., 2003; Peers et al., 1995; Serrano & Maschat, 1998).

Which factors Pbx interacts with depends partially on its different isoforms. Alternative splicing of *pbx* transcripts creates isoforms with similar DNA binding specificities but different recruitment preferences (Milech et al., 2001; Monica et al., 1991; Nourse et al., 1990). For example, Pbx1 has two isoforms: Pbx1a, which contains a C-terminal extension and binds N-CoR/SMRT, and Pbx1b,

which lacks the C-terminal extension and does not bind N-CoR/SMRT (Asahara et al., 1999). Regulation of Pbx itself appears to be at the level of nuclear access. In order to access the nucleus, Pbx proteins often must bind other factors, such as Prep proteins, or undergo phosphorylation by PKA (Abu-Shaar et al., 1999; Aspland & White, 1997; Berthelsen et al., 1999; Kilstrup-Nielsen et al., 1998; Rieckhof et al., 1997; Vlachakis et al., 2001; Waskiewicz et al., 2001).

The Prep/Meis family

The other subgroup comprising the TALE family is the *prep*/*meis* family. Like the *pbx* family, the *prep*/*meis* family is well conserved across species, with vertebrates having two *prep* genes and three *meis* genes (except for zebrafish, which have four *meis* genes) (Imoto et al., 2001; Nakamura et al., 1996; Waskiewicz et al., 2001). Orthologous genes also exist in *C. elegans* and *Drosophila*, known as *unc-62* and *homothorax* (*hth*), respectively (Bürglin, 1997; Rieckhof et al., 1997; Van Auken et al., 2002). Like Exd, Hth disruption causes homeotic phenotypes typically associated with Hox proteins but without any disruption to Hox expression (Rieckhof et al., 1997). Like *pbx*, the discovery of *meis* genes came through investigation of leukemia as proto-oncogenes co-activated with Hox (Moskow et al., 1995; Nakamura at al., 1996). The discovery of *prep* genes came from a different study of transcription factors that regulated the urokinase plasminogen activator gene (Berthelsen et al., 1998). In terms of

expression patterns, *hth* and the *prep* genes are ubiquitously expressed, and in zebrafish Prep1 is maternally deposited (Ferretti et al., 1999; Imoto et al., 2001; Rieckhof et al., 1997; Waskiewicz et al., 2001). Conversely, the *meis* genes turn on later in development and are tissue restricted (Nakamura et al., 1996; Waskiewicz et al., 2001). Although some research suggests unique functions for specific Prep and Meis proteins, others suggest that the two are generally interchangeable (Choe et al., 2002). In addition to their roles as Hox cofactors, Prep/Meis proteins can also serve as cofactors for other transcription factors such as Engrailed and MyoD (Berkes et al., 2004; Kobayashi et al., 2003).

Like *pbx* mutants, loss of Prep1 or Meis1 function leads to death of mice *in utero*, with hematopoietic, angiogenic, and hypoplastic organ defects. Similarly, loss of Meis2 results in death *in utero* due to hemorrhaging. More selective loss of function experiments, such as hypomorphic mutation of Prep1 and conditional loss of Meis2 function in neural crest cells, display more localized effects. These result in a reduction of Pbx and Meis levels and defects in the cranial and cardiac neural crest cell lineages, respectively (Azcoitia et al., 2005; Cai et al., 2012; Ferretti et al., 2006; Hisa et al., 2004; Machon et al., 2015).

Pbx and Prep/Meis dysfunction causes many human deformities and diseases. As previously mentioned, the discovery of Pbx came in the context of human leukemia, where a (1:19) chromosomal translocation created a chimeric fusion protein containing the Pbx1 homeodomain and E2A transactivation

domain, leading to pre-B acute lymphoblastic leukemias (Nourse et al., 1990).

Indeed, both Pbx and Prep/Meis are implicated in many cancers and oncogenic

capacities, including breast, lung, liver, and ovarian cancers (Morgan et al.,

2017). Other research points to Pbx dysregulation as playing a role in cleft lip

and palate deformities as well as diabetes (Ferretti et al., 2011; Kim et al., 2002;

Losa et al., 2018; Welsh et al., 2018).
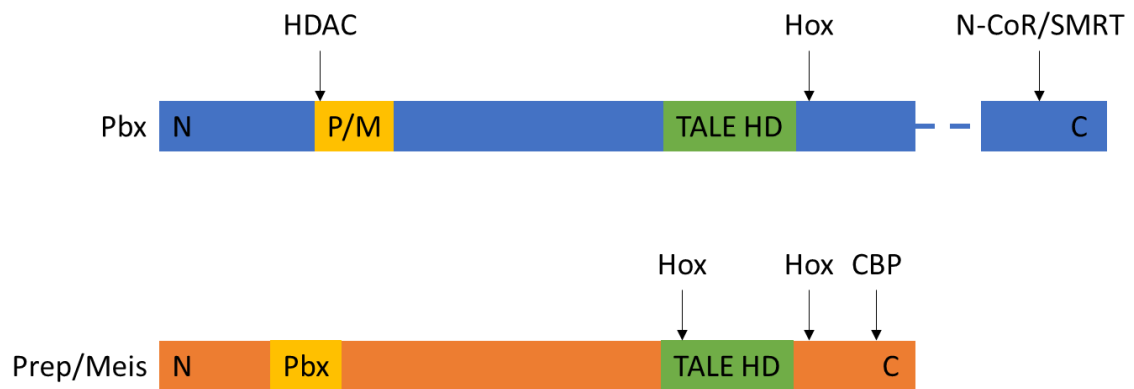
Figure 1.2: Schematics of TALE proteins. Representative Pbx (blue) and Prep/Meis (orange) proteins. N-termini are represented by N and C-termini by C. The homeodomains (HD) are represented in green and the interacting domains in yellow. Other significant factor interacting sites are pointed out above each schematic. In the case of Pbx, some proteins have elongated C-termini, indicated by the dashed line.

TALE transcription factor dynamics

The TALE factors are mostly studied in the context of their role as

cofactors for Hox proteins. Hox can form complexes with just Pbx, just

Prep/Meis, or both Pbx and Prep/Meis proteins (Berthelsen et al., 1998; Choe et

al., 2009; Ferretti et al., 2000; Jacobs et al., 1999; Penkov et al., 2013; Pöpperl et

al., 1995; Ryoo & Mann, 1999; Shen et al., 1997; Vlachakis et al., 2000). By

themselves, Hox proteins have moderate DNA affinity and sequence specificity.

Pbx proteins enhance the affinity of Hox proteins for DNA and also increase the

specificity of the sequences they bind, perhaps by altering the conformation of

the region N-terminal to the Hox homeodomain to a state allowing greater minor

groove contacts (Pöpperl et al., 1995). Indeed, many Pbx binding sites are

characterized as Pbx/Hox half-sites, generally with the sequence TGATNNAT

(Grieder et al., 1997; Ryoo & Mann, 1999). Prep/Meis factors, on the other hand,

appear to have little effect on Hox DNA binding specificity, though they do have

indirect effects that include stabilization and nuclear localization of Pbx proteins

and mediating Hox/Pbx interactions (Hudry et al., 2012; Vlachakis et al., 2001;

Waskiewicz et al., 2001). Although Prep/Meis proteins do not appear to have

direct effects on Hox complex DNA binding, they do alter the composition of Hox

complexes. For example, Meis competes with HDACs for binding to Pbx, helping

to maintain more active chromatin profiles near Hox complex binding sites (Choe

et al., 2009). Prep/Meis proteins bind a DNA motif known as the HEXA motif on

account of its six nucleotides with the sequence TGACAG. HEXA sites are often

located a short distance from Pbx/Hox sites (Amin et al., 2015; Ferretti et al.,

2005; Ferretti et al., 2000; Jacobs et al., 1999; Tümpel et al., 2007).

In addition to the Pbx/Hox hybrid sites and HEXA sites, Pbx and

Prep/Meis proteins can bind the so-called DECA motif, a ten-nucleotide motif

composed of Pbx and Prep/Meis half-sites with the sequence TGATTGACAG.

Originally identified *in vitro* as a binding site for TALE dimers, the DECA motif

also appeared in ESCs and the mouse trunk (Chang et al., 1997; De Kumar et al., 2017; Knoepfler & Kamps, 1997; Laurent et al., 2015; Penkov et al., 2013). Further study of the DECA motif in zebrafish revealed that Pbx4 and Prep1 bound the DECA motif early in development, occupying other binding sites such as Pbx/Hox and HEXA sites later. Pbx4 and Prep1 bound to DECA sites persisted later into development, but their early occupancy relative to that of the more traditional Pbx/Hox and HEXA motifs suggests an alternative function for these motifs (Ladam et al., 2018).

Oftentimes, TALE factors will bind DNA well before recruiting binding partners such as Hox. A prime example is the zebrafish *hoxb1a* gene. In zebrafish, Pbx4 and Prep1 are maternally deposited and bind target enhancer regions within the first few hours post-fertilization. These factors assemble poised transcription complexes by recruiting factors to create a permissive chromatin landscape with modifications such as H3K27ac and H3K4me3 as well as RNA polymerase II. Transcription does not begin, however, until Hoxb1b, the earliest expressed *hox* gene in zebrafish, binds the complex. Once transcription at *hoxb1a* is started, Hoxb1a can drive its own transcription in the fourth rhombomere region of the hindbrain (Choe et al., 2014; Choe et al., 2009).
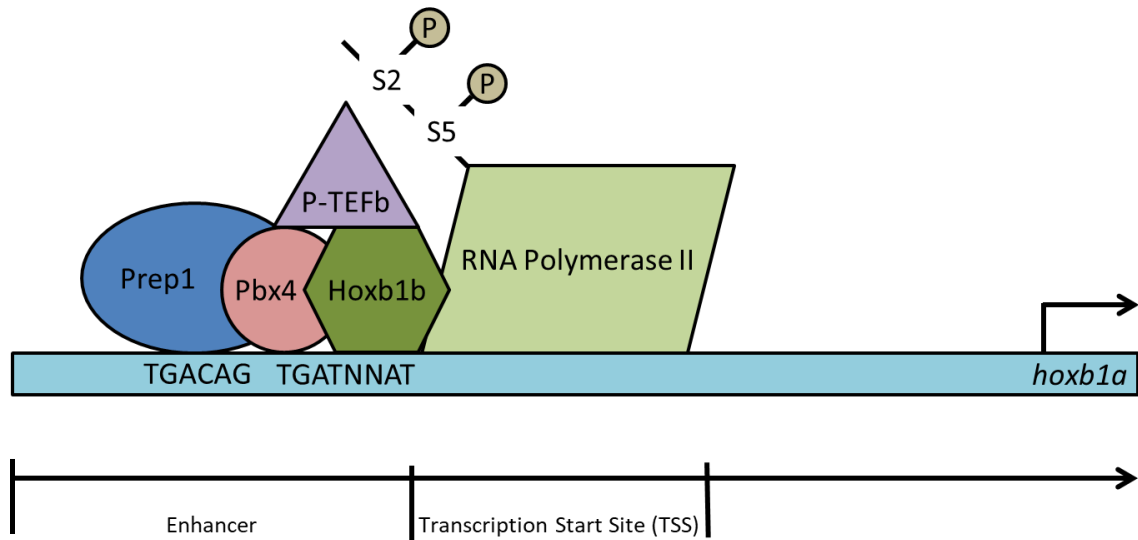
Figure 1.3: Transcriptional activation at *hoxb1a*. Formation of the transcription complex at *hoxb1a* exemplifies the classical model of TALE function. The process begins with the binding of the TALE factors Pbx4 and Prep1 to the *hoxb1a* enhancer, with Prep1 binding the HEXA (TGACAG) motif and Pbx4 binding the TGAT portion of a composite Pbx/Hox motif (TGATNNAT). Pbx4 and Prep1 recruit general transcription factors such as RNA polymerase II and P-TEFb, and the transcription complex remains poised until Hoxb1b binds, initiating transcription.

The CCAAT box and Nuclear Factor Y

Recently, our lab conducted ChIP-seq experiments for Prep1 in zebrafish embryos. These ChIP-seq experiments took place at 3.5 hpf, which is around ZGA, and 12 hpf, which is during segmentation when tissue differentiation is underway. These ChIP-seq experiments revealed that Prep1 was predominantly bound to DECA sites at 3.5 hpf. After this point, Prep1 bound other motifs such as HEXA sites through 12 hpf. DREME, a sequence motif recognition program, identified another sequence motif near these Prep1-bound DECA sites with the sequence CCAAT. This motif, known as the CCAAT box, consistently appeared

within about 20 base pairs (bp) of the DECA sites bound by Prep1, suggesting a potential biological relevance (Ladam et al., 2018).

The CCAAT box is one of the most frequently observed DNA sequence motifs at eukaryotic promoters, found preferentially at the -80 region. For many years the CCAAT box was considered a core promoter element. It correlates well with CpG islands and is present at many well-studied promoters. Advances in high-throughput sequencing technology, however, showed that the CCAAT box is only present in about 30% of eukaryotic promoters. Furthermore, the CCAAT box counter-correlates with other well-known core promoter elements, such as the TATA box. While it is not unheard of for a promoter to have both a CCAAT box and a TATA box, they are uncommon (Dolfini et al., 2009).

A protein complex known as *n*uclear *f*actor *Y* (NFY) recognizes and binds to the CCAAT box. NFY, also referred to as CCAAT Binding Factor (CBF) or CCAAT Protein 1 (CP1), is a heterotrimeric transcription factor comprising the subunits Nfya, Nfyb, and Nfyc. The first description of NFY was as the *n*uclear *f*actor binding the *Y* box at promoters of MHC Class II genes, providing its predominant moniker. At these promoters, NFY interacts with RFX and CREB to form a functional unit capable of recruiting the master MHC Class II coactivator CIITA and drive transcription of genes involved in antigen presentation. To investigate the wider regulatory targets of NFY, gene ontology (GO) analysis of the genes associated with promoters containing a CCAAT box revealed an

enrichment for basic cellular functions. Cell cycle ontologies were the top result, with others including DNA metabolism, chromatin, and endoplasmic reticulum. At these promoters, NFY is a key factor, playing a significant role in the recruitment of RNA polymerase II and general transcription factors (Dolfini et al., 2009).

As with the CCAAT box, advances in high-throughput sequencing provided new insight into NFY function. A ChIP-seq experiment in mouse embryonic stem cells (ESCs) revealed that only about half of NFY peaks were present within 500 bp of a transcription start site, demonstrating that NFY binding was not limited to promoters. Further study on the surrounding characteristics of the distal peaks suggested that they were bound to enhancers. GO analysis of these distal peaks revealed that they were enriched for functions related to ESC maintenance and neuronal function. Indeed, NFY is present in all transformed and immortalized cell lines but absent in many terminally differentiated cells such as myocytes, circulating monocytes, and cells undergoing senescence, suggesting a role in maintaining pluripotency (Oldfield et al., 2014).

At the protein level, all three NFY subunits are conserved from yeast through humans. They are maternally deposited and, initially, ubiquitously expressed (reviewed in (Dolfini et al., 2012)). Although there are two isoforms of Nfya and Nfyc, they have no known functional differences (Ceribelli et al., 2009; Chen et al., 2002; Li et al., 1992). In *Saccharomyces cerevisiae* (yeast), the genes HAP2, HAP3, and HAP5 are orthologous to Nfya, Nfyb, and Nfyc,

respectively. Yeast also have a fourth subunit, HAP4, which provides activating

activity to the complex (Forsburg & Guarente, 1989; Hahn et al., 1988; McNabb

et al., 1995; McNabb & Pinto, 2005; Pinkham et al., 1987). In higher eukaryotes,

the activating properties of HAP4 have been incorporated into Nfya and Nfyc in

the form of two large glutamine- and hydrophobics-rich domains. The HAP

complex binds the upstream activating sequence of many cytochrome genes,

making it a master regulator of respiratory metabolism (Buschlen et al., 2003).

NFY in embryonic development

    While the majority of NFY functions relate to basic cell functions, NFY also

plays roles in embryonic development. In *Drosophila*, increasing or decreasing

Nfya expression leads to embryonic death. Furthermore, Nfyc is a target of

Dorsal, which is a vital transcription factor in early embryogenesis. It is involved

in a variety of roles, including activation of mesodermal genes, R7/R8

photoreceptor axonal guidance, and the JNK pathway (Stathopoulos et al., 2002;

Yoshioka et al., 2007). The roles for NFY in development continue into the

planarian *Schmidtea mediterranea*, where male germ cell development relies on

NFY (Wang et al., 2010). Finally, in *C. elegans*, NFY targets a CCAAT box in the

promoter of the *abdominal-B hox* ortholog *egl-5*. Egl-5 controls patterning in the

tail, and mutation of *nfya-1* results in ectopic expression of *egl-5* as well as

myriad developmental abnormalities. Although generally accepted to be a

transcriptional activator, this suggests that NFY may also be able to function as a repressor. Further evidence of a repressive role for NFY is its interaction with MES-2 and MES-6, subunits of the Polycomb repressive complex in worms. (Deng et al., 2007).

In vertebrates, NFY's roles grow even more complex. In zebrafish, Nfyb is maternally deposited but at later stages of development becomes restricted to the head cartilage and notochord. Knockdown by MOs results in smaller heads, sharpened Meckel's cartilage, loss of ceratobranchial cartilage, and enlarged angles of ceratohyal cartilage due to apoptosis of neural crest cells. (Chen et al., 2009). Mutations in zebrafish *nfyc* resulted in eye defects between the outer and inner nuclear layers due to a loss of photoreceptor and breakdownin lamination (Gross et al., 2005). In mice, NFY is essential for embryonic development. *nfya* knockout mice die by e8.5 thanks to rampant apoptosis and defects in s-phase progression (Bhattacharya et al., 2003). NFY expression is relatively high in the inner cell mass of the mouse blastocyst and there is evidence that it is required for maintenance of pluripotency, which may explain this phenotype (Yoshikawa et al., 2006).

Factor	Morpholino	Dominant Negative	Germline
pbx4	Reduced head and eye size, pericardial edema/ swollen pericardium, hindbrain mispatterning (loss of r3) (Ladam et al., 2018; Waskiewicz et al., 2002)	Reduced head and eye size, pericardial edema/ swollen pericardium, hindbrain mispatterning (loss of r3) (Choe et al., 2002)	Hindbrain mispatterning anterior to r4/r5 boundary, loss of Mauthner neurons, loss or reduction of NC-derived cartilages, reduced jaw, loss of aortic arch vasculature, poor circulation, swollen pericardium, thin and weak heart, loss of pectoral fins, death by 6-7 dpf (Pöpperl et al., 2000)
prep1	Hindbrain mispatterning, improper reticulospinal neuron migration (except Mauthner neurons), cell death/ degeneration of neuroectoderm and hindbrain, impaired motor coordination, reduced head and eye size, loss or reduction of NC-derived cartilages due to inability of NC cells to properly differentiate, loss of jaw and swim bladder, pericardial edema, atrophy of pectoral fins, abnormal distribution of melanocytes, death by 6-7 dpf (Deflorian et al., 2004)	N/A	N/A
nfya	N/A	Reduced head and eye size, pericardial edema/ swollen pericardium (Mantovani et al., 1994; Ladam et al., 2018)	N/A
nfyb	Reduced head size, loss of branchial arc, sharpened Meckel's cartilage, loss of ceratobranchial cartilage, enlarged angles of ceratohyal cartilage, apoptosis of NC cells (Chen et al., 2009)	N/A	N/A
nfyc	N/A	N/A	Loss of photoreceptor and breakdownin lamination between outer and inner nuclear layers (Gross et al., 2005)

Table 1.1: Phenotypes caused by TALE and NFY disruption. Summary of phenotypes associated with disruption of key TALE and NFY factors in zebrafish. NC stands for neural crest.

NFY transcription factor dynamics

All three subunits of NFY are required for DNA binding, and all three subunits contact the DNA when bound. In order to form a complex, Nfyb and Nfyc must first associate. This often happens in the cytoplasm, as Nfyc requires Nfyb to enter the nucleus while Nfya can enter on its own (Frontini et al., 2004). Once inside, the Nfyb/Nfyc heterodimer can bind Nfya, which is the least

abundant of the three subunits, to form the complete heterotrimer. Nfya is the subunit that recognizes the CCAAT box and provides the factor's sequence specificity while Nfyb and Nfyc interact with DNA, as well as one another, through their histone fold domains (HFD). In total, there are 41 points of contact between NFY and DNA, making the complex extraordinarily stable (Nardini et al., 2013; Romier et al., 2003).

Sequence-specific recognition of the CCAAT box by Nfya is managed by its α2 helix, which inserts into the minor groove of DNA, making contact via a GxGGRF motif and inducing a bend of about 80° in the DNA. This bend could promote the binding of other transcription factors in the exposed major grooves. The initial DNA contacts of NFY are likely made via the HFDs of the Nfyb/Nfyc heterodimer, allowing Nfya to interact with those subunits via is α1 helix and search for a CCAAT box with its α2 helix (Nardini et al., 2013).

The HFDs of Nfyb and Nfyc are key components to NFY's function. Nfyb and Nfyc are related to the histone proteins H2B and H2A, respectively (Baxevanis et al., 1995). Despite only 15%-18% sequence similarity, the structure of the interaction of Nfyb and Nfyc with the DNA sugar-phosphate backbone mimics that of the H2B/H2A heterodimer as well as several other well-characterized factors such as TATA-binding protein (TBP)-associated factors (TAFs), the TBP/TATA-binding negative cofactor 2 (NC2a/b), and the CHRAC15/CHRAC17 subunits of the nucleosome remodeling complex CHRA

(Kamada et al., 2001; Poot et al., 2000; Romier et al., 2003). Furthermore, NFY can undergo many post-translational modifications, many of which appear to mirror modifications made to histone proteins. For example, H2B-K120 monoubiquitination is one of the earliest chromatin marks associated with active genes. Nfyb-K138 monoubiquitination could be a paralogous modification that plays a role in establishing permissive chromatin environments around NFY binding sites by mimicking H2B-K120 monoubiquitination, as RNAi of NFY reduces H2B-K120 monoubiquitination and a Nfyb-K138R mutant reduces transcription (Nardini et al., 2013). Indeed, loss of NFY leads to a loss of many chromatin marks associated with transcription activation and elongation, including H3K4me3, H3K36me3, and H3K79me2 (Donati et al., 2008; Gurtner et al., 2008). For example, NFY contacts the hASH2L subunit of the MLL complex, which deposits H3K4me3 (Fossati et al., 2011). NFY also recruits the lysine demethylase KDM1 and its cofactor coREST, which demethylate H3K4me2 into H3K4me1 (Qureshi et al., 2010).

A growing body of evidence suggests that NFY could be a PTF. NF-YC interacts with OCT4, and knockdown of NFY can reduce OCT4 binding at co-occupied sites. This apparent dependency on NFY suggests that NFY could be a PTF or that it could promote the binding of other PTFs, such as OCT4, by mimicking nucleosomes (Oldfield et al., 2014). Research into the mammalian zygotic genome activation demonstrated that the process happened in several steps, with large increases in open chromatin regions at the 2-cell stage and 8-

cell stage. The factor responsible for the increase in open chromatin at the 8-cell stage was OCT4 while the factor responsible for the increase in open chromatin at the 2-cell stage was NFY, lending credence to the notion that NFY possesses PTF capabilities (Lu et al., 2016).

NFY in disease

Like many transcription factors, NFY is a suspect in several pathologies. There are no significant examples of mutant or grossly abnormal NFY subunits implicated in disease, but deviation from normal functions, such as expression pattern or DNA binding, do implicate NFY in some pathologies. At the top of the list is cancer, where the role of NFY in driving the expression of many cell cycle genes makes it a frequent suspect. Although evidence of a direct link between NFY and cancer remains elusive, changes in the transcriptome during transformation of oncogenic cells require the activation of many genes that contain CCAAT boxes in their promoters (reviewed in (Dolfini et al., 2012)).

Polyglutamine diseases, such as Huntington disease, spinocerebellar ataxia type 17 (SCA17), and spinal and bulbar muscular atrophy (SBMA), are another group of pathologies in which NFY may play a role. In particular, the glutamine-rich activations domains of Nfya and Nfyc can aggregate in the cytoplasm of affected neurons, leading to a loss of NFY function in the nucleus. Among its target genes are several chaperones, such as HSP40 and HSP70,

which are crucial to sequestering misfolded proteins that would otherwise contribute to the aggregates. The result is a feedback loop in which NFY subunits, entangled in the aggregates, are unable to activate the genes required for controlling the aggregates, allowing the progression of these polyglutamine diseases (Huang et al., 2011; Katsuno et al., 2010; Yamanaka et al., 2008). In addition to cancer and polyglutamine diseases, NFY is a target in Leigh syndrome, schizophrenia, and diabetes, though its role in these is not well characterized. Despite evidence that conditional deletion of Nfya in postmitotic mouse neurons induces progressive neurodegeneration, there is no link between NFY and neurodegenerative diseases (reviewed in (Dolfini et al., 2012)).

Dominant negatives: a method to study TALE and NFY

Dominant negative constructs provide a unique method for studying the function of both the TALE and NFY complexes. In the case of the TALE factors, the dominant negative factor is a truncated Pbx4 known as PBCAB. PBCAB contains the PBC-A and PBC-B domains required for binding Prep/Meis proteins but lacks the C-terminal domain responsible for interacting with Hox proteins and nuclear localization (Choe et al., 2002). The PBCAB protein competes with endogenous Pbx4 for interactions with Prep/Meis proteins, which lack a nuclear localization signal and cannot enter the nucleus on their own (Berthelsen et al., 1999). When present in excess, PBCAB shifts the equilibrium of TALE factors in

the nucleus versus the cytoplasm drastically to the cytoplasm, causing the TALE factors to become scarce in the nucleus and unable to affect their target genes.

The dominant negative NFY, Nfya-DN, works with a similar mechanism to PBCAB. This construct is a mutant Nfya containing an α2 helix in which is unable to recognize the CCAAT box. The result is a non-functional Nfya subunit which, like PBCAB, competes with endogenous Nfya for binding to the Nfyb/Nfyc dimer. When present in saturating amounts, the equilibrium shifts drastically in favor of the non-functional Nfya subunit, preventing functional Nfyb/Nfyc dimers from binding to DNA and resulting in a loss of NFY function (Mantovani et al., 1994).
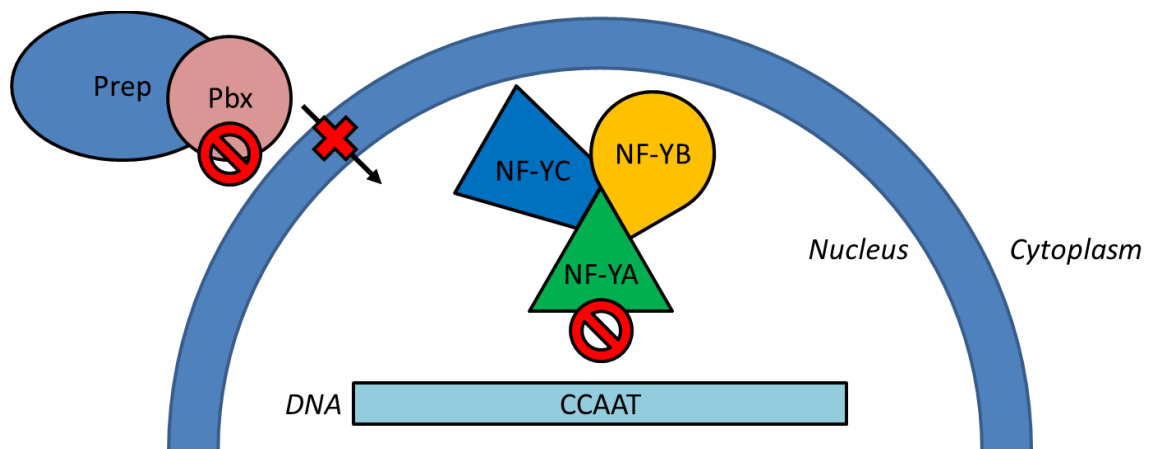
Figure 1.4: Mechanism of action for dominant negative Pbx (PBCAB) and Nfya (Nfya-DN). PBCAB is a truncated zebrafish Pbx4, lacking its C-terminus containing a Hox interacting domain and nuclear localization sequence. PBCAB is still able to bind Prep/Meis proteins in the cytoplasm but unable to enter the nucleus, and when present in saturating quantities sequesters both factors in the cytoplasm and prevents them from activating their target genes. Nfya-DN contains a mutated α2 helix, preventing its recognition of the CCAAT box, but can still interact with the Nfyb/Nfyc dimer. In saturating amounts, Nfya-DN outcompetes endogenous Nfya for Nfyb/Nfyc trimerization, preventing the complete complex from binding its target genes.

In this project, I investigated cooperation between NFY and the TALE factors Pbx4 and Prep1 in driving gene expression at ZGA in zebrafish embryos. Through co-immunoprecipitation of epitope-tagged zebrafish proteins followed by western blotting, I find that Nfya and Nfyb interact with Pbx4 and Prep1 on a protein level. I show that disruption of TALE and NFY function by injecting the dominant negative constructs into zebrafish embryos at the 1-cell stage leads to similar phenotypes, including smaller heads, smaller eyes, and deformed anterior cartilage. In addition, RNA-seq at 12 hpf on embryos injected with the dominant negatives shows that the two factors co-regulate many genes involved in transcription regulation and embryonic development but also independently regulate other processes, including homeobox genes for the TALE factors and cilia for NFY.

Using ChIP-seq at 3.5 hpf, just before ZGA, I find that 94% of Pbx-occupied sites are also occupied by Prep1 (TALE sites), whereas only 37% of regions occupied by Prep1 are also occupied by Pbx4. These Pbx4- and TALE-bound sites are located at DECA motifs. I also find that 17% of TALE sites are co-occupied by Nfya at a motif comprising a DECA site and a CCAAT box. The genes nearest these TALE/NFY regions are largely involved in transcription regulation, while genes downregulated by the dominant negatives and near TALE/NFY ChIP peaks are involved in transcription and development.

To determine whether the TALE factors and NFY work together to drive transcription, I created stable zebrafish lines containing putative enhancer regions with CCAAT boxes and DECA sites upstream of GFP. These lines showed that TALE and NFY could drive GFP expression. In addition, I made stable zebrafish lines containing mutant versions of these enhancers with scrambled CCAAT boxes and DECA sites driving RFP expression. These mutant lines lacked RFP expression, showing that TALE and NFY are required to drive transcription. Taken together, this research demonstrates new roles for the TALE factors and NFY in early embryogenesis.

CHAPTER II: MATERIALS AND METHODS

Experimental Model and Subject Details

Care of zebrafish

The Institutional Animal Care and Use Committee (IACUC) of the University of Massachusetts Medical School approved all procedures involving zebrafish. Adult EkkWill zebrafish were maintained at 28°C in groups at a maximum density of 12 individuals per liter with constant flow. To collect embryos for timing-sensitive experiments, one adult male fish and one adult female fish were placed in separate chambers of a 500mL tank overnight then placed together the following morning for no more than 30 minutes. For experiments that were not timing-sensitive, both adults were placed in the same chamber overnight. Eggs were collected in 10cm dishes, immersed in egg water (60μg/mL Instant Ocean, 0.0002% methylene blue), and maintained in an incubator at 29°C. Dead and unfertilized eggs were manually removed after two hours.

HEK293T cells

HEK293T cells were maintained in growth medium (Dulbecco's Modified Eagle Medium + 10% heat-inactivated fetal bovine serum (FBS) + 0.5% penicillin/ streptomycin) in a humidified incubator with 5% CO_2 at 37°C. Frozen stocks were stored submerged in liquid nitrogen in freezing medium (FBS + 10% DMSO).

Method Details

Transfection for co-immunoprecipitation

HEK293T cells were seeded on 10cm dishes and grown overnight in growth medium. Transfection mixes with Lipofectamine 2000 (ThermoFisher) were prepared according to the manufacturer's instructions with 2µg of total plasmid DNA in Opti-MEM medium. The HEK293T cells were rinsed twice with PBS then immersed in growth medium 10X the volume of transfection mix. The transfection mix was then added to the growth medium, swirled gently to mix, and the dishes were returned to the incubator overnight.

Co-immunoprecipitation

Aliquots of 50µL and 40µL of agarose protein A/G beads (Roche) were rinsed in PBS and then blocked in chilled Blocking Solution (1% bovine serum albumin (BSA) in Co-IP Buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.2mM EDTA 1mM DTT 0.5% Triton X100, 1X C0mplete Protease Inhibitor (Roche)) for at least one hour. The growth medium was aspirated from the transfected cells, which were then rinsed in chilled PBS on ice. The PBS was aspirated and 4mL of Co-IP Buffer was added to each dish of cells. The cells were scraped into a conical tube, incubated on a rotating rack at 4°C for 30 minutes, and centrifuged at 2,000 RPM for 10 minutes at 4°C. The supernatant was transferred to a new conical

tube and 1% was set aside as Input. The blocked pre-clearing beads (50μL) were added to the remainder and incubated on a rotating rack for 2 hours at 4°C. The pre-cleared samples were centrifuged at 2,000 RPM for 3 minutes at 4°C and the supernatant divided into two new tubes. Control antibody was added to one tube and target antibody to the other, then the tubes were incubated on a rotating rack overnight at 4°C. The next morning, the blocked IP beads (40μL) were added to the lysate and incubated on a rotating rack for 4 hours at 4°C. The samples were centrifuged at 2,000 RPM for 3 minutes at 4°C, the supernatant removed, and the beads washed five times in 1mL of Co-IP Buffer each time. The beads were then resuspended in 40μL of Laemmli Buffer (Bio-Rad) + 5% β-mercaptoethanol and incubated at 95°C for 5 minutes.

Western Blot

Samples were run on 4%-20% polyacrylamide gels at 200 V. When complete, the gels were equilibrated in Transfer Buffer (Bio-Rad) for 15 minutes and then the samples were transferred to an activated PVDF membrane at 100 V for 1 hour. The membranes were then blocked in blocking solution (5% nonfat dehydrated milk in TBST (20mM Tris-HCl pH 7.6, 150mM NaCl, 0.1% Tween20) at room temperature on a rocker for 1 hour. The membranes were then transferred to their respective primary antibody solution (5% BSA in TBST) and incubated on a rotating rack overnight at 4°C. The next day, the membranes

were washed four times in TBST for at least 10 minutes each time. The membranes were then placed in 20mL of secondary antibody solution and incubated for 2 hours at 4°C. The membranes were then washed four more times in TBST for at least 10 minutes each time, developed in ECL solution (ThermoFisher), and imaged.

Generation of mRNA for injection

Capped messenger RNAs encoding the dominant negative Nfya (Nfya-DN; (Mantovani et al., 1994)), dominant negative Pbx4 (PBCAB; (Choe et al., 2002)) proteins were generated from 2µg of NotI-digested linearized pCS2+ plasmids using the mMessage mMachine SP6 Transcription Kit (ThermoFisher Scientific) according to the manufacturer's guidelines. The RNA was purified using the RNeasy column with DNase treatment (Qiagen) according to the manufacturer's guidelines. RNA quality was assessed on a 1% agarose gel and its concentration was measured on a NanoDrop instrument. 300pg of RNA injection mix containing water and 0.1% phenol red was injected into zebrafish embryos at the 1-cell stage. Injected embryos were raised to the proper stage according to animal care guidelines.

Characterization of TALE and NFY phenotypes

For gross phenotype assessment, 24 hpf zebrafish embryos were placed on

glass slides in 80% glycerol. For Alcian blue staining, all incubations and washes

took place on a nutator. 5 dpf zebrafish embryos were fixed overnight in 4%

phosphate-buffered paraformaldehyde. Following fixation, the embryos were

washed in 0.1% phosphate-buffered Tween-20 (PBST) and bleached in 30%

hydrogen peroxide for 2 hours. Once bleached, the embryos were rinsed twice in

PBST and then stained overnight in Alcian blue solution (1% hydrochloric acid

(HCl), 70% ethanol, 0.1% Alcian blue). After staining, the embryos were washed

five times in acidic ethanol (HCl-EtOH; 5% HCl, 70% ethanol) with the final wash

lasting 20 minutes. The embryos were then rehydrated in a series of 10-minute

incubations of 75% HCl-EtOH/25% water, 50% HCl-EtOH/50% water, 25% HCl-

EtOH/75% water, and 100% water and imaged. For *in situ* hybridizations, all

incubations and washes took place on a nutator. 24 hpf zebrafish embryos were

fixed overnight in 4% phosphate-buffered paraformaldehyde. Following fixation,

the embryos were washed in a 1:1 methanol:PBST solution, then PBST, and

then treated with 1 µg/mL Proteinase K in PBST for 2 minutes. The embryos

were washed once with -20°C acetone and twice with PBST then incubated at

70°C for 1 hour in Hyb+tRNA Buffer (50% formamide, 5X saline sodium citrate

(SSC), 9.2mM citric acid, 0.5% Tween-20, 50 µg/mL heparin, 500 µg/mL tRNA).

Next, the embryos were transferred to *pax2*/*krox20*/*hoxd4a* probe solution and

incubated at 70°C overnight. After probe incubation, the embryos were washed

sequentially for 10 minutes each at 70°C in Hyb Wash Buffer (50% formamide,

5X saline sodium citrate (SSC), 9.2mM citric acid, 0.5% Tween-20, 50 µg/mL

heparin), 2:1 Hyb:2xSSC, 1:2 Hyb:2xSSC, 2xSSC, 0.2xSSC, and 0.1xSSC, then

blocked in Blocking Solution (2% lamb serum and 2 µg/µL bovine serum albumin

in PBST) at 4°C for 1 hour. The embryos were then incubated with 0.01% α-DIG

antibody at 4°C overnight. Following antibody treatment, the embryos were

washed four times with PBST and two times with Staining Buffer (0.1M Tris pH

9.5, 50mM MgCl2, 125mM NaCl, 0.5% Tween20) then stained with Staining

Solution (100 mg/mL polyvinyl alcohol, 0.35% 5-Bromo-4-chloro-3-indolyl

phosphate, 0.45% 4-Nitro blue tetrazolium) at 37°C until the color developed.

The embryos were then washed four times in PBST and scored. Sample size for

phenotypic analyses was based on previous published reports that these

dominant negative constructs produce phenotypes in >85% of injected embryos

(Choe et al., 2002; Deflorian et al., 2004; Ladam et al., 2018; Waskiewicz et al.,

2001). Embryos were randomly selected for inclusion in injected or control pools.

No animals were excluded, and experiments were not blinded.

RNA extraction

Zebrafish embryos were injected with either PBCAB, Nfya-DN, GFP, or antisense

Nfya-DN mRNA as described above. At the desired timepoint, embryos were

collected into three biological replicates of 50-100 embryos per condition. Dead

animals were counted but excluded from RNA extraction procedures. No other animals were excluded, and selection was not blinded. Each sample was placed in 1mL of Trizol and frozen at -80°C to help break up embryos. Once thawed, the embryos were broken up by pipette and 250µL of chloroform was added to each sample followed by vigorous shaking and a 3-minute incubation at room temperature. The samples were then centrifuged at 12,000*g for 15 minutes at 4°C and the aqueous phase was transferred to a new tube with 500mL of isopropanol and 10µg of GlycoBlue (ThermoFisher Scientific). The samples were vortexed, incubated at room temperature for 10 minutes, and then centrifuged at 12,000*g for 15 minutes at 4°C. The supernatant was removed, and the pellet washed in 75% ethanol then centrifuged at 7,500*g for 5 minutes at room temperature. The supernatant was once again removed, and the pellet was air-dried at room temperature for 10 minutes before resuspension in 50µL of water. The samples were then further purified and treated with DNase using the RNeasy Column kit (Qiagen) and eluted in 30µL of water.

RNA-seq library preparation and deep sequencing

The concentration and quality of each sample was assessed on a Bioanalyzer (Agilent), with all samples having a minimum RNA Quality Number of 8.0 and 28S/18S ratio of 1.0. 4µg of each sample of RNA was shipped to BGI for library preparation and sequencing. Polyadenylated RNAs were selected using oligo dT

beads and then fragmented. N6 random primers were then used to reverse transcribe the library into double-stranded cDNA. A minimum of 20 million single-end 50 bp reads were then generated with the BGISEQ-500 platform.

RT-qPCR

The concentration of each sample was assessed on a NanoDrop instrument. 1µg of RNA was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (ThermoFisher Scientific). To measure the quantity of select mRNAs, 25µL samples were prepared using 2µL of cDNA, 0.2mM of forward and reverse primer for each pair, and SYBR Green qPCR Master Mix (Bimake) according to the manufacturer's guidelines. Measurements were made on a 7300 Real-Time PCR System (Applied Biosystems).

Generation of Nfya Antiserum

Zebrafish Nfya antiserum was prepared by ABClonal Technology. DNA encoding amino acids 1-328 of zebrafish Nfya was cloned into the vector pET-28a-SUMO, containing a 12aa SUMO tag and a 6aa His tag. The vector was transformed into the E. coli Rosetta strain and the antigen peptide was induced with 0.8mM IPTG at 37°C for 4 hours. Small-scale antigen expression was confirmed by Western blot, showing a band at about 58 kDa corresponding to the peptide. The peptide

was purified, appearing in both the supernatant and inclusion bodies. The concentration in the supernatant was 2mg/mL, which was deemed appropriate for immunization. Two rabbits were used for immunization and serum was collected on Day 52. The antiserum was tested by ELISA and deemed of sufficient quality with an OD450 > 0.4 at a 1:64,000 dilution. The antibody was purified via antigen affinity purification, with the polyclonal antibody concentration from animal #E7260 at 4.25mg/mL and from animal #E7621 at 4.66mg/mL. The antibodies were tested via Western blot at a 1:1000 dilution with 10, 5, 1, and 0.5ng of antigen. Bands of about 60 kDa were observed for antibodies from both animals at all four antigen concentrations.

ChIP-seq

Groups of about 5,000 embryos (for Pbx4) and of about 10,000 embryos (for Nfya) were collected at 3.5 hpf and dechorionated in 1X pronase. The embryos were then dissociated by pipette, fixed in 2% formaldehyde in PBS for 10 minutes at room temperature, quenched with 125mM glycine, and flash-frozen in liquid nitrogen. Processing of cell pellets followed the protocol previously described (Amin et al., 2015). Nuclei were isolated in L1 Buffer (50mM Tris-HCl pH 8.0, 2mM EDTA, 0.1% NP-40, 10% glycerol, 1mM PMSF) then lysed in SDS Lysis Buffer (50mM Tris-HCl pH 8.0, 10mM EDTA, 1% SDS). Chromatin was sheared to an average length of 300 bp using a Palmer immersion sonicator

(Three 1-minute rounds of 10s on/2s off at 40% amplitude) and diluted 1:10 in ChIP Dilution Buffer (50 mM Tris-HCl pH8.0, 5 mM EDTA, 200 mM NaCl, 0.5% NP-40, 1 mM PMSF). The samples were pre-cleared with 50µL of Protein A Dynabeads (ThermoFisher Scientific) at 4°C for 3 hours, then Input samples were set aside and stored at -80°C. Next, 10µL of the appropriate antiserum was added (α-Pbx4 or α-Nfya) and the samples were incubated rotating at 4°C overnight. The immune complexes were precipitated onto 50µL of Protein A Dynabeads, which were washed five times with Wash Buffer (20 mM Tris-HCl pH8.0, 2 mM EDTA, 500 mM NaCl, 1% NP-40, 0.1% SDS, 1 mM PMSF), three times with LiCl Buffer (20 mM Tris-HCl pH8.0, 2 mM EDTA, 500 mM LiCl, 1% NP-40, 0.1% SDS, 1 mM PMSF), and three times with TE Buffer (10 mM Tris-HCl pH8.0, 1 mM EDTA, 1 mM PMSF). To elute chromatin, the beads were incubated in 50µL of fresh Elution Buffer with shaking at 1,500 RPM for 15 minutes at 25°C then 15 minutes at 65°C. To reverse crosslinks, 2µL of 5M sodium chloride was added to the samples, which were then incubated at 65°C overnight. Purification of the DNA was accomplished using the MicroChIP Dia Pure Column kit (Diagenode) according to the manufacturer's guidelines with an 11µL elution. To quantify the concentration of DNA, 1µL of each sample was passed through the dsDNA HS Assay (ThermoFisher Scientific) according to the manufacturer's guidelines and quantified on a Qubit device.

ChIP-seq library preparation and deep sequencing

ChIP-seq libraries were prepared using the MicroPlex Library Preparation Kit v2 (Diagenode) according to the manufacturer's guidelines. The entirety of each ChIP sample was used, and Input samples were either diluted to the same concentration as their corresponding ChIP sample or, if the concentration of the corresponding ChIP sample was below the Qubit's range, diluted to 0.2 ng/μL. Following library synthesis, an Illumina HiSeq4000 Sequencer was used to sequence the libraries.

E1b-GFP-Tol2 cloning

Putative enhancers of about 500 bp centered on Prep1 peaks near DECA sites and CCAAT boxes were amplified via PCR from 24 hpf wild-type zebrafish genomic DNA using specific primers with XhoI sites (*tcf3a*, *tle3a*, *dachb*, *fgf8a*, *pax5*, *her6*, *prdm14*) or BglII sites (*yap1*) flanking either end (Table S1). The fragments were ligated into the E1b-GFP-tol2 (Birnbaum et al., 2012; Li et al., 2010) empty backbone digested with XhoI or BglII and transformed into competent DH5alpha E. coli cells (New England Biolabs). The amplified vector was validated by Sanger sequencing and purified using the Plasmid Maxi Kit (Qiagen).

Generation of pTransgenesis donor vectors

Mutant enhancers were generated by changing DECA sites contained within each enhancer to the sequence CGGTTGGTGC, which has been shown to prevent TALE binding (Vlachakis et al., 2000), and CCAAT boxes to the sequence ATGCG. Both mutant and wild-type versions of each enhancer were generated using gBlock technology (Integrated DNA Technologies). Due to limitations in gBlock synthesis, a 34 bp AT-rich region at the 3' end of the *tcf3a* enhancer could not be included compared to the E1b-GFP-tol2 version. A-tails were added to each end of the gBlock fragments using OneTaq Hot Start DNA Polymerase (NEB) (50ng of gBlock DNA, 1 unit of OneTaq Hot Start DNA Polymerase, 1X OneTaq Standard Reaction Buffer, 0.05mM dATP, 1.5mM MgCl2) and incubating the samples at 70°C for 30 minutes. 1µL of A-tailed gBlock fragment solution was then cloned into the pCR8 vector using the pCR8/GW/TOPO TA Cloning Kit (ThermoFisher Scientific) according to the manufacturer's guidelines. The product was transformed into TOP10 chemically competent cells, validated by Sanger sequencing, and then purified using the Plasmid Midi Kit (Qiagen).

Generation of pTransgenesis vectors

pTransgenesis vectors were assembled using the LR Clonase II Plus enzyme mix (ThermoFisher Scientific). Four cassettes were assembled in one reaction, with ɣ-crystallin:venusGFP as the p1 cassette (European Xenopus Resource Center (EXRC)), gBlock enhancers in pCR8 as the p2 cassette and Tol2/I-SceI-CH4-SAR/I-SceI/Tol2/P-element (EXRC) as the pDest-4 cassette. The p3.13 cassette was generated by ligating a BamHI/BglII-digested gBlock (containing the SV40 minimal promoter) into BamHI-digested p3.13 Katushka-RFP plasmid (EXRC). 10fmol of each of the p1, p2, and p3 cassettes were combined with 20fmol of p4 cassette and 2µL of LR Clonase II Plus enzyme mix for the LR reaction in 10µL. The reaction was incubated at 25°C for 16 hours then treated with Proteinase K at 37°C for 10 minutes. 2µL of LR reaction product was transformed into Top10 chemically competent cells, validated by Sanger sequencing, and then purified using the Plasmid Maxi Kit (Qiagen).

Generation and observation of transgenic animals

Injection mixes containing 100ng/µL of E1b-GFP-Tol2 or pTransgenesis vector, 100ng/µL of Tol2 mRNA, and 0.1% phenol red were injected into wild-type zebrafish embryos at the 1-cell stage. The animals were observed for transient fluorescence for the first week, then raised to adulthood. Mature fish were

crossed with wild-type fish and the offspring were observed for fluorescence. For

E1b-GFP-Tol2 fish, GFP was observed as early as 18 hpf. For pTransgenesis

fish, RFP expression and GFP expression overlap was best observed at 32 hpf,

with RFP being apparent sooner and disappearing by about 48 hpf while GFP

persisted after that time. Thus, any fish that appeared to be RFP+/GFP- were

separated and observed for GFP expression at a later timepoint.

Quantification and Statistical Analysis

RNA-seq analysis

RNA-seq analysis was performed using the University of Massachusetts Medical

School Dolphin web interface. Ribosomal RNA reads were filtered out and

FastQC was used to assess the quality of the remaining reads. RSEM_v1.2.28

with parameters -p4 --bowtie-e 70 --bowtie-chunkmbs 100 (Li & Dewey, 2011)

was used to align the reads to the DanRer10 zebrafish transcriptome and

normalize gene expression to transcripts per million (TPM). This revealed that

PBCAB replicate 2 underperformed relative to the other samples and was

excluded from further analysis. DeSeq2 (Anders & Huber, 2010) was used to

identify differentially-expressed genes between three independent biological

replicates of 12 hpf embryos injected with GFP and three independent biological

replicates of 12 hpf embryos injected with Nfya-DN or between three

independent biological replicates of 12 hpf embryos injected with GFP and two independent biological replicates of 12 hpf embryos injected with PBCAB. DEBrowser was used to identify outliers among the replicates. To compensate for the exclusion of one replicate in GFP versus PBCAB analysis, only differentially expressed genes with a p-adj ≤ 0.01 (Benjamini and Hochberg FDR) were considered for analysis.

ChIP-seq Data Processing

All eight ChIP-seq fastq files (two independent 3.5 hpf Pbx4 biological replicates, two independent 3.5 hpf Nfya biological replicates, and matched input DNA controls for each) contained 76 bp paired-end sequences. The raw sequence quality was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Fastq-screen (https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/). Next, remaining adapter reads were filtered out and poor-quality 3' end sequences were trimmed with Trimmomatic version 0.36 (Bolger et al., 2014) using default parameters for ILLUMINACLIP and SLIDINGWINDOW and MINLENGTH set to 50 bp. Using Bowtie2 version 2.2.3 (Langmead & Salzberg, 2012), the processed reads were then mapped to UCSC browser zebrafish genome release GRCz10 (danRer10/September 2014) (Tyner et al., 2017), and the mapped reads were further filtered with SAMtools view version 0.1.19 (Li et al., 2009) (with flags used

-f 2 -q30) to remove reads with poor mapping quality and discordant mapped read pairs. To call peaks, the data, excluding reads that mapped to the mitochondrial genome and unassembled contigs in the assembly, was next passed through MACS2 version 2.1.0.20140616 (Zhang et al., 2008) with the q-value threshold set to 0.05 and default parameters except that the effective genome size was set to 1.03e9 (this equates to 75% of the total genome sequence, excluding 'N' bases).

ChIP-seq Analysis

Since the biological replicates for each factor demonstrated robust overlap, the sum of the two replicates was used for all subsequent analyses, by including all peaks meeting the selected cutoff in at least one of the biological replicates. Three different cutoffs were considered: all peaks with a fold enrichment (FE) ≥ 10, all peaks with a FE ≥ 4, and the top 10% of all peaks in each data set. The FE ≥ 10 cutoff showed the highest overlap between Pbx4 and Prep1 peaks as a percentage of the total Pbx4 peaks and was selected as the best cutoff for ChIP-seq analysis (Table 3.7). For a larger set of peaks, FE ≥ 4 peaks were considered for comparison to RNA-seq data (Figure 3.12; Table 3.8).

ChIP peak overlap analysis

In the text, we use the term 'overlap' to indicate peaks identified as follows: ChIP peaks shared between different data sets were identified with the Intersect tool and exclusive peaks were identified using the Subtract tool in Galaxy (Goecks et al., 2010). All coordinates used were 200 bp in length centered on peak summits and considered overlapping if they shared one or more base pairs.

qPCR Analysis

ddCt values were calculated from raw Ct values according to the formula 0.5^{Ct}. Average ddCt values were then calculated by taking the mean of all three biological replicates. The ddCt of each GFP replicate was then normalized to the average *gapdh* ddCt according to the formula ddCt$_{gfp}$/average ddCt$_{gapdh}$ and then the mean of the normalized values was determined. Error bars were calculated based on the standard deviation of the three normalized GFP replicates in Excel. To determine whether the dominant negative conditions were significantly different from the control condition, an unpaired t-test was used in Excel, with p-values < 0.05 considered significant.

Determination of nearest genes to ChIP peaks

The number of Ensembl zebrafish transcription start sites within 5 kb or 30 kb of the summit of ChIP peaks was determined using the bedtools suite (Quinlan & Hall, 2010) in the Galaxy toolshed (Goecks et al., 2010). ChIP peak coordinates in danrer10 were converted to danrer7 (Zv9) using the LiftOver tool in the UCSC browser. The identities of genes near ChIP peaks were determined by the GREAT software version 3.0.0 (Hiller et al., 2013; McLean et al., 2010) using the default settings of basal plus extension with proximal set to 5 kb upstream and 1 kb downstream and distal set to 1,000 kb.

GO term analysis

Gene ontology (GO) terms enriched within different sets of genes were determined using DAVID version 6.8 (Huang et al., 2009a, 2009b). GO terms were ranked according to the EASE score, which was calculated based on a modified Fisher's exact p-value and graphed as the -log10 of that value.

DNA binding motif analysis

Significantly enriched binding motifs were identified using MEME and DREME within the MEME-Suite version 4.11.1 (Bailey et al., 2009; Machanick & Bailey,

2011). Both MEME and DREME were run according to their default settings. CENTRIMO was also run with default settings to determine the distribution of discovered motifs relative to ChIP peaks.

Chromatin feature analysis

Version 2.0 of the Deeptools (Ramírez et al., 2014) toolset in the Galaxy toolshed was used to create mean score profiles and heatmaps. Using the computeMatrix tool with region inputs of BED files containing ChIP coordinates and sample inputs of wiggle files from previously published data sets downloaded from GEO (Key Resources Table), signal matrices were generated in reference-point mode with the center set as the reference point. The distance upstream of the start sites and downstream of the end sites were set to 1000 bp with a bin size of 25 bp and ranked by mean signal when necessary. Heatmaps and profiles were generated from the matrices using the plotHeatmap and plotProfile tools respectively. The previously published H3K27ac, H3K4me1, and H3K4me3 data sets were all performed at 4.5 hpf, which is somewhat later than the Pbx4, Nfya, and Prep1 ChIP-seq experiments performed at 3.5 hpf; however, asynchronous development in zebrafish embryos and large sample sizes make considerable overlap likely.

Data and Code Availability

RNA-seq data is available in GEO under accession number GSE133459. ChIP-

seq data is available in ArrayExpress under accession number E-MTAB-8137.

Key Resources Table

Reagent or Resource	Source/ Reference	Identifier
Antibodies		
Rabbit polyclonal α-zebrafish Pbx	(Choe et al., 2009)	N/A
Rabbit polyclonal α-zebrafish Nfya	This paper	N/A
Bacterial and Virus Strains		
Subcloning Efficiency DH5α competent cells	Thermo Fisher	18265-017
OneShot Top10 chemically competent cells	Thermo Fisher	C404003
Chemicals, Peptides, and Recombinant Proteins		
Protein-A Dynabeads	Thermo Fisher	10001D
NotI	New England Biolabs	R0189S
XhoI	New England Biolabs	R0146S
BglII	New England Biolabs	R0144S
Critical Commercial Assays		
mMESSAGE mMACHINE SP6 Transcription Kit	Thermo Fisher	AM1340

RNeasy Mini Kit	Qiagen	74104
DIG DNA Labeling Mix	Millipore Sigma	11277065910
Trizol	Thermo Fisher	15596026
GlycoBlue	Thermo Fisher	AM9515
MicroPure DiaChIP Columns	Diagenode	C03040001
dsDNA HS Assay	Thermo Fisher	Q32851
MicroPlex Library Preparation Kit v2	Diagenode	C05010012
OneTaq Hot Start DNA Polymerase	New England Biolabs	M0481L
pCR8/GW/TOPO	Thermo Fisher	45-0642
Plasmid Midi/Maxi Kit	Qiagen	12143/12163
LR Clonase II Plus Enzyme Mix	Thermo Fisher	12538-120
High Capacity cDNA Reverse Transcription Kit	Thermo Fisher	4368814
SYBR Green qPCR Master Mix	Bimake	B21203
Deposited Data		
Pbx4 ChIP-seq and Inputs in 3.5 hpf zebrafish embryos	This paper	E-MTAB-8137
Nfya ChIP-seq and Inputs in 3.5 hpf zebrafish embryos	This paper	E-MTAB-8137
PBCAB and GFP RNA-seq in 12 hpf zebrafish embryos	This paper	GSE133459
Nfya DN and GFP RNA-seq in 12 hpf zebrafish embryos	This paper	GSE133459
Prep1 ChIP-seq and Inputs in 3.5 hpf zebrafish embryos	(Ladam et al., 2018)	E-MTAB-5967
H3K27ac ChIP-seq in dome zebrafish embryos, WIG files	(Bogdanović et al., 2012)	GSM915197

H3K4me1 ChIP-seq in dome zebrafish embryos, WIG files	(Bogdanović et al., 2012)	GSM915193
H3K4me3 ChIP-seq in dome zebrafish embryos, WIG files	(Bogdanović et al., 2012)	GSM915189
H3K27ac ChIP-seq in 80% epiboly zebrafish embryos, WIG files	(Bogdanović et al., 2012)	GSM915198
H3K4me1 ChIP-seq in 80% epiboly zebrafish embryos, WIG files	(Bogdanović et al., 2012)	GSM915194
H3K4me3 ChIP-seq in 80% epiboly zebrafish embryos, WIG files	(Bogdanović et al., 2012)	GSM915190
Experimental Models: Organisms/Strains		
Strain EKW	EkkWill breeders	http://www.ekkwill.com/
Oligonucleotides		
gapdh forward primer TGCTGGTATTG CTCTCAACG		N/A
gapdh reverse primer AACAGCAAAGG GGTCACATC		N/A
gfp forward primer ATGGTGAGCAA GGGCGAGGAG		N/A
gfp reverse primer TTACTTGTACA GCTCGTCCATG		N/A
Recombinant DNA		
Nfya-DN in pCS2+	(Ladam et al., 2018)	N/A
PBCAB in pCS2+MT	(Choe et al., 2002)	N/A
tcf3a element in E1b-GFP-Tol2	This paper	N/A
tle3a element in E1b-GFP-Tol2	This paper	N/A
dachb element in E1b-GFP-Tol2	This paper	N/A

fgf8a element in E1b-GFP-Tol2	This paper	N/A
yap1 element in E1b-GFP-Tol2	This paper	N/A
pax5 element in E1b-GFP-Tol2	This paper	N/A
her6 element in E1b-GFP-Tol2	This paper	N/A
prdm14 element in E1b-GFP-Tol2	This paper	N/A
tcf3a element in pCR8	This paper	N/A
tle3a element in pCR8	This paper	N/A
Mutant *tcf3a* element in pCR8	This paper	N/A
Mutant *tle3a* element in pCR8	This paper	N/A
pTransgenesis p1 ɣ-crystallin::VenusGFP	(Love et al., 2011)	N/A
pTransgenesis p3 sv40 minimal promoter:: Katushka RFP	This paper	N/A
pTransgenesis pDest4 Tol2/ I-SceI-CH4-SAR/I-SceI/ Tol2/P-element	(Love et al., 2011)	N/A
Software and Algorithms		
FastQC	Babraham Institute	https://www.bioinformatics. babraham.ac.uk/projects/fastqc/ RRID:SCR_014583
FastQ Screen	Babraham Institute	https://www.bioinformatics. babraham.ac.uk/projects/ fastq_screen/ RRID:SCR_000141
Trimmomatic v0.32	(Bolger et al., 2014)	http://github.com/timflutre/ trimmomatic RRID:SCR_011848
Bowtie v2.2.3	(Langmead & Salzberg, 2012)	http://github.com/BenLangmead/ bowtie2 RRID:SCR_005476
SAMtools v0.1.19	(Li et al., 2009)	http://github.com/samtools/ samtools RRID:SCR_002105
MACS v2.1.0.20140616	(Zhang et al., 2008)	http://github.com/taoliu/MACS

RSEM v1.2.28, Dolphin, Biocore, University of Massachusetts Medical School	(Li & Dewey, 2014)	http://www.umassmed.edu/biocore/introducing-dolphin/ RRID:SCR_013027
DESeq2, Dolphin, Biocore, University of Massachusetts Medical School	(Anders & Huber, 2010)	http://www.umassmed.edu/biocore/introducing-dolphin/ RRID:SCR_015687
DEBrowser v1.12.2	(Kucukural et al., 2019)	http://github.com/UMMS-Biocore/debrowser
Galaxy web interface	(Goecks et al., 2010)	http://usegalaxy.org RRID:SCR_006281
BedTools, Galaxy	(Quinlan & Hall, 2010)	http://usegalaxy.org RRID:SCR_006646
DeepTools, Galaxy	(Ramírez et al., 2014)	http://usegalaxy.org
MEME-ChIP	(Bailey et al., 2009; Machanick & Bailey, 2011)	http://meme-suite.org/tools/meme-chip RRID:SCR_00178
DAVID v6.8	(Huang et al., 2009a, 2009b)	http://david.ncifcrf.gov/ RRID:SCR_001881
GREAT v3.0.0	(Hiller et al., 2013; McLean et al., 2010)	http://bejerano.stanford.edu/great/public/html RRID:SCR_005807

A: *tcf3a*-WT enhancer

TACTGCGTTAATCGCGCGTTTACTTTGATATTTAATCCACAACCAACACAATT
AAAACGCCAAACATCAGCGACGACAGTATATGTAACTTTATCCTGATATTTC
CCGATTGTGCTTTAAATCACGCAGTACTAGACTCGCGCGCGGAATGACACG
ACGCACTGTTGAAGAGCGATGGACTGAGAAAAGTGCGAGATGGCACGATA
GACCCACTGAGCGGACCAATAGCGATCGGGGAAAGTTTGATTGACGTATTC
GGTGGCCAATCGAAGATCGTGTTAACACGAAAGCCAAGCCTCTCTTCCATG
CACACCCTAGCCAGGTTTTAAAAGAATGGCAACAGGAAGCCATGGAATACT
GTTGTGTTTTGTTGTTTGGTAAATGCTAATGTTTACCGCTAACCGCTCAAACT
AACTTCAAATGAATTCGACTCGAAACATAACATTGTTATTATTACATTTAGAC

B: *tcf3a*-mut enhancer

TACTGCGTTAATCGCGCGTTTACTTTGATATTTAATCCACAACCAACACAATT
AAAACGCCAAACATCAGCGACGACAGTATATGTAACTTTATCCTGATATTTC
CCGATTGTGCTTTAAATCACGCAGTACTAGACTCGCGCGCGGAATGACACG
ACGCACTGTTGAAGAGCGATGGACTGAGAAAAGTGCGAGATGGCACGATA
GACCCACTGAGCGGAATGCGAGCGATCGGGGAAAGTTCGGTTGGTGCATT
CGGTGGATGCGCGAAGATCGTGTTAACACGAAAGCCAAGCCTCTCTTCCAT
GCACACCCTAGCCAGGTTTTAAAAGAATGGCAACAGGAAGCCATGGAATAC
TGTTGTGTTTTGTTGTTTGGTAAATGCTAATGTTTACCGCTAACCGCTCAAAC
TAACTTCAAATGAATTCGACTCGAAACATAACATTGTTATTATTACATTTAGA
C

C: *tle3a*-WT enhancer

ATAGATGACATTACCAGGACTGTATTGTTATATGGGTAACATGCGATTATGA
GTGAGGGCTTTTTTTAATGTTATTAAGTGTTTGCATGCTCCTTTGCTCCTTTG
TTTTATGTAAGGCTCTCATTACCACGTGGTAGTAACAGATTGTTTGAACTGG
AAAGAAAGCCATTCGAAGCTAATTAAGCAGCCATTCCAGGCACTATTCACG
GGCAGAAGAGCGAGAAGCACAGGCATTTGTCAGCGCTTGACCCCGCGTGG
TATTGATTGACAACAAACCTTCTTGAATGACAGCCTTAACCTTTCCCGTCCAA
TTGCAGTCGAGAGAATATAGATGCTGCTCTGCGATTGGCTGAGAAGCTGTA
AAGCCGCAAAGGGATCCCACGTGGGTGCAGCAGAAGAAACGGCACAGGAT
TGGCCGCTTCTTCTGAGTTCAGACATGGCCGTTGTTCACGGAGATCAAACC
TGAACAATCATCGTATTCCCAGCGCTAGC

D: *tle3a*-mut enhancer

ATAGATGACATTACCAGGACTGTATTGTTATATGGGTAACATGCGATTATGA
GTGAGGGCTTTTTTTAATGTTATTAAGTGTTTGCATGCTCCTTTGCTCCTTTG
TTTTATGTAAGGCTCTCATTACCACGTGGTAGTAACAGATTGTTTGAACTGG
AAAGAAAAGCCATTCGAAGCTAATTAAGCAGCCATTCCAGGCACTATTCACG
GGCAGAAGAGCGAGAAGCACAGGCATTTGTCAGCGCTTGACCCCGCGTGG
TATTGATTGACAACAAACCTTCTCGGTTGGTGCCCTTAACCTTTCCCGTATG
CGTGCAGTCGAGAGAATATAGATGCTGCTCTGCGCGCATCTGAGAAGCTGT
AAAGCCGCAAAGGGATCCCACGTGGGTGCAGCAGAAGAAACGGCACAGCG
CATCCCGCTTCTTCTGAGTTCAGACATGGCCGTTGTTCACGGAGATCAAACC
TGAACAATCATCGTATTCCCAGCGCTAGC

E: sv40 minimal promoter

AAAGATCTGCGATCTGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTA
ACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCC
CATCGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCT
CTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTT
GCAAAAAGCTTGGCATTCCGGTACTGTTGGTAAAGGATCCAA

CHAPTER III: Results

This chapter includes material that was published in (Ladam et al., 2018) as well as material from the manuscript "Combinatorial Action of NF-Y and TALE at Embryonic Enhancers Defines Distinct Gene Expression Programs during Zygotic Genome Activation in Zebrafish," which has been submitted to eLife and is currently under review.

The TALE factors form complexes with NFY

Sequence motif analysis of 3.5 hpf Prep1 ChIP-seq data using the MEME Suite showed that Prep1 predominantly bound the DECA motif at that timepoint. Further, it identified the CCAAT box as appearing near about 30% of those DECA sites bound by Prep1. CCAAT boxes near DECA sites consistently appeared a set distance from the DECA site of roughly 20 bp. This raised the possibility that the TALE factors and NFY formed protein complexes at these regions. To test this, I performed pairwise co-immunoprecipitation analyses with Pbx4, Prep1, Nfya, and Nfyb. I transfected different combinations of epitope-tagged versions of the proteins into HEK293T cells, immunoprecipitated one of the factors, and performed a western blot for the other. In this context, I found that both Pbx4 and Prep1 co-immunoprecipitated with both Nfya and Nfyb. I conclude that the TALE factors and NFY form protein complexes at genomic loci containing a DECA motif and CCAAT box.
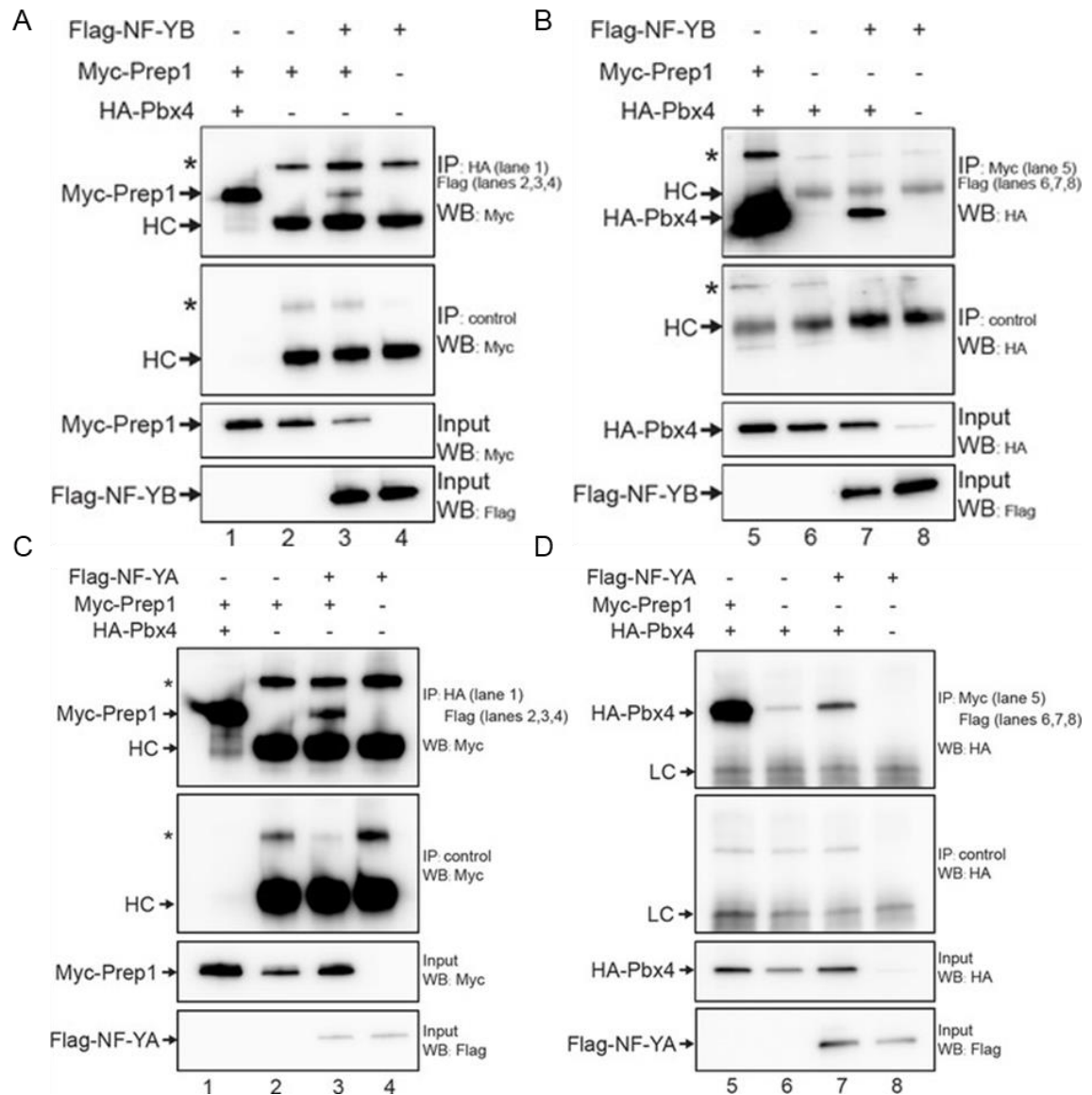
Figure 3.1: The TALE factors and NFY form protein complexes. HA::Pbx4, Myc::Prep1, FLAG::Nfya, and FLAG::Nfyb were transfected into HEK293T cells in pairwise combinations of FLAG::Nfyb and Myc::Prep1 (A), FLAG::Nfyb and HA::Pbx4 (B), FLAG::Nfya and Myc::Prep1 (C), or FLAG::Nfya and HA::Pbx4 (D). FLAG was then immunoprecipitated with HA or Myc as a positive control, followed by western blotting with α-HA or α-Myc. HC indicates heavy chain, LC indicates light chain, and asterisks indicate non-specific signal.

TALE and NFY are required for formation of anterior embryonic structures

In order to assess the roles of TALE and NFY during embryogenesis, I first set out to disrupt their function. Previous work demonstrated that the TALE factors are required for formation of anterior embryonic structures, such that loss of various combinations of TALE factors results in animals with smaller heads, small (or absent) eyes, swollen pericardium/cardiac edema, and hindbrain defects (Choe et al., 2002; Deflorian et al., 2004; Pöpperl et al., 2000; Pöpperl et al., 1995; Waskiewicz, et al., 2002; Waskiewicz et al., 2001). This similarity in phenotypes is likely due to the fact that multiple TALE factors frequently act together in larger protein complexes, which are rendered ineffective when one or more TALE factors are disrupted (reviewed in (Ladam & Sagerström, 2014; Merabet & Mann, 2016)). In preliminary experiments, we recently observed abnormal anterior development also upon disruption of NFY function (Ladam et al., 2018). Here, I extend this analysis to directly compare disruption of TALE factors (using the dominant negative PBCAB construct reported previously; (Choe et al., 2002)) to disruption of NFY function (using the previously reported Nfya dominant negative construct (Nfya-DN); (Mantovani et al., 1994)) and find smaller heads in both cases (Figure 3.2 A-D). A more detailed examination revealed abnormal head cartilage formation (53% of animals with disrupted NFY and 79% of animals with disrupted TALE function; Figure 3.3 A-H and Table 3.1) as well as loss of eyes (28% of animals with disrupted NFY and 19% of animals with disrupted TALE function; Table 3.2). Using *in situ* hybridization to detect

expression of *pax2* (at the midbrain-hindbrain boundary), *krox20* (in

rhombomeres 3 and 5 of the hindbrain), and *hoxd4* (in the spinal cord) in 24 hpf

embryos, I observed loss of r3 *krox20* expression upon TALE disruption (52% of

embryos; as reported previously; (Choe et al., 2002; Deflorian et al., 2004;

Pöpperl et al., 2000; Waskiewicz et al., 2002; Waskiewicz et al., 2001)), but did

not detect any effects of NFY disruption (Figure 3.4 A-H and Table 3.3). I

conclude that both the TALE factors and NFY function in formation of the anterior

embryo and that the TALE factors have a distinct role in hindbrain patterning.
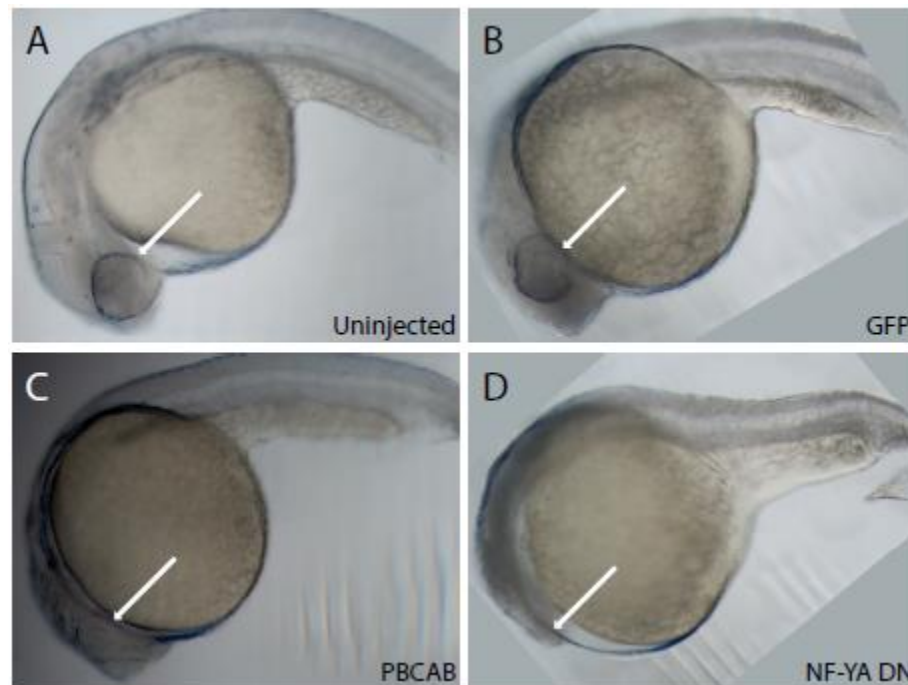
Figure 3.2: Disruption of TALE or NFY function affects anterior embryonic development.
Zebrafish embryos at the 1-cell stage were either left uninjected (A) or injected with mRNA
encoding a control (GFP; B), dominant negative TALE (PBCAB; C), or dominant negative Nfya
(Nfya-DN; D) and raised to 28 hpf. White arrows highlight differences in eye morphology.
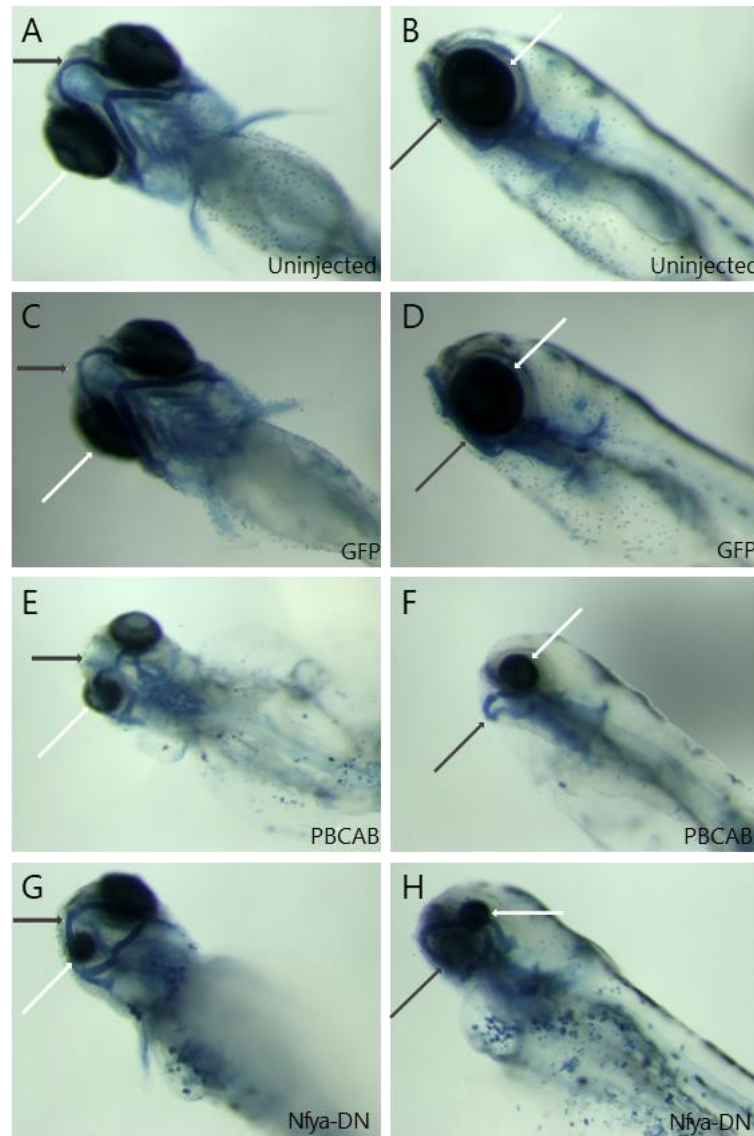
Figure 3.3: Disruption of TALE or NFY function affects anterior cartilage development.
Zebrafish embryos at the 1-cell stage were either left uninjected (A, B) or injected with mRNA encoding a control (GFP; C, D), dominant negative TALE (PBCAB; E, F), or dominant negative Nfya (Nfya-DN; G, H). The embryos were then raised to 5 dpf and stained with Alcian blue. White arrows highlight differences in eye morphology and black arrows highlight differences in cartilage morphology.
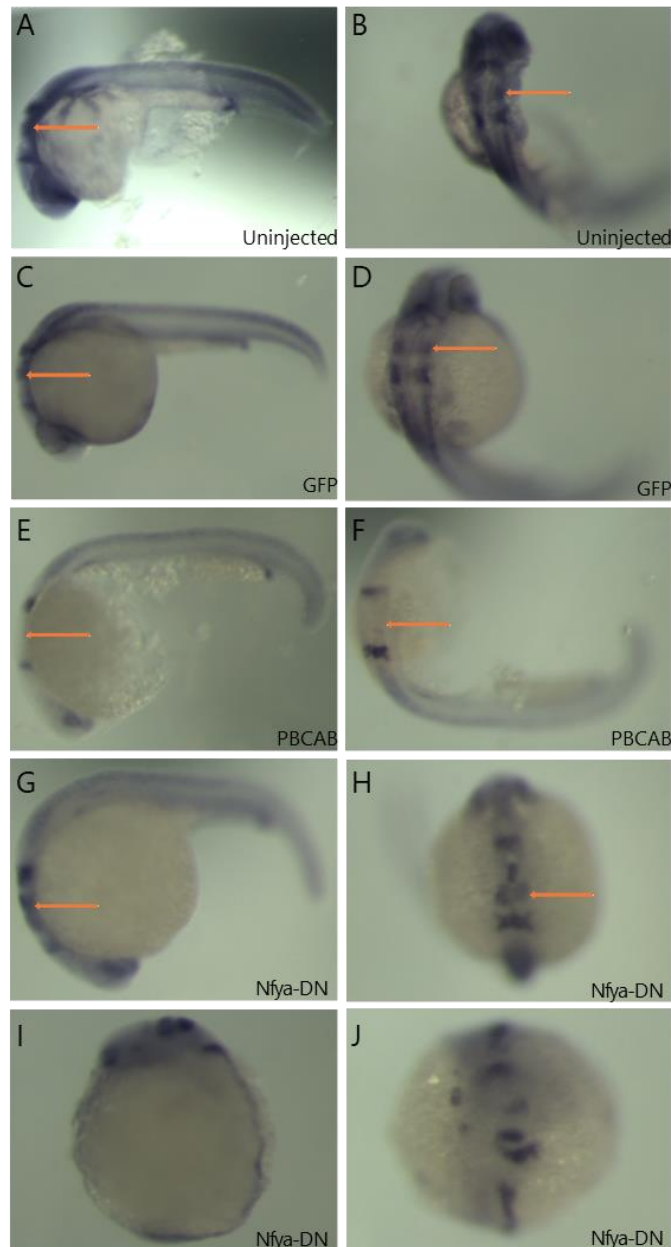
Figure 3.4: PBCAB disrupts hindbrain segmentation. Zebrafish embryos at the 1-cell stage were either left uninjected (A, B) or injected with mRNA encoding a control (GFP; C, D), dominant negative TALE (PBCAB; E, F), or dominant negative Nfya (Nfya-DN; G-J). The embryos were then raised to 24 hpf and processed for detection of *pax2* (at the mid/hindbrain boundary), *krox20* (in rhombomeres 3 and 5) and *hoxd4* (in the spinal cord) transcripts by *in situ* hybridization. Orange arrows highlight differences in *krox20* rhombomere 3 expression. Panels I and J show representative images of embryos scored as having gross abnormalities in Table 3.3.

	Uninjected	GFP	PBCAB	Nfya-DN
Normal	77 (100%)	54 (100%)	11 (21.2%)	26 (47.3%)
Abnormal	0 (0%)	0 (0%)	41 (78.8%)	29 (52.7%)
Total	77	54	52	55

Table 3.1: Dominant negatives cause abnormal cartilage development. Embryos were either left uninjected or injected with mRNA encoding GFP (control), PBCAB, or Nfya-DN. Embryos were raised to 5 dpf, stained with Alcian blue, and scored as having normal or abnormal cartilage development.

	Uninjected	GFP	PBCAB	Nfya-DN
Two Eyes	65 (100%)	48 (78.7%)	46 (73.0%)	18 (50.0%)
No Eye/Anterior Truncation	0 (0%)	0 (0%)	12 (19.0%)	10 (27.8%)
Grossly Abnormal	0 (0%)	13 (21.3%)	5 (8.0%)	8 (22.2%)
Total	65	61	63	36

Table 3.2: Dominant negatives cause abnormal anterior development. Embryos were either left uninjected or injected with mRNA encoding GFP (control), PBCAB, or Nfya-DN. Embryos were raised to 24 hpf, processed for detection of *pax2*, *krox20*, and *hoxd4* transcripts by *in situ* hybridization, and scored as having two eyes, no eyes/an anterior truncation, or grossly abnormal development.

	Uninjected	GFP	PBCAB	Nfya-DN
Normal	59 (90.8%)	48 (78.7%)	13 (20.6%)	28 (77.7%)
Missing r3	0 (0%)	0 (0%)	33 (52.4%)	0 (0%)
Poor Signal	6 (9.2%)	13 (21.3%)	17 (27.0%)	8 (22.2%)
Total	65	61	63	36

Table 3.3: PBCAB causes loss of r3. Embryos were either left uninjected or injected with mRNA encoding GFP (control), PBCAB, or Nfya-DN. Embryos were raised to 24 hpf, processed for detection of *pax2*, *krox20*, and *hoxd4* transcripts by *in situ* hybridization, and scored as having normal hindbrain development, lacking r3, or having too poor signal to score.

TALE and NFY have both shared and independent transcriptional targets

In order to identify shared and separate functions of the TALE factors and NFY, I carried out RNA-seq at 12 hpf of zebrafish development (Figure 3.5 A-C; Figure 3.6 A, B, and Table 3.4). I selected the 12 hpf timepoint for RNA-seq analysis in order to ensure that I would broadly capture gene expression changes resulting from disruption of TALE and NFY function. I find that disruption of TALE function affects the expression of 1,500 genes (646 downregulated, 854 upregulated; FC > 1.5, p-adj < 0.01; Figure 3.5 B, D; Figure 3.8 A) at 12 hpf. Since TALE factors are thought to act primarily as activators of transcription, I focused on genes downregulated upon disruption of TALE function. Applying the DAVID functional annotation tool, I find that TALE-dependent genes are enriched for functions related to transcription (particularly *hox* genes) as well as for factors controlling embryogenesis (Figure 3.7 A; GO terms for genes upregulated upon disruption of TALE function are shown in Figure 3.8 B). Accordingly, an examination of individual TALE-dependent genes identified members of several classes of transcription factors and developmental control genes (Figure 3.7 B). This result agrees with our previous analysis of gene expression changes in response to loss of TALE function (using antisense morpholino oligos targeting *pbx2*, *pbx4*, and *prep1.1*; (Ladam et al., 2018)) – confirming that disruption of various combinations of TALE factors produces similar phenotypes. The downregulation of transcriptional and developmental control genes upon TALE disruption is also consistent with the phenotype observed in Figures 3.2, 3.3, and

3.4. I next examined the effect of disrupting NFY function and find that 902 genes are affected (325 downregulated, 577 upregulated; Figure 3.5 C, D; Figure 3.8 A) at 12 hpf. An analysis of the GO terms associated with NFY-dependent genes revealed high enrichment in functions related to cilia and, to a lesser extent, in genes broadly controlling transcription and development (Figure 3.7 C; GO terms for upregulated genes are shown in Figure 3.8 C). Different classes of transcription factors, as well as both structural and motor proteins found in cilia, are downregulated upon disruption of NFY function (Figure 3.7 D).

Since disruption of either the TALE factors or NFY produces embryos with shared phenotypes, I next identified genes whose expression is dependent on both TALE and NFY function. I find that there are 201 such genes (74 downregulated, 127 upregulated; Figure 3.5 E). Strikingly, the annotation of genes downregulated upon disruption of both TALE and NFY function identifies transcriptional and developmental roles, but not roles associated with cilia, though several terms associated with tubulin function are retained (Figure 3.7 E, F; GO terms for upregulated genes are shown in Figure 3.8 D). I refer to this set of genes as 'TALE/NFY co-regulated genes' and note that it is a relatively small population, such that about 23% (74/325) of the NFY-dependent genes are also dependent on TALE function (Figure 3.5 E). These results indicate that TALE and NFY co-regulate a set of genes with functions that are distinct from those of genes regulated by either TALE or NFY. Accordingly, a GO term analysis of genes regulated exclusively by NFY revealed strong enrichment for cilia genes

(Figure 3.9 B), while genes exclusively dependent on TALE function return GO terms enriched for transcriptional regulation such as *hox* genes (Figure 3.9 A). Hence, in addition to TALE and NFY co-regulating a subset of transcriptional and developmental control genes, NFY controls expression of cilia-related genes independently of the TALE factors.
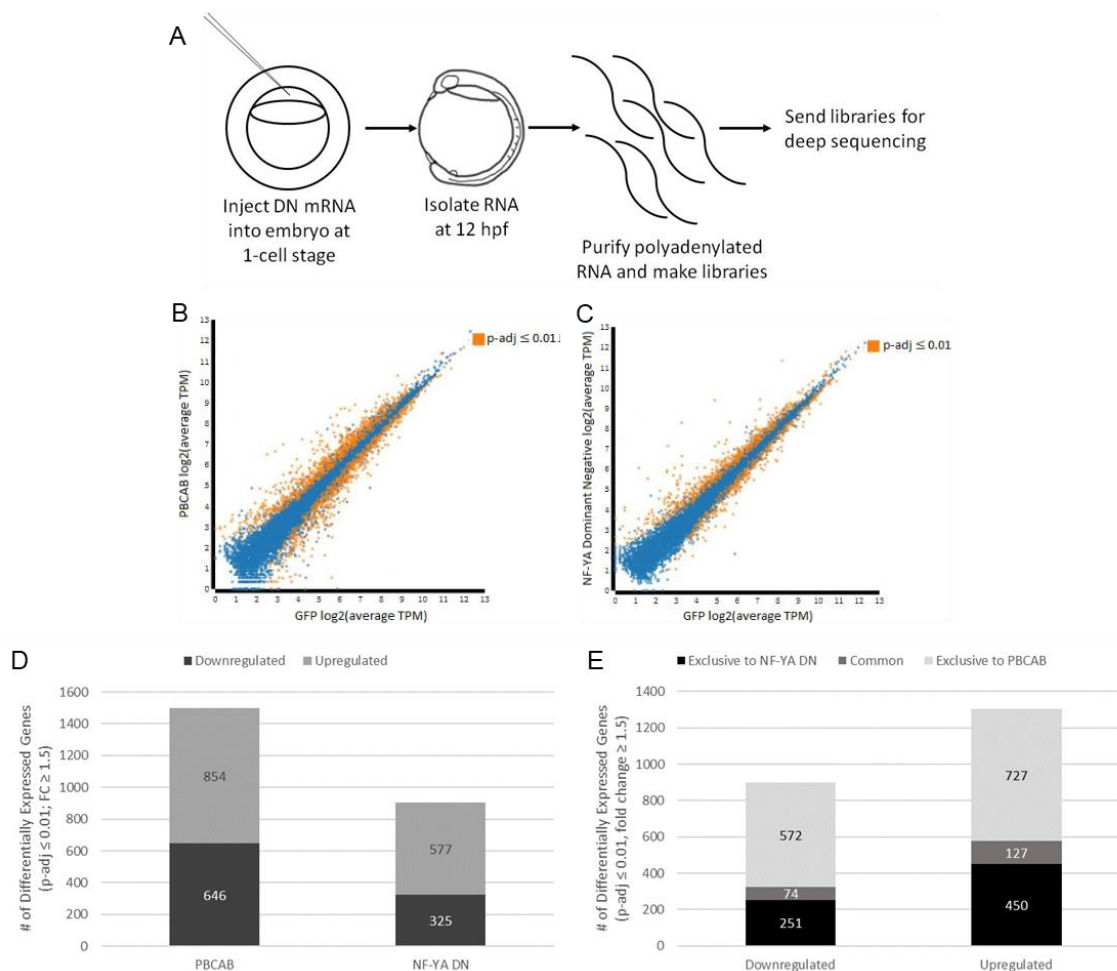
Figure 3.5: Dominant negatives disrupt gene expression. (A) Schematic of RNA-seq experiments. (B-C) Scatterplots of gene expression in PBCAB- vs GFP-injected (B) and Nfya-DN- vs GFP-injected (C) zebrafish embryos (expression presented as log2 of average TPM for multiple replicates; see methods). Expression of genes highlighted in orange is significantly different at 12 hpf (p-adj ≤ 0.01; Wald test in DESeq2). (D) Number of genes differentially expressed in PBCAB- (left) or Nfya-DN-injected (right) embryos relative to GFP-injected embryos (p-adj ≤ 0.01; fold-change ≥ 1.5). (E) Breakdown of downregulated (left) and upregulated (right) genes exclusive or common to each experimental condition.

Library	Total Reads	rRNA Reads	% rRNA	Reads - rRNA	Reads Aligned (RSEM)
NFYA DN1	25,570,095	118,566	0.46%	25,451,529	15,618,541
NFYA DN2	25,627,400	90,013	0.35%	25,537,387	15,884,091
NFYA DN3	25,648,288	103,952	0.41%	25,544,336	15,847,491
PBCAB1	22,951,454	6,211	0.03%	22,945,243	14,084,935
PBCAB2	25,649,514	7,074	0.03%	25,642,440	15,668,814
PBCAB3	25,660,266	7,991	0.03%	25,652,275	15,605,294
GFP1	25,565,467	4,756	0.02%	25,560,711	16,034,166
GFP2	25,488,663	4,725	0.02%	25,483,938	15,765,669
GFP3	25,696,572	6,392	0.02%	25,690,180	15,980,155

Table 3.4: RNA-seq read counts. RNA-seq library read counts, ribosomal RNA reads, rRNA percentage, read counts after rRNA filtering, and reads aligned.
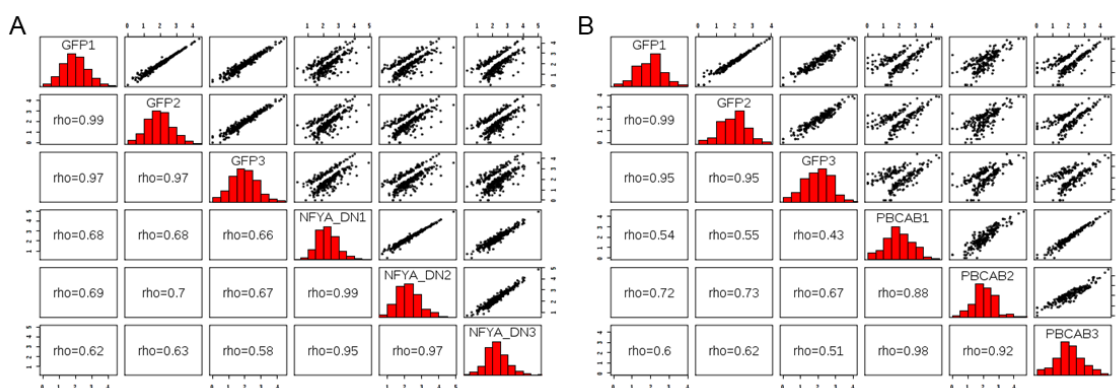
Figure 3.6: Comparison of RNA-seq biological replicates. Histograms, scatter plots, and Spearman's rank correlation coefficient comparing each biological replicate of PBCAB with GFP (A) or Nfya-DN with GFP (B).
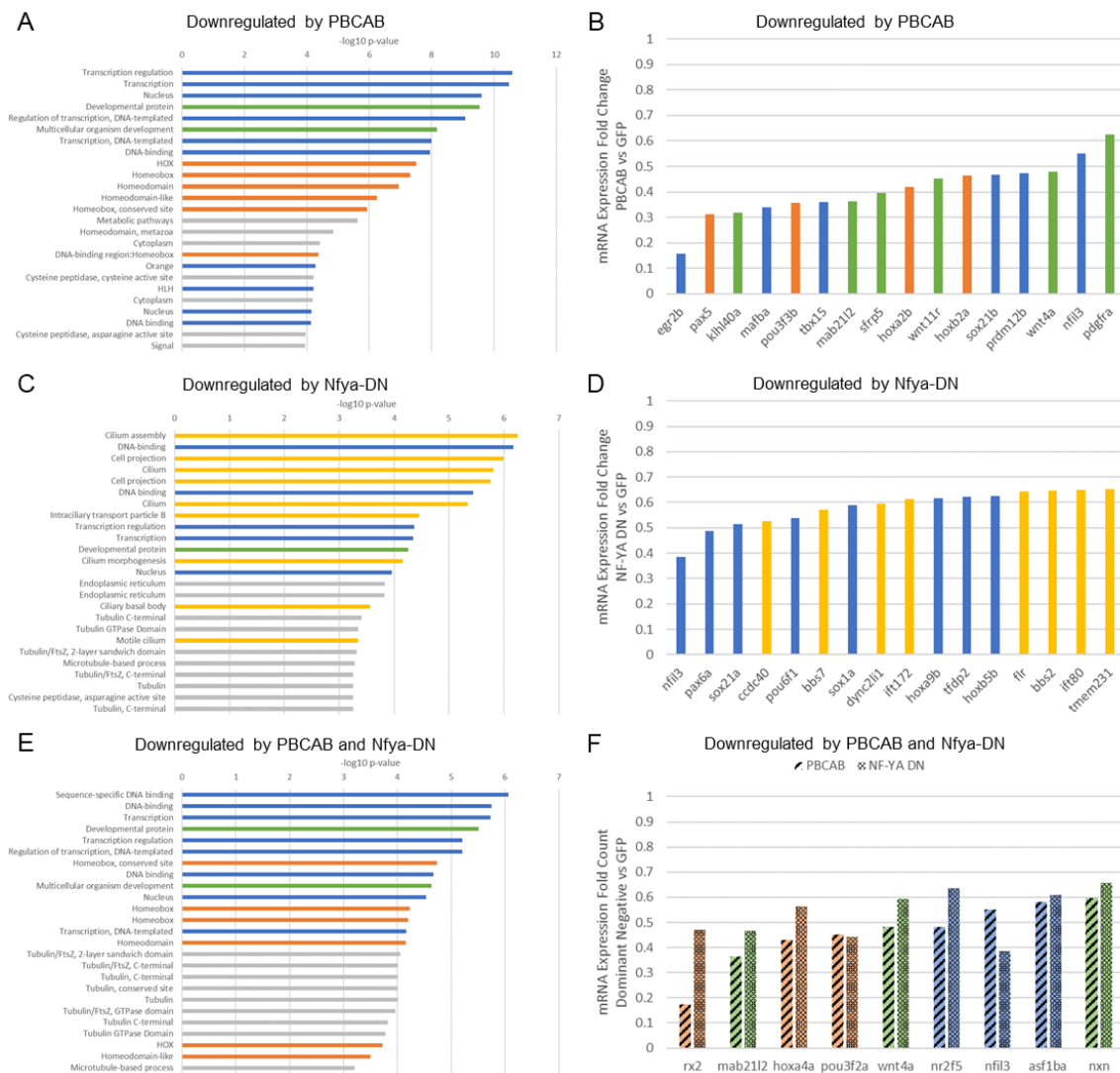
Figure 3.7: The TALE factors and NFY occupy genomic sites associated with developmental and transcriptional control genes. (A, C, E) DAVID analyses showing the 25 most significant GO terms (EASE Score) associated with genes downregulated by PBCAB (A), Nfya-DN (C), and common to both (E). Blue bars correspond to transcription-related, green to embryogenesis-related, orange to homeodomain-related, yellow to cilia-related, and gray to other ontologies. (B, C, F) Selected genes downregulated by PBCAB (B), Nfya-DN (D), or both (F). Color coding is the same as in (A, C, E).
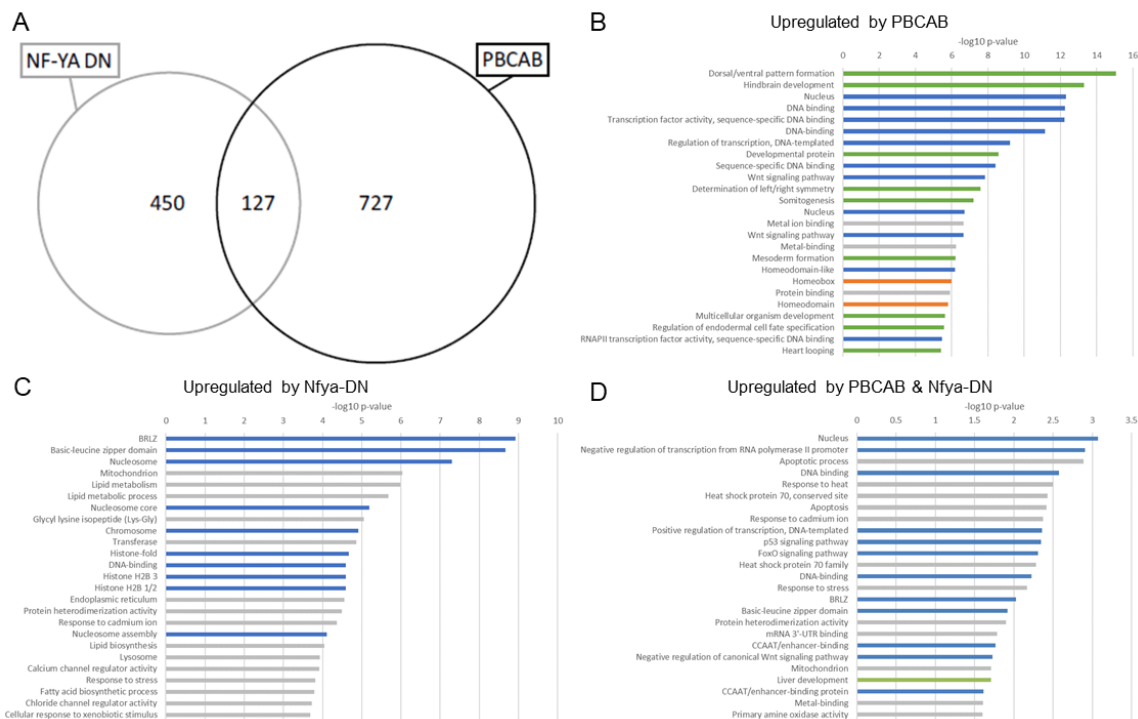
Figure 3.8: Genes upregulated by disruption of TALE and NFY function. (A) Venn diagram showing upregulated genes (p-adj ≤ 0.01; FC ≥ 1.5) in embryos injected with PBCAB or Nfya-DN. (B-D) GO terms associated with genes upregulated (p-adj ≤ 0.01, FC ≥ 1.5) by PBCAB (B), upregulated by Nfya DN (C), or upregulated by both PBCAB and Nfya-DN (D). Blue bars correspond to transcription-related, green to embryogenesis-related, orange to homeodomain-related, yellow to cilia-related, and gray bars to other ontologies.
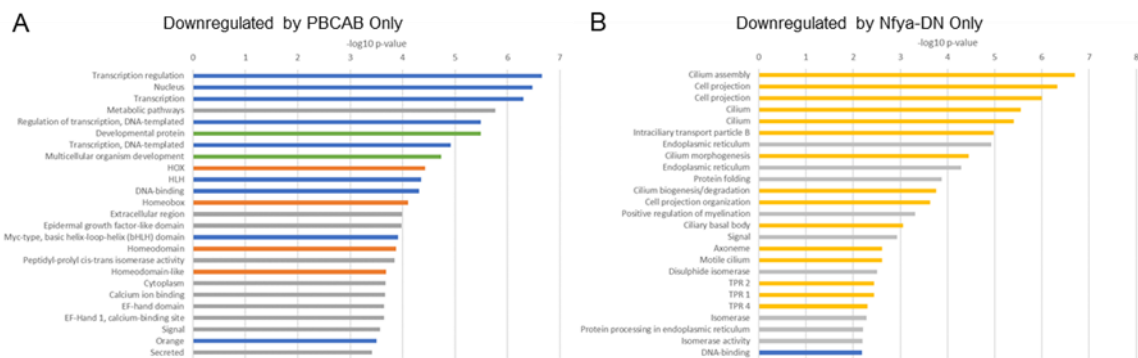
Figure 3.9: PBCAB and Nfya-DN disrupt genes independently. (A-B) GO terms of genes downregulated exclusively by PBCAB (A) or Nfya-DN (B). Blue bars correspond to transcription-related, green to embryogenesis-related, orange to homeodomain-related, yellow to cilia-related, and gray bars to other ontologies.

TALE and NFY occupy genomic sites associated with developmental and transcriptional control genes

Given that the TALE factors and NFY appear to have both shared and independent functions, I next examined binding of these transcription factors across the zebrafish genome. In order to determine if these transcription factors have a role during the onset of zygotic gene expression, I focused our analysis at 3.5 hpf, when zygotic genes are becoming active in zebrafish embryos. We previously used ChIP-seq to characterize Prep1 occupancy and found that this transcription factor is bound at many genomic elements at maternally controlled stages (3.5 hpf and earlier; (Choe et al., 2014; Ladam et al., 2018)), consistent with previous reports that the TALE factors are maternally transmitted in zebrafish (Choe et al., 2002; Deflorian et al., 2004; A. J. Waskiewicz et al., 2002). Specifically, our analysis identified a 10 bp motif (TGATTGACAG; termed the 'DECA motif') as the predominant element occupied by Prep1 in 3.5 hpf zebrafish embryos. The DECA motif contains two half-sites: one for Pbx proteins (TGAT) and one for Prep proteins (TGACAG). Pbx proteins form dimers with Prep proteins (reviewed in (Ladam & Sagerström, 2014)) and, using ChIP-qPCR, we previously demonstrated that zebrafish Pbx4 occupied 11 of 12 tested DECA sites in 3.5 hpf zebrafish embryos (Ladam et al., 2018). We have now extended this analysis to the entire zebrafish genome by performing ChIP-seq for Pbx4 at 3.5 hpf (Figure 3.10 A, B; Figure 3.11 A; Table 3.5). We find that the majority of

Pbx4 peaks overlap with a Prep1 peak (94% overlap at FE > 10; Figure 3.10 A,

B; Figure 3.11 D; Table 3.6; Table 3.7) and that the predominant sequence motif

at Pbx4 binding sites is indistinguishable from the DECA motif observed at

Pbx4:Prep1 co-occupied sites (Figure 3.11 C, F). We also find that the

distribution of Pbx4 peaks relative to TSSs is similar to that for Prep1 (Figure

3.10 C), with about 50% of all binding sites located within 30 kB of an annotated

promoter element (Nepal et al., 2013) in both cases. Accordingly, sites co-

occupied by both Pbx4 and Prep1 show a similar distribution (Figure 3.10 C). GO

term analyses revealed that genes associated with Pbx4 binding sites are

enriched for functions related to transcriptional regulation and embryogenesis

(Figure 3.11 B). This is similar to the functions we previously identified for genes

associated with 3.5 hpf Prep1 bound sites (Ladam et al., 2018). As expected,

genes associated with Prep1 and Pbx4 co-occupied sites return essentially the

same GO terms (Figure 3.11 E). Notably, many Prep1 binding sites do not

overlap with Pbx4 peaks (Figure 3.11 D). While this could indicate that Prep1 has

functions independent of Pbx4, it may also reflect different affinities of the two

antisera. Nevertheless, my observations indicate that Pbx4 binds primarily at

DECA sites in the context of Pbx:Prep heterodimers at this stage of

embryogenesis. I will focus on these Prep and Pbx co-occupied sites and will

refer to them as 'TALE sites.'

We previously reported that approximately 30% of Prep1-occupied DECA

sites observed at 3.5 hpf have a CCAAT motif nearby – usually at a distance of

about 20 bp (Ladam et al., 2018). In other systems, such CCAAT boxes serve as binding sites for the heterotrimeric NFY transcription factor. Since NFY is maternally deposited in zebrafish (Chen et al., 2009), we previously used ChIP-qPCR to test 15 CCAAT boxes located near DECA sites and found that nine were occupied by NFY (Ladam et al., 2018). However, the commercial antiserum we used for the ChIP-qPCR experiment is too low affinity for ChIP-seq. Therefore, in order to examine NFY binding genome-wide, we raised antiserum to zebrafish Nfya, which is the sequence-specific DNA binding component of the NFY heterotrimer, and carried out ChIP-seq for NFY on 3.5 hpf zebrafish embryos (Figure 3.10 A and B; Figure 3.11 G; Table 3.5). As expected, NFY-occupied genomic sites are highly enriched for the CCAAT box sequence motif (Figure 3.11 I), but I find that the distribution of NFY peaks in the genome is somewhat different than the distribution of TALE peaks, such that NFY appears to be preferentially bound closer to promoters (Figure 3.10 C). Further, I find that NFY-bound genomic elements are also associated with genes enriched for functions related to transcriptional regulation and embryogenesis (Figure 3.11 H). These functions are similar to those observed for genes associated with TALE-occupied elements, in support of the idea that NFY functions together with TALE transcription factors at this stage of development.

To further address the potential cooperation between the TALE factors and NFY, I next examined co-localization of TALE and NFY proteins in the zebrafish genome. I find that approximately 22% of the NFY-occupied sites

overlap with a TALE-occupied site (corresponding to 17% of the TALE-bound sites; Figure 3.10 A, B; Figure 3.11 J; Table 3.6; Table 3.7). Strikingly, motif analyses identified a roughly 27 bp sequence motif encompassing both a DECA motif and a CCAAT box (Figure 3.11 L) associated with the TALE/NFY co-occupied sites, while sites occupied by TALE alone display a DECA motif (Figure 3.11 M) and those occupied by NFY alone contain a CCAAT box (Figure 3.11 N). GO terms for genes associated with co-occupied sites are again enriched for functions related to transcriptional control, but less so for functions controlling embryogenesis (Figure 3.11 K). Lastly, sites co-occupied by TALE and NFY factors show high association with promoters (Figure 3.10 C). I conclude that, in 3.5 hpf zebrafish embryos, the TALE factors and NFY individually occupy genomic regions associated with both developmental and transcriptional regulators, but also co-occupy an extended binding motif that appears more selectively associated with transcriptional control genes.

Replicate	Total Peaks	FE≥10 Peaks	Common Peaks	FE≥10 Sum Peaks	Associated Genes (GREAT)
Pbx4-1	21,104	5,021			
Pbx4-2	34,119	3,640	3,427	5,234	5,955
Prep1-1	12,256	2,058			
Prep1-2	47,831	13,328	2,058	13,342	9,798
NF-YA1	21,784	3,056			
NF-YA2	22,196	3,252	2,588	3,720	4,398

Table 3.5: ChIP-seq read counts. Data for Pbx4 and Nfya ChIP-seq biological replicates with Prep1 ChIP-seq data (Ladam et al., 2018) included as reference.

ChIP-seq Data Sets	Overlapping Peaks	Associated Genes (GREAT)
Pbx4/Prep1	4,907	5,701
Pbx4/NF-YA	834	1,183
Prep1/NF-YA	937	1,332
Pbx4/Prep1/NF-YA	820	1,161

Table 3.6: ChIP-seq peak overlaps. Number of peaks that overlap (at least 1 bp shared between 200 bp fragments centered on peaks) between Prep1, Pbx4 and Nfya ChIP-seq data sets. Only peaks with a 10-fold or greater enrichment over input are considered.

ChIP-seq Overlap	FE≥4	FE≥10	Top 10%
Pbx4/Prep1	74.8% (13,836/18,508)	93.8% (4,907/5,234)	85.7% (2,960/3,455)
TALE/NF-YA	13.2% (2,014/15,270)	22.0% (820/3,720)	23.5% (612/2,599)

Table 3.7: ChIP-seq fold enrichment cutoffs. Extent of overlap of Pbx4 peaks with Prep1 peaks and TALE peaks with Nfya peaks at three different cutoffs (FE ≥ 4, FE ≥ 10 and top 10% of peaks).
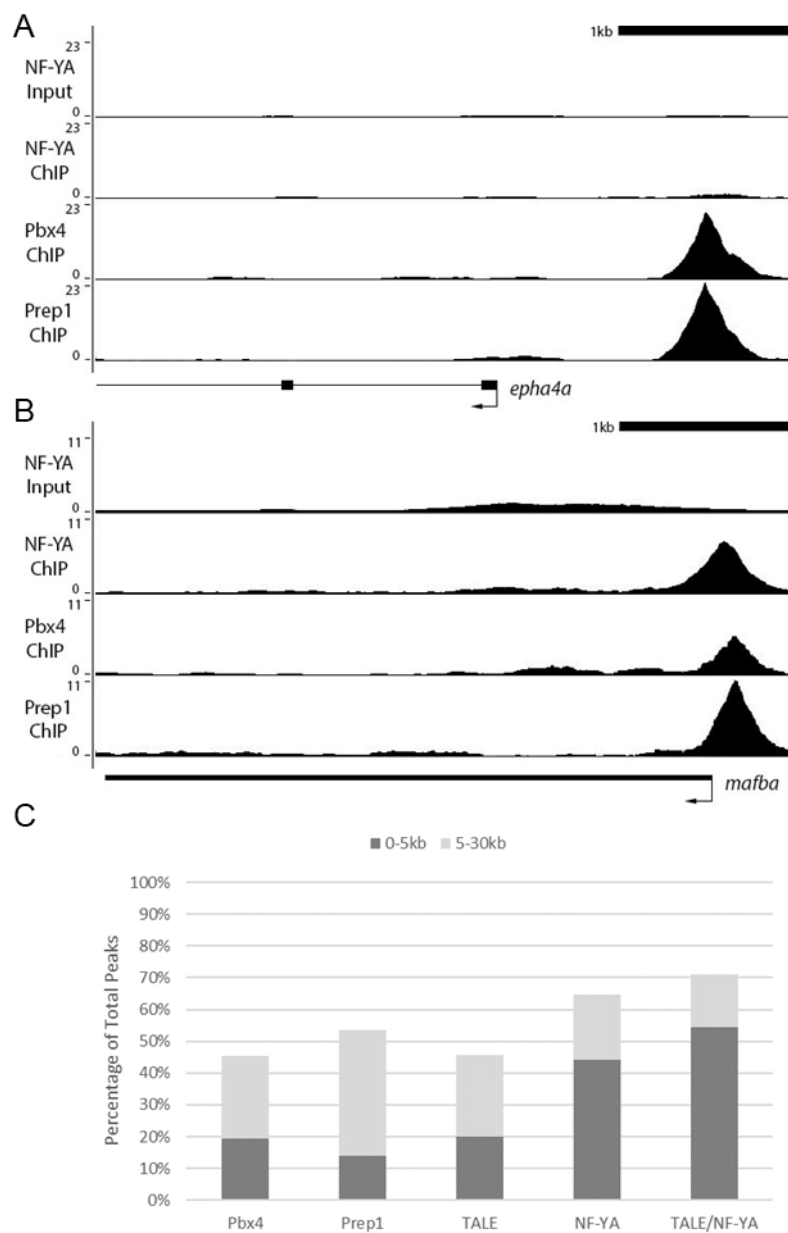
Figure 3.10: The TALE factors and Nfya occupy genomic sites near promoters. (A-B) Representative UCSC Genome Browser tracks for Nfya, Pbx4 and Prep1 ChIP-seq analyses at 3.5 hpf. (C) Chart showing percent of ChIP-seq peaks found within 5 kb or 30 kb of a promoter.
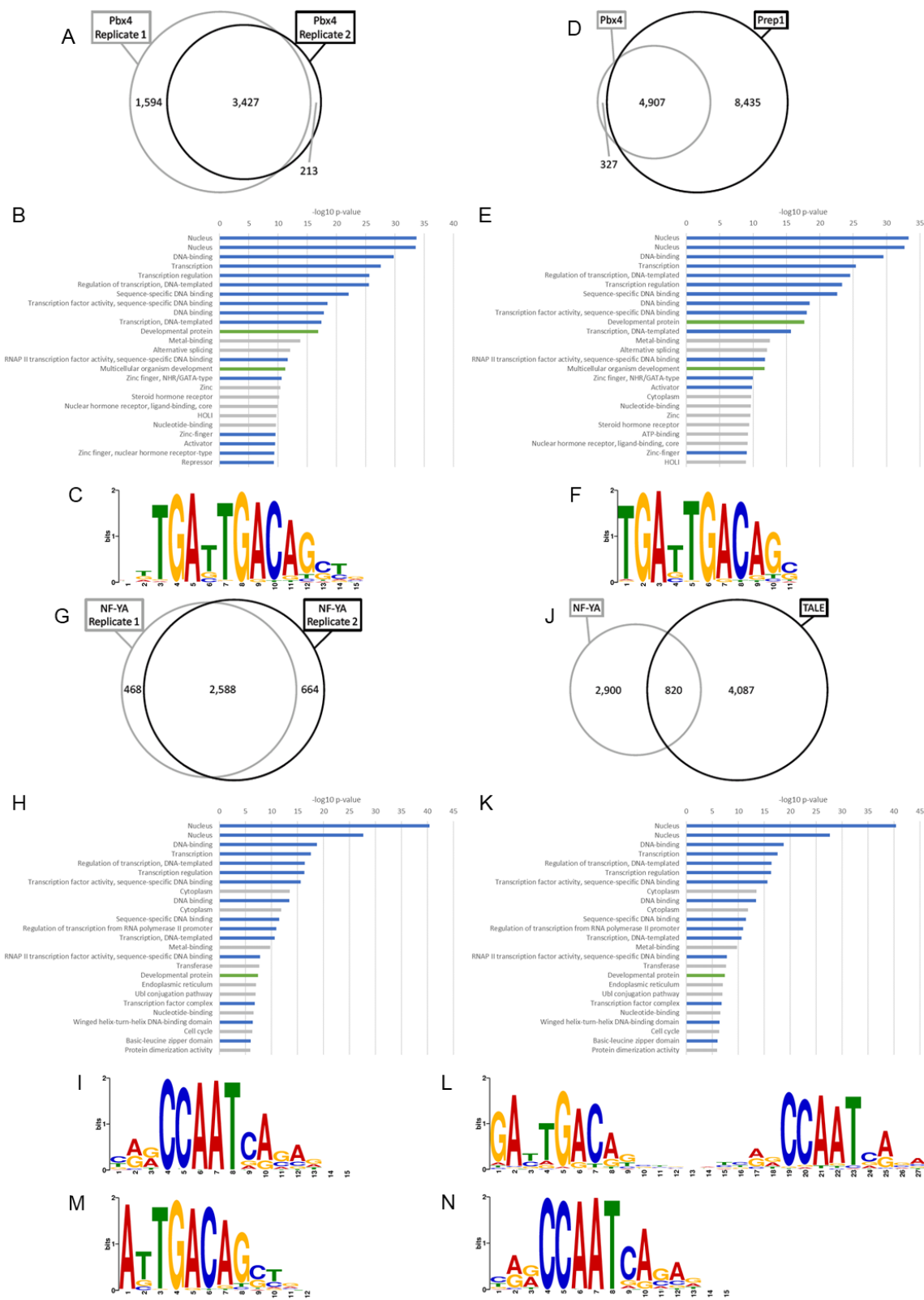
Figure 3.11: The TALE factors and NFY occupy genomic sites associated with developmental and transcriptional control genes. (A, D, G, J) Venn diagrams showing the overlap (at least 1 bp shared between 200 bp fragments centered on peaks) of two Pbx4 ChIP-seq replicates (A), the overlap of Pbx4 and Prep1 ChIP-seq peaks (D), the overlap of two Nfya ChIP-seq replicates (G) and the overlap of TALE and NFY ChIP-seq peaks (J). (B, E, H, K) The top sequence motif returned by MEME for Pbx4-occupied sites (B), Pbx4/Prep1 co-occupied sites (E), Nfya occupied sites (H) and TALE/Nfya co-occupied sites (K). (C, F, I, L) The top 25 gene ontology (GO) terms returned by the GREAT analysis tool for genes associated with Pbx4-occupied sites (C), Pbx4/Prep1 co-occupied sites (F), Nfya occupied sites (I) and TALE/Nfya co-occupied sites (L). (M, N) Top sequence motif returned by MEME for peaks bound by TALE, but not Nfya (M) and peaks bound by Nfya, but not TALE (N). Only peaks with a 10-fold or greater enrichment over input (FE ≥ 10) were considered for the analyses.

TALE and NFY co-regulate a subset of early-expressed transcriptional regulators

While my RNA-seq analysis identified genes regulated by the TALE factors and NFY, it is neither clear how direct this regulation might be nor evident whether some of these genes are activated at the ZGA. To begin addressing these questions, I first examined whether TALE-dependent genes are associated with genomic elements bound by the TALE factors. I find that, of the 646 genes I identified as being TALE-dependent, 52% (335/646) are found near (as defined using default parameters in GREAT; see Chapter II: Materials & Methods) a TALE occupied element (Table 3.8; Figure 3.12 A), and these genes are enriched for functions related to embryonic development and transcriptional regulation with a specific emphasis on *hox* genes (Figure 3.12 B). Similarly, of the 325 genes my RNA-seq analysis showed to be downregulated upon disruption of NFY function, I find that 61% (199/325) are near an NFY-occupied element (Table 3.8; Figure 3.12 A). The GO terms for these genes are enriched for functions related to transcriptional regulation, as well as for cilia structure and function (Figure 3.12 C). Thus, 50-60% of TALE- and NFY-dependent genes are associated with a binding site for the corresponding transcription factor and the functional annotations of these genes show specific enrichment for the same terms as I observed in my RNA-seq analysis, which include Hox transcription factors for TALE-dependent genes and cilia structure and function factors for

NFY-dependent genes. Since initial cilia formation and Hox activity occurs at gastrula and segmentation stages in zebrafish (Essner et al., 2002; Prince et al., 1998a; Prince et al., 1998b), these results indicate that the TALE factors and NFY control separate gene expression programs by 12 hpf of zebrafish development.

In order to begin assessing co-regulation by the TALE factors and NFY, I carried out the reciprocal analysis. In doing so, I find that 55% of the 646 TALE-dependent genes are associated with an NFY occupied site (358/646) and 49% of the 325 NFY dependent genes are associated with a TALE-occupied element (158/325), indicating that the TALE factors and NFY co-regulate a subset of their target genes (Table 3.8; Figure 3.12 A). To examine this co-regulation further, I next focused specifically on the TALE/NFY co-regulated genes defined in Figure 3.5 E and Figure 3.7 E and F. I find that of the 74 genes in this category, 70% are associated with an NFY- (52/74) occupied site and 50% (37/74) with a TALE-occupied site (Table 3.8; Figure 3.12 A). Indeed, 49% of co-regulated genes are found near both TALE- and NFY-occupied sites (36/74). Strikingly, the GO terms of co-regulated genes associated with both TALE- and NFY-occupied elements converge on functions related to transcriptional regulation and embryonic development (Figure 3.12 D). Further, if I specifically focus on genes that are near regulatory elements with overlapping TALE/NFY peaks (as defined in Figure 3.11 J, K), I find that they function in transcription and regulation of development, but the categories related to cilia and homeobox functions are no longer

represented (Figure 3.12 E). Importantly, broad expression of transcriptional and

developmental control genes is the first event to take place after MZT.

Accordingly, by analyzing RNA-seq data collected at 6 hpf, which is shortly after

the MZT, I find that genes associated with both TALE- and NFY-occupied

elements are expressed at higher levels than genes that lack such an association

(Figure 3.12 F). Hence, my results indicate that genes co-regulated by TALE and

NFY act uniquely in transcriptional control of embryogenesis shortly after MZT,

while TALE and NFY each controls a distinct gene expression program at later

stages.

ChIP-seq	TALE-Dependent	NF-Y-Dependent	TALE/NF-Y-Dependent
TALE	51.9% (335/646)	48.6% (158/325)	50% (37/74)
NFYA	55.4% (358/646)	61.2% (199/325)	70.3% (52/74)
TALE/NF-Y	40.9% (264/646)	41.5% (135/325)	48.6% (36/74)

Table 3.8: Overlap of TALE- and NFY-dependent genes with ChIP peaks. Summary of the correlation between TALE and/or NFY-dependent genes and binding by the corresponding transcription factor at a nearby site. (ChIP peaks enriched by 4-fold or greater over input were considered).
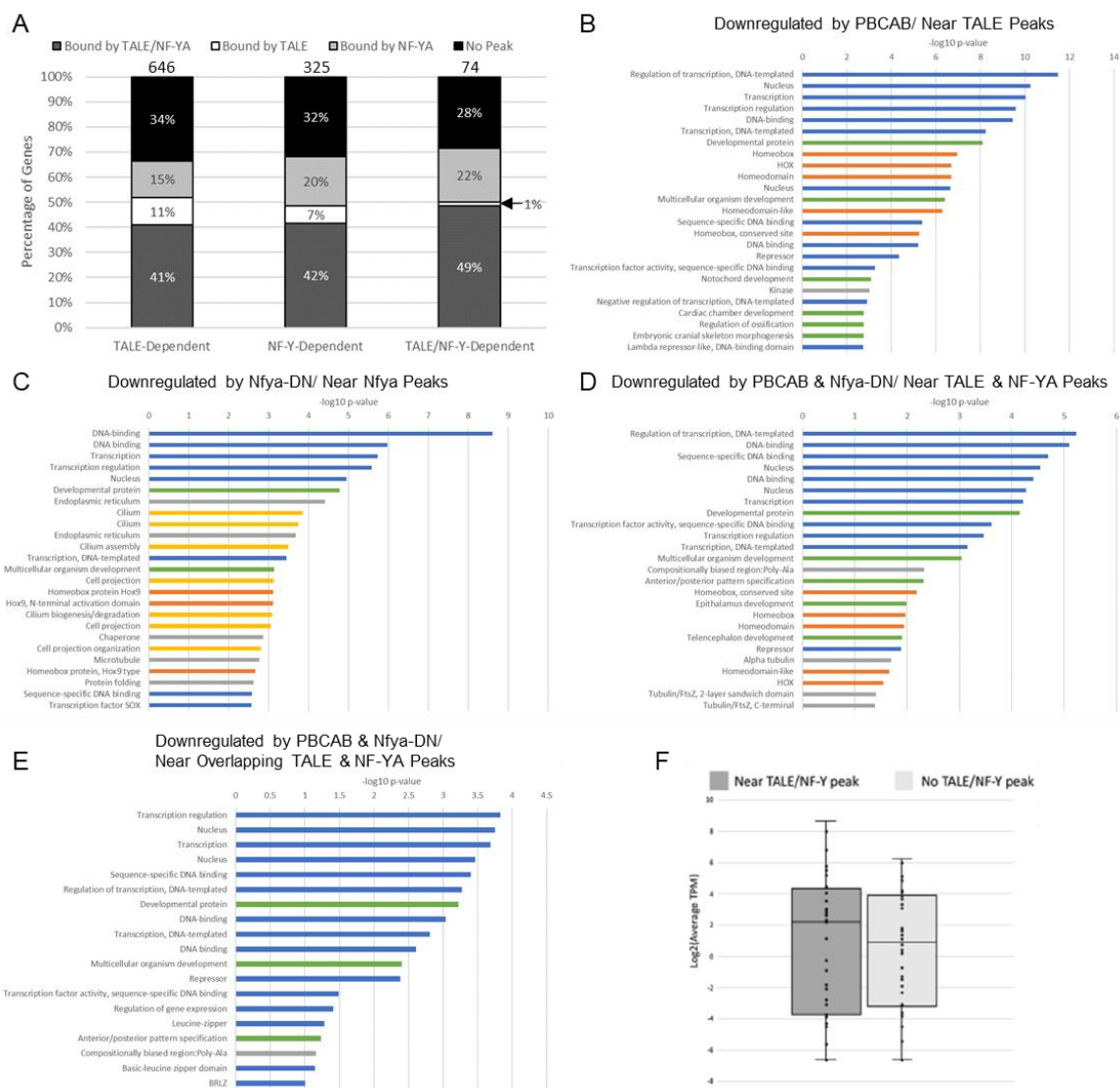
Figure 3.12: TALE and NFY co-regulate a subset of early-expressed transcriptional regulators. (A) Graphical breakdown of TALE and/or NFY occupancy near TALE and/or NFY-dependent genes. (B-E) Top GO terms returned by DAVID for TALE-dependent genes associated with TALE peaks (B), NFY dependent genes associated with Nfya peaks (C), TALE/NFY-dependent genes associated with both TALE and Nfya peaks (D), and TALE/NFY-dependent genes associated with overlapping TALE and Nfya peaks (E). Blue bars correspond to transcription-related, green to embryogenesis-related, orange to homeodomain-related, yellow to cilia-related, and gray bars to other ontologies. (F) Box chart showing expression levels as log2(average TPM) for TALE/NFY-dependent genes with and without nearby TALE/Nfya peaks.

Genomic elements co-occupied by TALE and NFY act as enhancers *in vivo*

Although my data show that many genomic elements co-occupied by

TALE and NFY are found near promoters, the TALE factors often act at

enhancers (Ferretti et al., 2005; Ferretti et al.,2000; Grieder et al., 1997; Jacobs

et al., 1999; Pöpperl et al., 1995; Ryoo & Mann, 1999; Tümpel et al., 2007).

Further, while NFY was originally identified as acting at promoters (reviewed in

(Maity & de Crombrugghe, 1998)), more recent work revealed an important role

for NFY at tissue-specific enhancers (Oldfield et al., 2014). To explore these

relationships in greater detail, I examined the chromatin state at TALE/NFY co-

occupied elements and found that both H3K4me1 and H3K27ac, which mark

enhancers and promoters, are highly enriched already at 4.5 hpf (at ZGA) and

persist at 9 hpf (Figure 3.13 A-D). I note that elements bound by NFY alone and,

to a lesser extent, TALE alone have the same characteristics. In agreement with

TALE/NFY co-occupied elements driving gene expression at this stage of

development, I also find a dramatic increase in H3K4me3 modifications (a mark

of active promoters) between 4.5 hpf and 9 hpf (Figure 3.13 E, F).
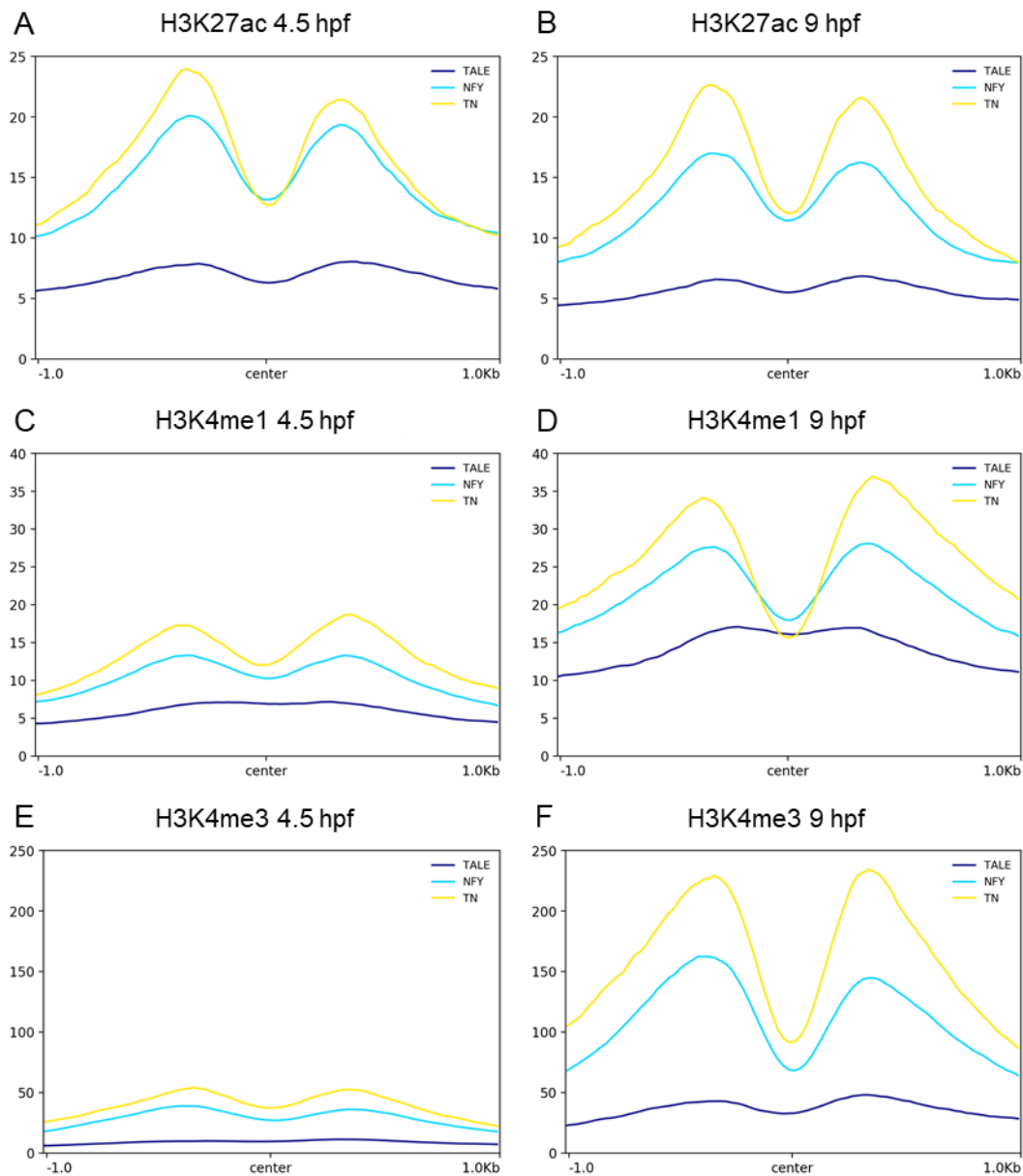
Figure 3.13: Genomic elements co-occupied by the TALE factors and NFY contain enhancer chromatin marks. Average histone mark signals at genomic regions containing only TALE peaks (dark blue), only Nfya peaks (light blue), or TALE/Nfya peaks (yellow) for H3K27ac at 4.5 hpf (A) and 9 hpf (B), H3K4me1 at 4.5 hpf (C) and 9 hpf (D), or H3K4me3 at 4.5 hpf (E) and 9 hpf (F).

To directly test if TALE/NFY co-occupied elements act as enhancers *in vivo*, I used a previously published enhancer assay (Li et al., 2010) and inserted individual genomic elements upstream of the E1b minimal promoter and the GFP reporter gene. I selected eight genomic elements that contain adjacent TALE/NFY motifs (as in figure 3.11 K) and that are associated with genes expressed in the anterior embryo (Figure 3.14 A, C, E, G, I; Figure 3.15 A-C) and used these to generate transgenic zebrafish. Of the eight constructs, named for the identity of the nearest gene, five showed expression in the F0 generation and GFP-positive embryos were raised to generate stable lines (summarized in Table 3.9). The remaining three constructs did not show F0 expression and I did not consider them further. In stable lines for each of the five constructs, I detected tissue-restricted GFP expression with each construct producing a distinct pattern (Figure 3.14 B, D, F, H, J). I screened at least two independent founders for each stable line and find that GFP expression is indistinguishable between founders carrying the same construct (Figure 3.16 C, D; Table 3.9), indicating that each element imparts a unique tissue specificity to the basal E1b-GFP reporter that is independent of its integration site. In some instances, the observed expression pattern is comparable to that of the nearest gene (e.g. *fgf8a*; Figure 3.14 H), suggesting that it represents an enhancer element controlling expression of the nearby gene. In other instances, the enhancer drives expression in a novel pattern (e.g. *yap1*; Figure 3.14 J), suggesting that it may control a gene further away, or that the enhancer element tested, which is about 500 bp in length, lacks

some inputs required for proper expression of the nearby gene. These results

indicate that TALE/NFY co-occupied elements act as enhancers *in vivo*.

Element	Coordinates	Size	NF-YA Peak	Pbx4 Peak	Prep1 Peak	Cloned in E1b::GFP	GFP+ Founders	GFP+ F1 0 hpf	GFP+ F1 24 hpf
tcf3a	chr2:57244797-57245296	500bp	Strong	Strong	Strong	Yes	♀ #3	11/26 (42.3%)	11/24 (45.8%)
							♂ #4	0/84 (0.0%)	20/84 (23.8%)
							♂ #7	0/333 (0.0%)	13/185 (7.0%)
tle3a	chr18:20235689-20236181	493 bp	Strong	Strong	Strong	Yes			
dachb	chr1:9669399-9669898	500 bp	Moderate	Strong	Strong	Yes	♂ #6	0/364 (0.0%)	26/308 (8.4%)
							♂ #8	0/390 (0.0%)	12/359 (3.3%)
							♀ #15	25/198 (12.6%)	25/174 (14.4%)
							♀ #17	64/451 (14.2%)	64/436 (14.7%)
fgf8a	chr13:28356309-28356808	500 bp	Moderate	Moderate	Moderate	Yes	♀ #1	59/92 (64.1%)	59/88 (67.0%)
							♂ #10	0/77 (0.0%)	31/72 (43.1%)
yap1	chr18:37345724-37346223	500 bp	Weak	Moderate	Moderate	Yes	♂ #4	0/221 (0.0%)	33/156 (21.2%)
							♂ #5	0/313 (0.0%)	47/218 (21.6%)
							♂ #11	0/25 (0.0%)	3/21 (14.3%)
pax5	chr1:21037015-21037514	500 bp	Moderate	Strong	Strong	Yes	0		
her6	chr6:36600313-36600812	500 bp	Weak	Moderate	Strong	Yes	0		
prdm14	chr24:14342429-14342928	500 bp	None	Weak	Moderate	Yes	0		

Table 3.9: Putative enhancer details. Summary of information about each putative enhancer element.

A — tcf3a
0.5 kb
NF-YA Input (35–0)
NF-YA ChIP (35–0)
Pbx4 ChIP (35–0)
Prep1 ChIP (35–0)

B

C — tle3a
0.5 kb
NF-YA Input (38–0)
NF-YA ChIP (38–0)
Pbx4 ChIP (38–0)
Prep1 ChIP (38–0)

D

E — dachb
0.5 kb
NF-YA Input (37–0)
NF-YA ChIP (37–0)
Pbx4 ChIP (37–0)
Prep1 ChIP (37–0)

F

G — fgf8a
0.5 kb
NF-YA Input (5–0)
NF-YA ChIP (5–0)
Pbx4 ChIP (5–0)
Prep1 ChIP (5–0)

H

I — yap1
0.5 kb
NF-YA Input (5–0)
NF-YA ChIP (5–0)
Pbx4 ChIP (5–0)
Prep1 ChIP (5–0)

J
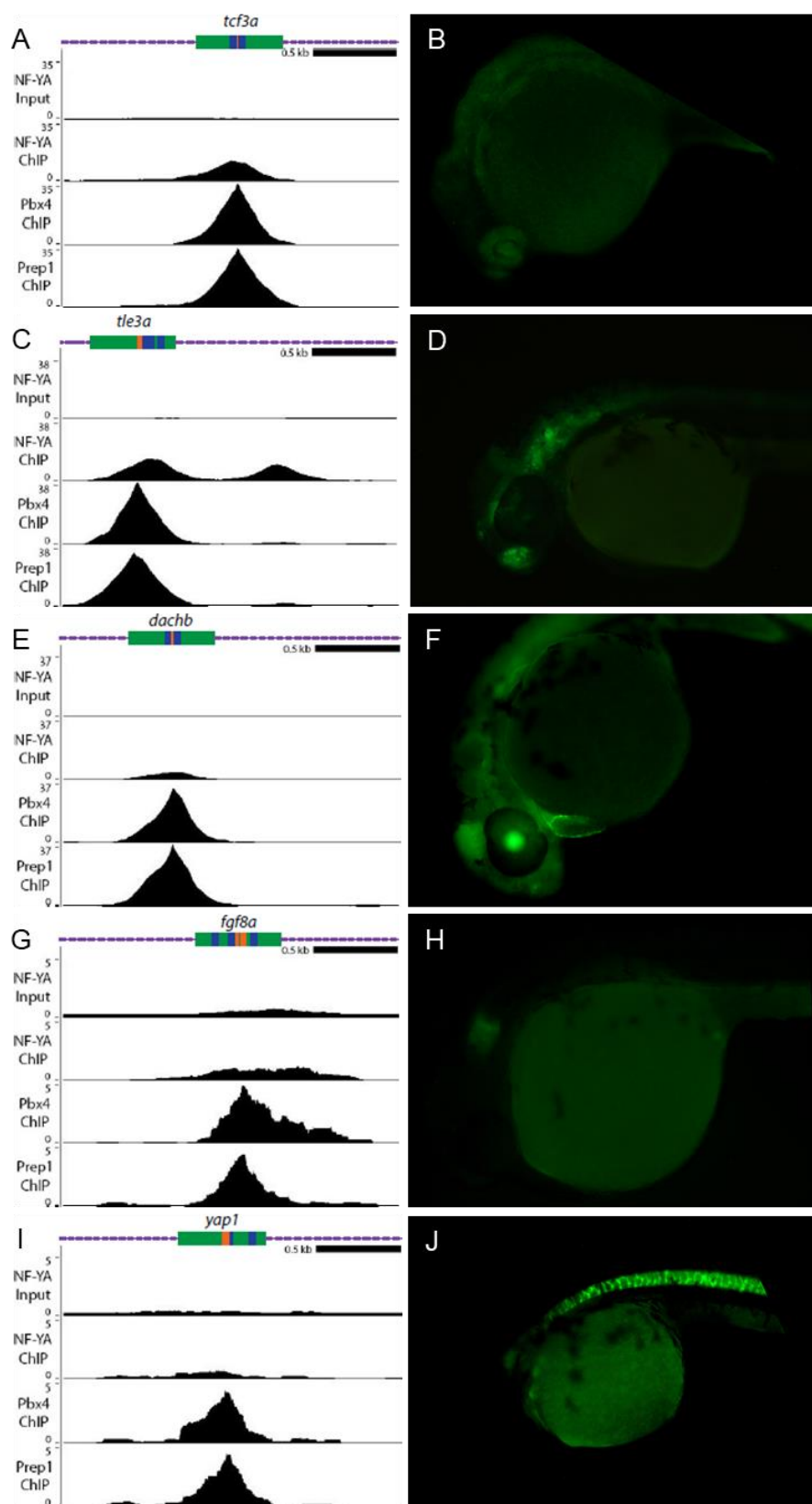
Figure 3.14: Genomic elements co-occupied by the TALE factors and NFY act as enhancers *in vivo*. (A, C, E, G, I) UCSC Genome Browser tracks showing Nfya, Pbx4, and Prep1 ChIP-seq data for the *tcf3a* (A), *tle3a* (C), *dachb* (E), *fgf8a* (G) and *yap1* (I) loci. The diagrams above the tracks show the putative enhancer region in green, DECA sites in orange, and CCAAT boxes in blue. (B, D, F, H, J) GFP expression in 24 hpf F1 *tcf3a*::E1b-GFP (H), *tle3a*::E1b-GFP (J), *dachb*::E1b-GFP (L), *fgf8a*::E1b-GFP (N) and *yap1*::E1b-GFP (P) transgenic embryos resulting from crosses between male founders and wild-type females.

I next took two approaches to confirm that the observed expression patterns are dependent on TALE and NFY function. First, I expressed the dominant negative TALE and NFY constructs in embryos from a cross of the F1 offspring of the *tcf3a*::E1b-GFP transgenic line (Figure 3.18 A). I find that GFP expression is dramatically reduced in embryos expressing either dominant negative construct (Figure 3.18 B-E), indicating that expression from the *tcf3a* genomic element requires both TALE and NFY function. This observation was further confirmed by RT-qPCR analysis (Figure 3.18 F). Second, I made use of a distinct transgenesis strategy that allows me to test the effect of mutating the TALE and NFY binding sites in a given enhancer element. Specifically, my transgenic construct includes the ɣ-crystallin promoter driving GFP along with the candidate enhancer element driving RFP. I find that the wild-type *tcf3a* and *tle3a* enhancers drive tissue-specific RFP expression (Figure 3.19 A, E), as expected based on my results in Figure 3.14. However, when I test mutated versions of these elements (where the TALE and NFY binding sites have been disrupted), I find that transgenic animals (defined by GFP expression in the eye; Figure 3.19 D, H) lack RFP expression (Figure 3.19 C, G). I conclude that TALE/NFY co-occupied elements possess enhancer activity and that this activity requires TALE and NFY function.
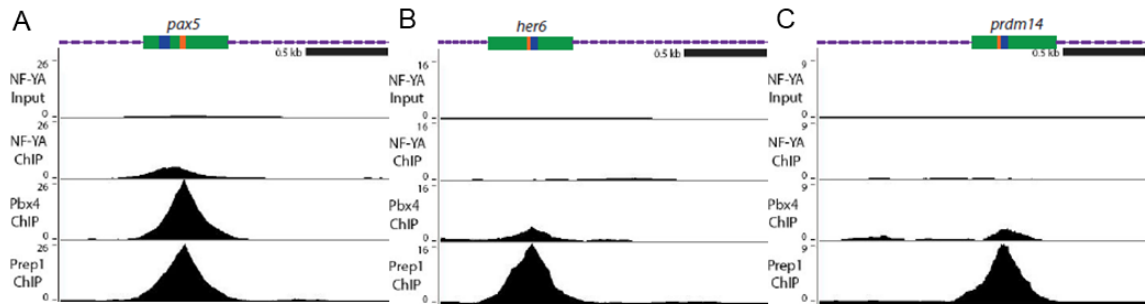
Figure 3.15: ChIP peaks at inactive putative enhancers. UCSC Genome Browser tracks showing Nfya, Pbx4, and Prep1 ChIP-seq data for the *pax5* (A), *her6* (B) and *prdm14* (C) loci. The diagrams above the tracks show the putative enhancer region in green, DECA sites in orange, and CCAAT boxes in blue.
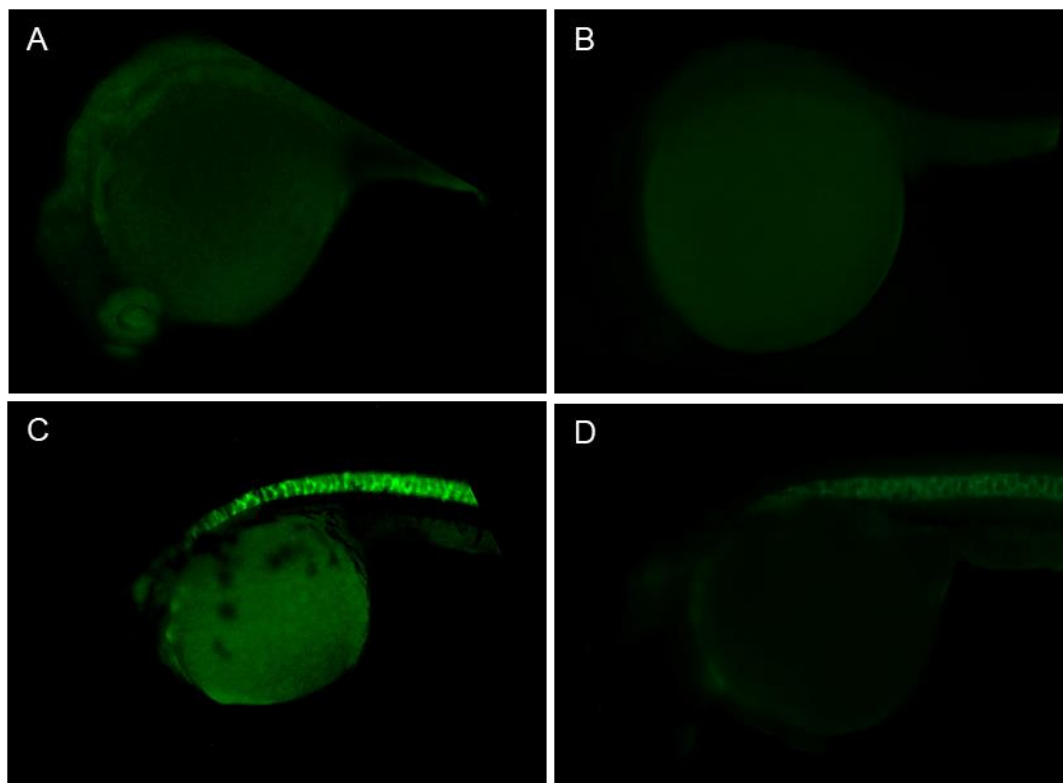
Figure 3.16: Characterization of TALE/NFY-regulated enhancers in zebrafish. (A-B) Side-by-side comparison of a 24 hpf *tcf3a*::E1b-GFP GFP-positive (A) and GFP-negative (B) embryo. (C-D) Side-by-side comparison of a 24 hpf *yap1*::E1b-GFP embryo from male founder #11 (C) and male founder #5 (D).
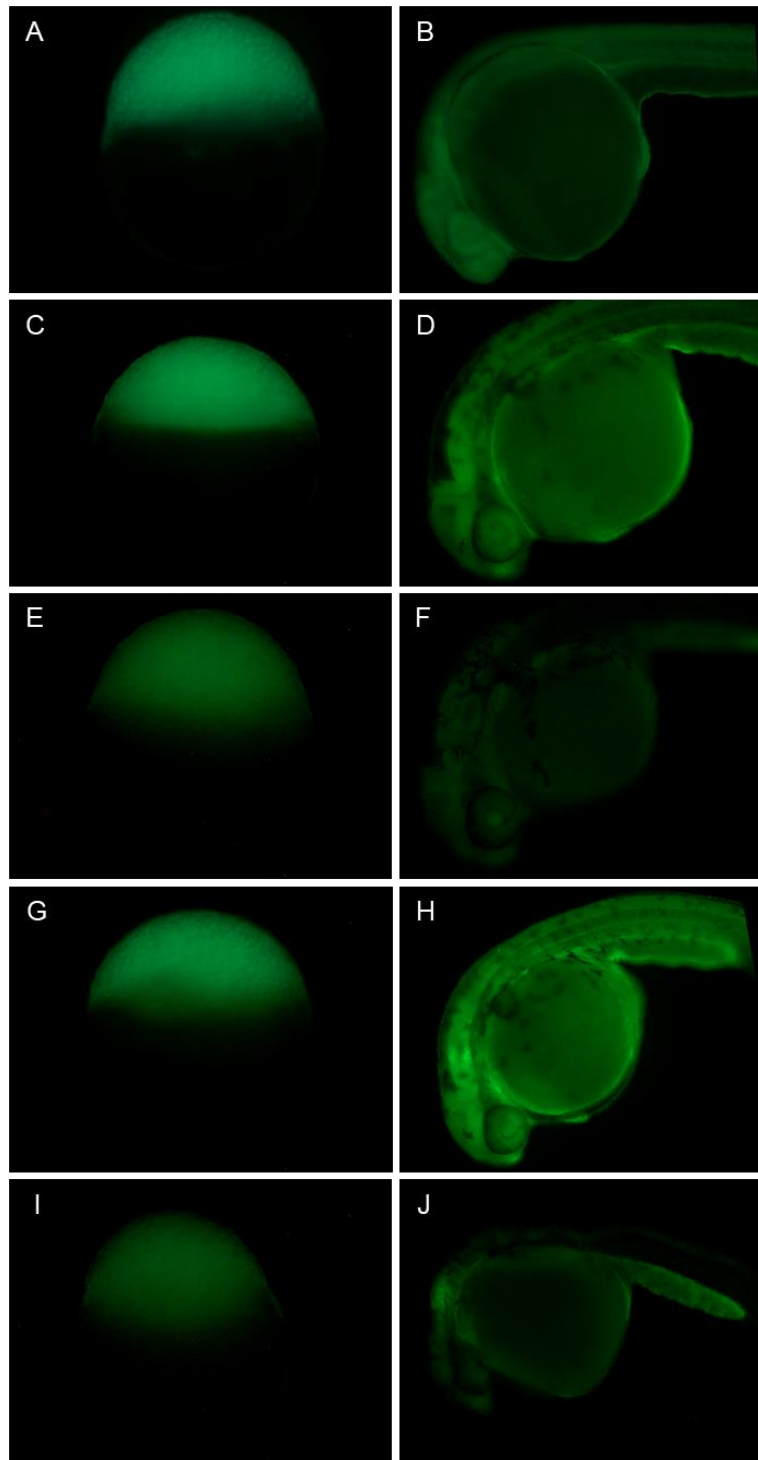
Figure 3.17: The GFP transgene is maternally deposited. GFP-positive offspring from female transgenic fish for *tcf3a*::E1b-GFP (A, B), *tle3a*::E1b-GFP (C, D), *dachb*::E1b-GFP (E, F), *fgf8a*::E1b-GFP (G, H), and *yap1*::E1b-GFP (I, J) at 3.5 hpf (A, C, E, G, I) and 24 hpf (B, D, F, H, J).
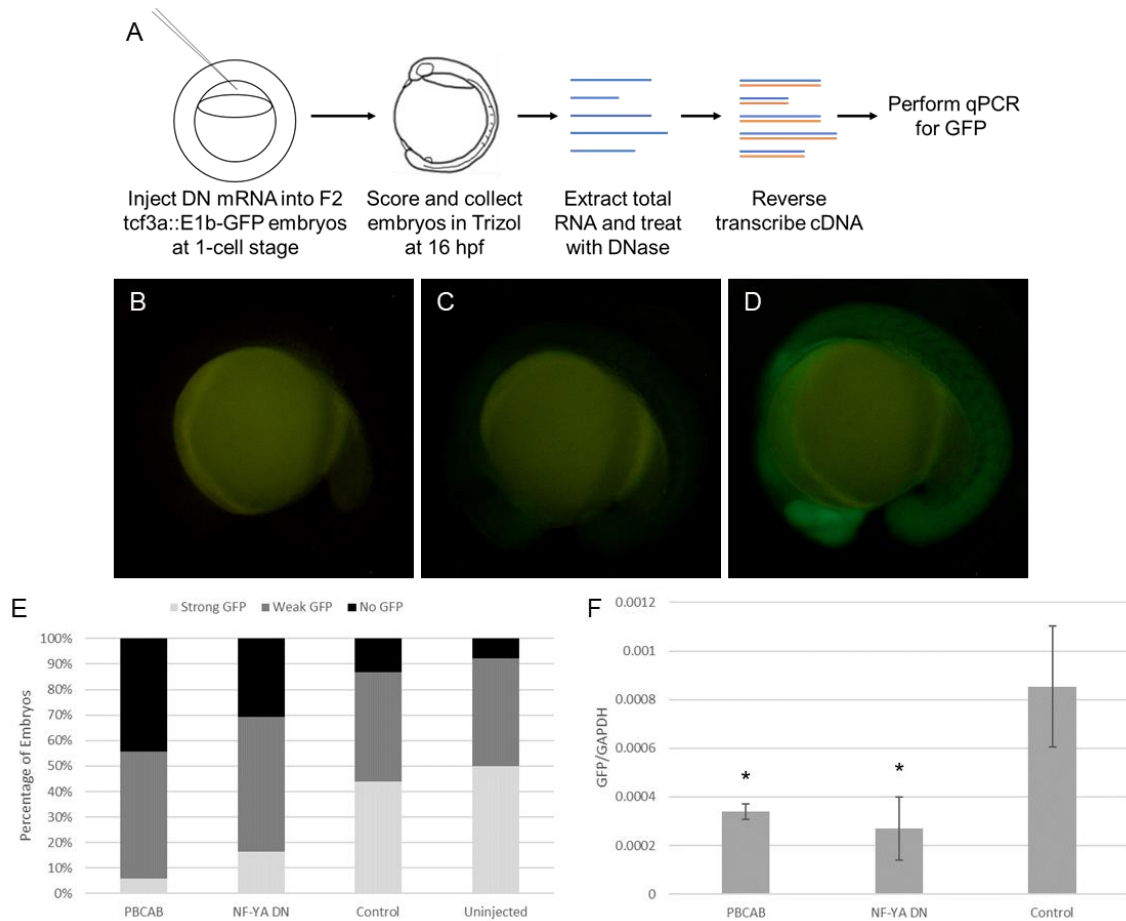
Figure 3.18: Disruption of TALE and NFY function reduces enhancer activity. (A) Schematic showing workflow for dominant negative disruption of *tcf3a*::E1b-GFP. (B-D) Representative images showing no GFP (B), weak GFP (C), and strong GFP (D) of dominant negative-injected embryos. (E) Distribution of GFP expression in uninjected embryos and embryos injected with PBCAB, Nfya- DN or control RNA. (F) RT-qPCR-based detection of GFP expression in embryos injected with PBCAB, Nfya-DN, or control RNA. Data are shown as mean +/- SEM. Statistical test: unpaired t-test.
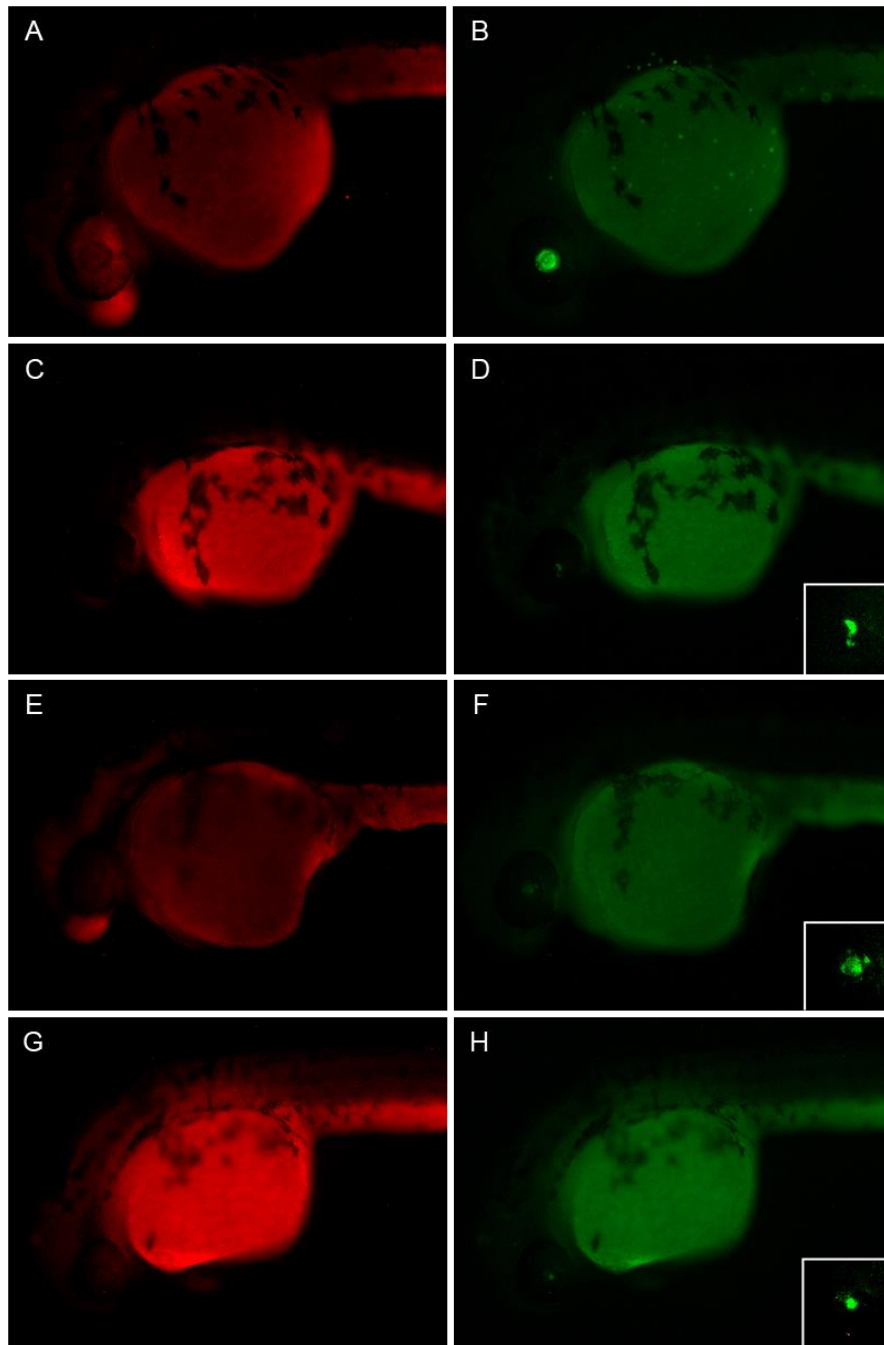
Figure 3.19: Mutation of DECA and CCAAT sites reduces enhancer activity. Representative examples of RFP (A, C, E, G) and GFP (B, D, F, H) signal in *tcf3a*-WT::sv40 (A, B), *tcf3a*-mut::sv40 (C, D), *tle3a*-WT::sv40 (E, F) and *tle3a*-mut::sv40 (G, H) embryos at 32 hpf. Insets in panels D, F, and H show higher magnification of GFP expression in the lens. Note that the embryo in panels A and B is at a later stage than embryos in panels C-H.

Line	Founders	GFP+ (32 hpf)	RFP+ (32 hpf)
tcf3a-WT::sv40	♂ #2	23/382 (6.0%)	23/382 (6.0%)
	♂ #3	6/534 (1.1%)	6/534 (1.1%)
	♂ #6	7/355 (2.0%)	7/355 (2.0%)
tle3a-WT::sv40	♂ #2	50/810 (6.2%)	50/810 (6.2%)
	♂ #5	4/262 (1.5%	4/262 (1.5%
tcf3a-mut::sv40	♂ #5	3/115 (2.6%)	0/115 (0.0%)
	♂ #6	2/260 (0.8%)	0/260 (0.0%)
	♂ #10	5/455 (1.1%)	0/455 (0.0%)
tle3a-mut::sv40	♀ #2	14/215 (6.5%)	0/215 (0.0%)

Table 3.10: Quantification of RFP-positive and GFP-positive offspring. Total number of RFP-positive and GFP-positive offspring from transgenic founders.

CHAPTER IV: DISCUSSION

This research reveals a novel mechanism for the regulation of a subset of genes at zygotic genome activation (ZGA) in zebrafish. In addition, it provides insight into many other processes related to TALE and NFY function. This data shows that the TALE factors and NFY drive the expression of distinct sets of genes at different developmental stages, demonstrating that these factors are critical to ensuring not only the activation of genes involved in key processes but also the proper timing of the expression of these genes which is essential to proper embryonic development. I also show that NFY plays a role in activating cilia-related genes, a process in which its role was previously unknown. Finally, I show that the TALE factors and NFY bind a novel DNA sequence motif prior to ZGA which is essential to the transcription-activating activity of the DNA elements in which it is present, suggesting that they may cooperate to drive many of the earliest-expressed genes in zebrafish.

NFY Drives the Expression of Cilia-Related Genes

An unexpected result obtained in this work was a role for NFY in cilia-related processes. To date, no other research has implicated NFY in the development of cilia. However, several aspects of this work support such a role for NFY. Analysis of the GO terms of the genes downregulated upon disruption of NFY at 12 hpf shows significant enrichment for cilia-related processes (Figure 3.7 C). As NFY predominantly acts as a transcriptional activator, this suggests

that NFY is responsible for directly activating genes required for proper cilia development. The ChIP-seq data provides further evidence of a direct role for NFY in cilia development, with several cilia-related GO terms appearing among genes which are both near Nfya ChIP peaks and downregulated upon disruption of NFY function (Figure 3.12 C). As cilia formation in zebrafish takes place beginning at late gastrula and segmentation stages, it is likely that a greater number of Nfya peaks would appear near cilia-related genes in embryos collected at those stages. Alternatively, it is possible that NFY drives a smaller number of master transcription factors which in turn activate larger gene regulatory networks that include many genes involved in cilia development. This finding warrants further research, which could include ChIP-seq of Nfya at 12 hpf and more detailed examination of the specific genes downregulated upon disruption of NFY function and near Nfya peaks.

In addition to the genes downregulated upon disruption of NFY and near Nfya peaks, the phenotypes of embryos injected with Nfya-DN also support a role for NFY in cilia development. Cilia fit into two predominant categories: motile cilia, which beat rhythmically to move fluids throughout the body, and sensory cilia, which send and receive signals between cells (reviewed in (Drummond, 2012; Sreekumar & Norris, 2019)). There are also modified cilia, such as the outer segments of photoreceptor cells, which are different from true cilia but nonetheless appear to share a common origin and rely on many of the same genes as true cilia for their proper development (Wolfrum & Schmitt, 2000).

Interestingly, one of the phenotypes observed upon disruption of NFY function is smaller eyes (Figure 3.2 D; Figure 3.3 G, H). It is possible that disruption of NFY results in the downregulation of cilia genes which photoreceptor cells rely upon for proper development, and abnormal photoreceptor development causes the smaller eyes observed in these injected fish.

Regarding true cilia, both motile and sensory cilia play roles in development. Motile cilia in the Kupfer's vesicle of fish embryos beat to form a leftward flow of fluid which is critical for proper development along the left-right body axis. This work shows that disruption of NFY function causes left-right asymmetry, which could stem from a defect in motile cilia development. Disruption of NFY could downregulate genes required for proper cilia development, resulting in defective motile cilia. This would prevent the leftward flow of fluid in the Kupfer's vesicle and lead to abnormal development along the left-right body axis. Sensory cilia also play a critical role in embryonic development, including in *sonic hedgehog* (*shh*) signaling. In mice, knockout of *shh* signaling pathway components results in a variety of developmental defects, including skeletal abnormalities (reviewed in (Sreekumar & Norris, 2019)). This work shows that disruption of NFY function causes abnormal anterior cartilage development (Figure 3.3 G, H). This phenotype could manifest in neural crest cells, which migrate to compose the anterior skeleton in fish. Shh plays a role in guidance of neural crest cell migration, so it is possible that defective development of sensory cilia impedes the ability of neural crest cells to detect

Shh and migrate to the proper location in skeletal formation. Alternatively,

defective development of the anterior skeleton could also be due to a joint role of

the TALE factors and NFY. Previous research indicates that disruption of the

TALE factors or NFY results in apoptosis and an inability to differentiate among

neural crest cells, suggesting that the two factors may play additional roles in the

lineage between neural crest cells and cartilage development (Chen et al., 2009;

Ferretti et al., 2006). Of course, a combinatorial effect is also possible in which

survival, differentiation, and migration of neural crest cells are all disrupted in the

absence of the TALE factors or NFY.

The TALE Factors and NFY Regulate Different Processes at Different

Developmental Stages

Both the TALE factors and NFY appear to have distinct roles at different

stages of development. This makes a great deal of sense since accurate timing

of gene expression is crucial to proper embryonic development. At 12 hpf, the

RNA-seq data shows that disruption of TALE function results in downregulation

of genes involved with transcription and embryonic development as well as

homeodomain-containing genes, the latter of which includes many tissue-specific

transcription factors (Figure 3.7 A). Comparing this data to the ChIP-seq data for

overlapping Pbx4 and Prep1 loci at 3.5 hpf shows increased enrichment of genes

involved with transcription and embryonic development while homeodomain-

containing genes become less enriched, suggesting that at 3.5 hpf the TALE factors are predominantly driving the expression of genes involved with transcription and development (Figure 3.12 B). In terms of developmental timing, this makes a great deal of sense, as the developing embryo's earliest needs include transcription factors and developmental proteins to spur it into the next stages of development (Aanes et al., 2011; Lee et al., 2013). Conversely, many of the homeodomain-containing factors are not only not essential at ZGA but could be detrimental to proper embryonic development. Were they to be expressed at ZGA, tissue-specific homeodomain-containing factors could cause premature differentiation of cells, resulting in incorrect tissue types in abnormal regions. This could then cause surrounding tissues to take on incorrect identities and prevent correct tissue types from appearing in the right location.

A similar pattern emerges upon disruption of NFY. At 12 hpf, the RNA-seq data shows that disruption of NFY causes downregulation of genes involved with transcription, embryonic development, and cilia-related processes (Figure 3.7 C). Comparison of the RNA-seq data to the 3.5 hpf Nfya ChIP-seq data shows increased enrichment of genes involved in transcription and development while enrichment of cilia-related genes declines (Figure 3.12 C). As before, genes involved with transcription and development are likely to be the most critical at ZGA while cilia-related functions are less critical than at later stages of development. Signaling molecules such as *shh* are not yet active at ZGA and organs such as Kupfer's vesicle have yet to develop, making cilia extraneous.

Consistent with this explanation, cilia development in zebrafish does not begin until the late gastrulation and segmentation stages.

The TALE Factors and NFY Cooperate to Drive Transcription at ZGA

The TALE factors and NFY appear to cooperate at ZGA to drive the expression of genes involved with transcription and embryonic development. The factors form a protein complex and bind a DNA sequence motif containing a DECA motif spaced about 20 bp from a CCAAT box (Figure 3.1 A-D; Figure 3.11 L). Examination of our previously published Prep1 ChIP-seq data shows that Prep1 preferentially binds the DECA motif prior to ZGA and that about 30% of Prep1-bound DECA motifs are within the 20 bp spacing, suggesting that they are binding the DECA/CCAAT motif I identified in this work (Ladam et al., 2018). Since almost 95% of Pbx4 sites at 3.5 hpf overlap with a Prep1 site, it is safe to assume that both Prep1 and Pbx4 are bound to these DECA sites (Figure 3.11 D). The DECA/CCAAT motif appears to be a unique motif which allows the TALE factors and NFY to have distinct functions at different stages of development. As Prep1 binds the DECA motif preferentially prior to ZGA, this suggests that the DECA motif has a unique role at ZGA which is distinct from the roles of other TALE and NFY binding sites. The data collected in this work supports this notion. Analysis of the GO terms associated with genes which are both downregulated upon disruption of TALE function at 12 hpf and near TALE peaks at 3.5 hpf

reveals enrichment for transcription, embryonic development, and homeodomain-containing genes (Figure 3.12 B). These same GO terms, as well as cilia-related GO terms, appear among the genes which are both downregulated upon disruption of NFY at 12 hpf and near Nfya peaks at 3.5 hpf (Figure 3.12 C). However, examination of the genes which are downregulated upon disruption of both the TALE factors and NFY at 12 hpf and are near overlapping TALE and Nfya binding sites at 3.5 hpf shows a complete loss of all GO terms related to homeodomain-containing and cilia-related genes, suggesting that prior to ZGA the TALE factors and NFY bound to the DECA/CCAAT motif are only driving genes related to these functions (Figure 3.12 E). Consequently, it is probable that the role of the TALE factors in driving homeodomain-containing genes and the role of NFY in driving cilia-related genes happen later and independently.

In addition to the distinct set of genes that the DECA/CCAAT motif appears to regulate, it is also essential to the ability of the enhancers it is a part of to activate transcription. Five distinct DNA elements containing the DECA/CCAAT motif were capable of driving expression of GFP *in vivo* with two different minimal promoters (Figure 3.12 B, D, F, H, J; Figure 3.19 A, E). Furthermore, each enhancer displayed a unique expression pattern, which was consistent across independent lines of the same enhancer. These expression patterns were also consistent with different minimal promoters. Taken together, this confirms that the activity observed from each enhancer was specific to the enhancer itself and not to some other facet such as integration site.

A Model for Transcription Activation by the TALE Factors and NFY at ZGA

Interestingly, each enhancer element displays a unique, tissue-restricted expression pattern (Figure 3.14 B, D, F, H, J). As the TALE factors and NFY are maternally deposited and ubiquitously expressed in embryonic tissues, this suggests that the factors themselves are not solely responsible for the activity of the enhancer elements. In the event that the TALE factors and NFY were solely responsible for driving the reporter gene expression, the reporter gene should also be ubiquitously expressed. This observation suggests that the role of the TALE factors and NFY at the DECA/CCAAT motif is broader. This fits a model in which the TALE factors and NFY possess pioneer transcription factor (PTF) activity. PTFs can recognize and attach to their binding sites in non-permissive chromatin environments, often subsequently recruiting enzymes that alter chromatin structure and form a more permissive chromatin state. This open chromatin state exposes binding sites for other transcription factors, allowing them to bind and carry out their functions. Other research suggests that both the TALE factors and NFY possess PTF activity, although whether they do is far from a consensus.

In the model I propose, the DECA/CCAAT motif possesses PTF activity, allowing the TALE factors and NFY to recognize and bind their DNA sequence motifs in the non-permissive chromatin environment of the zygotic genome prior to ZGA. This agrees with data from our previous publication, in which Prep1

preferentially binds the DECA motif prior to ZGA and then the HEXA motif later in development. It could be that Prep1 on its own lacks PTF activity and therefore cannot bind the HEXA motif in non-permissive chromatin. A complex of Prep1, Pbx4, and NFY, however, gains PTF activity, allowing all three factors to bind the DECA/CCAAT motif. Once bound, the TALE factors and NFY recruit histone modifying enzymes which open the chromatin around them and expose binding sites for other transcription factors, including, perhaps, the HEXA motif. Each of the DECA/CCAAT enhancer elements tested contains a unique sequence and thus binds a unique set of transcription factors. This is the likely explanation for the unique tissue-restricted patterns of each enhancer element, where reporter gene expression is present only in those tissues where the necessary transcription factors are expressed. It also explains the observation that the TALE factors and NFY are essential to the activity of these elements. Upon disruption of TALE or NFY activity, or mutation of the DECA and CCAAT box sequences, the enhancers lose their ability to drive reporter gene expression. In the model, this is because the TALE factors and NFY are unable to form a complex and recognize the DECA/CCAAT motif, which prevents the opening of the chromatin around the DECA/CCAAT site and exposure of binding sites for other transcription factors (Figure 4.1). This model represents a novel function for the TALE factors and NFY and provides insight into a new mechanism which spurs zygotic genome activation in zebrafish.
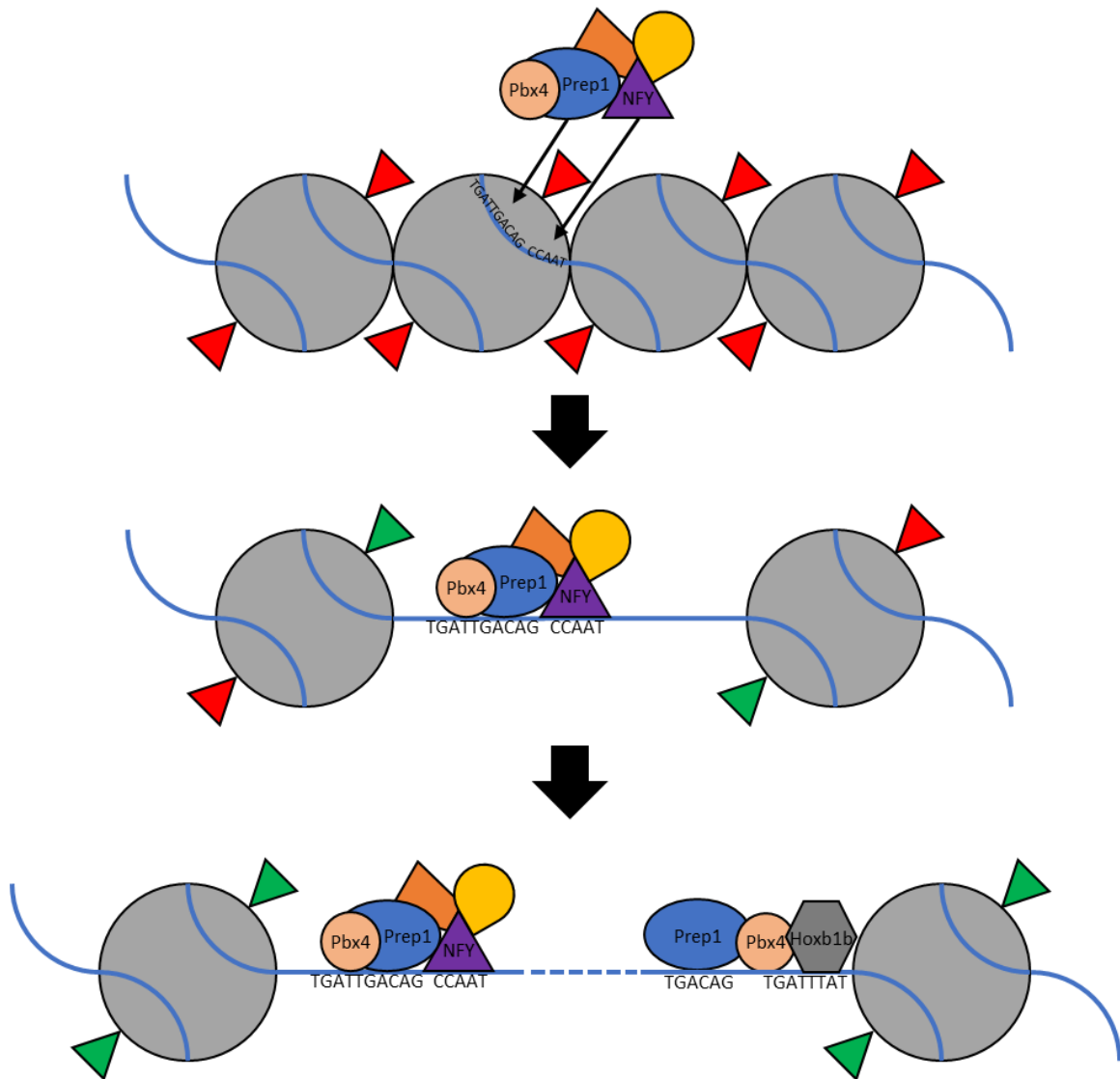
Figure 4.1: The TALE factors and NFY cooperate at ZGA to open chromatin at DECA/CCAAT enhancers. At ZGA, the TALE factors and NFY identify enhancers containing adjacent DECA/CCAAT motifs in dense chromatin occupied by nucleosomes (gray spheres) and modified by repressive chromatin marks (red triangles). The TALE factors and NFY bind there, forming a protein complex that causes the chromatin to take on a more open state marked by permissive marks such as H3K27ac (green triangles) and drives transcription of nearby early-expressed genes involved with transcription and development. As the chromatin decondenses, binding sites for other transcription factors become exposed, such as TALE/Hox sites, further activating transcription throughout the genome.

APPENDIX A: SUPPLEMENTARY DISCUSSION

The goal of this project was to expand upon previous research which suggested a cooperative role for the TALE factors and NFY. In this previous research, ChIP-seq experiments showed where Prep1 bound in the zebrafish genome at 3.5 hours post-fertilization (hpf). DNA sequence motif analysis of these binding sites using the MEME and DREME tools showed that Prep1 bound a motif with the sequence TGATTGACAG, known as the DECA motif (Ladam et al., 2018). The DECA motif comprises adjacent Pbx and Prep binding half-sites, with Pbx recognizing the TGAT portion and Prep recognizing the TGACAG portion (Chang et al., 1997; Knoepfler & Kamps, 1997). Previous *in vitro* studies of TALE binding had identified the DECA motif as a sequence recognized by the TALE factors, but it had no assigned biological function.

About 30% of the DECA sites bound by Prep1 appeared within about 20 bp of a well-characterized DNA sequence motif known as the CCAAT box (Ladam et al., 2018). Originally considered a core promoter element, technological advances such as high throughput sequencing later showed that the CCAAT box was only present at about 30% of eukaryotic promoters. Furthermore, it was also present at many enhancers (Dolfini et al., 2009; Oldfield et al., 2014). The fact that the CCAAT box appeared such a consistent distance away from such a sizeable fraction of DECA sites raised the possibility that the arrangement had a biological significance.

The majority of Prep1 association with DECA sites takes place around 3.5 hpf in zebrafish, coinciding with zygotic genome activation (ZGA). This bound Prep1 persists to 12 hpf, but Prep1 binds few additional DECA sites between 3.5 hpf and 12 hpf. Although many new Prep1 binding sites appear between 3.5 hpf and 12 hpf, the vast majority of these sites fit the more classically understood arrangement of Prep1 binding (Figure 1.3) (Ladam et al., 2018). In this arrangement, Prep1 binds the HEXA motif containing the sequence TGACAG (Amin et al., 2015; Ferretti et al., 2000; Ferretti et al., 2005; Jacobs et al., 1999; Tümpel et al., 2007). This sequence is frequently located a short distance from a composite binding site for Pbx and a tissue-specific transcription factor such as Hox (Grieder et al., 1997; Pöpperl et al., 1995; Ryoo & Mann, 1999). Furthermore, the fact that Prep1 and Pbx4 are bound to DECA sites at 3.5 hpf indicates that the proteins originate from a maternal source. Indeed, previous research has shown that the TALE factors bind as early as 2 hpf, well before ZGA (Choe et al., 2014). Taken together, these observations suggest that TALE factors bound to the DECA motif have a separate biological function from those bound to the more classically understood arrangement.

Like the TALE factors, the literature shows that NFY is maternally deposited across species (Dolfini et al., 2009). Besides its well understood role as a promoter-binding factor regulating housekeeping genes related to processes such as the cell cycle, recent research implicates NFY in a role in embryonic development. Specifically, this research describes it as an enhancer-binding

pioneer transcription factor (PTF) essential to ZGA in mice (Dolfini et al., 2012; Lu et al., 2016). The association of CCAAT boxes with DECA sites and the fact that the TALE factors and NFY interact on a protein level suggests a potential cooperative role for the two factors in ZGA at enhancers (Ladam et al., 2018).

The TALE Proteins Pbx4 and Prep1 Interact with NFY

The co-immunoprecipitation experiment demonstrated that Nfya and Nfyb both immunoprecipitated with Pbx4 and Prep1 (Figure 3.1). The selection of these proteins for this experiment, as well as subsequent ones, stems from the fact that all of them are maternally deposited in zebrafish (Choe et al., 2002; Deflorian et al., 2004; Dolfini et al., 2009; Ferretti et al., 2000; Vlachakis et al., 2000). Our previous Prep1 ChIP-seq data at 3.5 hpf and 12 hpf showed that Prep1 preferentially bound DECA motifs prior to ZGA (Ladam et al., 2018). As the DECA site comprises immediately adjacent Pbx4 and Prep1 half-sites and Pbx4 forms protein complexes with Prep1, Pbx4 seemed the obvious choice to complete the set of potential TALE factor interactors with the NFY subunits.

The finding that the TALE factors and NFY form a protein complex is significant since it implies that the two factors work together in some capacity. A great deal of the TALE factors' function derives from their ability to form complexes with unique compositions of proteins that will have specific effects at different genes (reviewed in (Ladam & Sagerström, 2014)). Thus, it is possible

that forming a complex with NFY allows the formation of yet another class of TALE complexes which fulfill this role at a key subset of genes. Alternatively, the formation of a TALE/NFY complex may impart some new function which the proteins lack independently. The TALE factors bind DNA very early in development and previous research implicates NFY as an essential factor for proper ZGA in mice (Choe et al., 2014; Lu et al., 2016). Due to this early genomic binding, many reports have suggested that the TALE factors and NFY possess PTF activity, but this is far from a consensus. Given the early association of the TALE factors with the DECA motif and the association of the CCAAT box with the DECA motif, it is possible that by associating with one another the TALE factors and NFY gain PTF ability which each factor lacks independently.

Although this experiment showed that the TALE factors and NFY do indeed interact on a protein level, the determination of the exact basis of the interaction will require more work. The fact that the proteins are well conserved leaves open the possibility that the fish proteins can form complexes with the human proteins endogenous to the HEK293T cells. It is possible, for example, that Nfyb and Pbx4 physically interact, but tagged Prep1 co-immunoprecipitates with tagged Nfyb on account of endogenous human PBX1 mediating the interaction between the two tagged proteins. The choice of system is another potential caveat to these results; there is no guarantee that the fish proteins will fold correctly in human cells due to the presence of different chaperones and environments between the two species. This is of particular concern on account

of the activation domains of Nfya and Nfyc, which are rich in glutamine and hydrophobic residues and can form pathological aggregations (reviewed in (Dolfini et al., 2012)). Finally, the epitope tagged versions of the proteins used for this experiment introduce the possibility of non-specific interactions or aggregation caused by the tags themselves. With the creation of the *nfya* antiserum used for ChIP-seq, it would be possible to repeat the co-immunoprecipitation in fish embryos to confirm the interaction. In addition, it would be useful to repeat the experiment with a tagged version of Nfyc to determine whether that subunit can also interact with Prep1 and Pbx4.

Disruption of TALE or NFY function causes developmental abnormalities

With the knowledge that the TALE factors and NFY form a protein complex at DECA/CCAAT sites during ZGA, I disrupted each factor's function to determine whether they caused similar defects in early embryonic development. To this end, I injected dominant negative versions of Pbx4 (PBCAB) or NFY (Nfya-DN) into zebrafish embryos at the 1-cell stage. PBCAB disrupts TALE function by sequestering endogenous Prep1 in the cytoplasm, and Nfya-DN disrupts NFY function by eliminating recognition of the CCAAT box, trapping endogenous Nfyb and Nfyc in a non-functional complex (Choe et al., 2002; Mantovani et al., 1994). Each of these dominant negative factors causes gross developmental defects. Compared to uninjected or GFP-injected animals,

PBCAB- and Nfya-DN-injected embryos displayed small or missing heads and small or missing eyes by 28 hpf (Figure 3.2). By 5 days post-fertilization (dpf), surviving embryos displayed gross deformities in anterior cartilage development (Figure 3.3). Although PBCAB and Nfya-DN cause many similar phenotypes, the affected animals are not identical. Most obviously, about half of the PBCAB-injected embryos lacked the third rhombomere of the hindbrain (r3) as indicated by *krox20* staining *in situ* (Figure 3.4 E, F). None of the Nfya-DN-injected embryos displayed this loss of r3, nor did they display any abnormalities in the midbrain-hindbrain boundary, r3, r5, or the spinal cord. (Figure 3.4 G, H). Based on these results, hindbrain development likely falls under the purview of a TALE function independent of NFY, while cooperation between the TALE factors and NFY affects the development of more anterior structures.

Dominant negatives versus morpholinos and germline mutants

The phenotypes caused by the dominant negatives fit with those previously described (Chen et al., 2009; Choe et al., 2002; Deflorian et al., 2004; Mantovani et al., 1994; Heike Pöpperl et al., 2000; Waskiewicz et al., 2002). In addition to the dominant negatives, prior experiments have disrupted TALE and NFY function using antisense Morpholino oligos (MO) and germline mutants. MOs bind specific mRNAs to block their translation or splicing. Recent research calls the specificity of MOs into question, however, observing that morphant

phenotypes are not always identical to the phenotypes observed in germline

mutants (Kok et al., 2015). This research posits that many observed phenotypes

are due to off-target effects, such as activation of p53 in response to MO

injection. Proponents of MOs suggest that genes with redundant function can

compensate for the loss of gene function in germline mutants, which could

explain the discrepancy between morphant and germline mutant phenotypes

(Rossi et al., 2015). This potential lack of specificity calls for careful consideration

of morphant phenotypes.

The dominant negative constructs appear to be a superior alternative to

MOs for hindering TALE and NFY function in zebrafish. Outwardly, the dominant

negative and morphant phenotypes compare favorably. Like PBCAB, TALE MOs

cause smaller heads and smaller eyes as well as a loss of r3. For a more in-

depth analysis of PBCAB versus the TALE MOs, see Appendix B, Figure B.1. In

addition, the phenotype caused by PBCAB mirrors that of the *pbx4* germline

mutant (Pöpperl et al., 2000; Waskiewicz et al., 2002), indicating that the effects

are specific. The same is true of Nfya-DN and NFY MOs, which both display

similar phenotypes in zebrafish (Chen et al., 2009; Ladam et al., 2018).

The TALE factors and NFY are essential for proper head cartilage formation

Cartilage deformities are one of the chief similarities shared between the

TALE and NFY phenotypes. Previous research using NFY MOs in zebrafish

described sharper Meckel's cartilage, loss of ceratobranchial cartilage, and enlarged angles of ceratohyal cartilage. This report attributed the cartilage abnormalities to apoptosis of neural crest cells (Chen et al., 2009). Interestingly, *prep1.1* MOs cause a similar dysfunction, with *prep1.1* morphants losing neural crest-derived cartilages in the anterior skeleton. Among the possible reasons for the loss of this cartilage was an inability of these cells to properly differentiate, causing them to undergo apoptosis (Deflorian et al., 2004). In addition, *lazarus* (*pbx4*) mutants also demonstrate anterior cartilage deformities (Pöpperl et al., 2000). Just as the *prep1.1* and *pbx4* abnormalities may be related to the intimate cooperation between those proteins, it is possible that the deformities observed in NFY MOs are related as well. It is also possible that, among other early development roles, the TALE factors and NFY cooperate in neural crest cell differentiation and migration. This finding is significant as it suggests a direct cooperative role for the TALE factors and NFY, although confirming it will require further research.

The abnormalities in anterior cartilage, as well as the apoptosis and differentiation deficiencies of neural crest cells, of TALE and NFY disruption make sense in the context of each gene's expression pattern. In zebrafish, *pbx4*, *prep1.1*, and *nfyb* expression transitions from a ubiquitous expression pattern early in development to a tissue-restricted pattern, becoming limited to the anterior part of the embryo (Deflorian et al., 2004; Dolfini et al., 2012; Pöpperl et al., 2000). Furthermore, in the case of NFY, its role in regulating cell cycle genes

is well documented (Dolfini et al., 2012). Thus, it makes sense that disruption of NFY with MOs could cause apoptosis of cells in a region to which an essential subunit is restricted. Although I did not look for apoptosis with Nfya-DN, the similarity of the phenotypes suggests similar mechanisms for it and MOs.

Severity of TALE and NFY phenotype

While disruption of the TALE factors and NFY results in substantial deformity, one might expect factors involved in activation of the zygotic genome to lead to more severe abnormalities. For example, the pluripotency factors Oct4/ Pou5f3, SoxB1, and Nanog also play roles in ZGA in zebrafish (Lee et al., 2013; Leichsenring et al., 2013). Disruption of these factors causes embryos to stall during blastula stages. I do not observe such a severe phenotype in response to the dominant negatives, but there are several potential explanations as to why this is the case. As I will explain further in the next section, the function of the dominant negatives has a lag time while the injected mRNA is translated. This, as well as maternally deposited proteins, allows the formation of functional TALE complexes, meaning that the dominant negatives do not cause a complete loss of TALE function. Furthermore, the developmental arrest observed in the pluripotency factors requires disruption of all three proteins, and recent evidence suggests that the effect of one, Nanog, may be due to its role in extraembryonic tissue formation (Gagnon et al., 2018). This same principle applied to DUX,

which is implicated in mammalian ZGA. Like the individual pluripotency factors, disruption of DUX does not lead to developmental stalling (Chen & Zhang, 2019). Thus, the lack of a more severe phenotype in the disruption of the TALE factors and NFY may be because vertebrate ZGA requires the activity of several transcription factors and no single one can block the process.

Potential caveats of the dominant negatives

Although the dominant negatives appear to be superior to MOs in terms of mitigating off-target effects while still showing phenotypic changes associated with the disruption of their target factors, they are not ideal substitutes for knockout mutants. In the case of Pbx and NFY, both factors are maternally deposited. Although mRNA encoding the dominant negatives is injected at the 1-cell stage within about 45 minutes after fertilization, time is still required for these factors to undergo translation and begin to disrupt TALE and NFY function. During this initial lag time, endogenous TALE and NFY factors are forming normal complexes and carrying out their normal functions. In this way, the dominant negatives may more accurately portray hypomorphic mutations than knockouts and may be inconsistent between individual embryos depending on which genomic loci normal endogenous complexes are able to bind. Nonetheless, they are a remarkably useful tool in studying the disruption of the TALE factors and NFY.

Whatever the selected method of gene disruption, it is important to bear in mind that I extracted total RNA for the RNA-seq from whole embryos, which at 12 hpf have differentiating tissues with distinct gene expression profiles. Therefore, while the direct effects of the TALE factors or NFY disruption may be very large in specific tissues, they are diluted out by the inclusion of other tissues where the effect is not as drastic or not present at all. For example, the dominant negatives seem to have a drastic effect on anterior development, suggesting that much of their activity is in that region of the embryo. However, genes whose expression is restricted to the anterior region, or whose expression is only affected in the anterior region, may have this effect diluted by the inclusion of the posterior region. This could result in the expression fold change not meeting the 1.5-fold cutoff. This effect could be corrected by dissecting the embryos before preparing the RNA libraries and only including anterior regions for RNA-seq analysis.

The TALE factors and NFY have cooperative and independent functions

The results of the previous experiments added support to the notion that the TALE factors and NFY cooperate to regulate genes involved in early embryogenesis. To ascertain which genes these factors may co-regulate, I performed RNA-seq on zebrafish embryos injected with PBCAB or Nfya-DN mRNA normalized to embryos injected with GFP mRNA (Figure 3.5 A). Since the TALE factors and NFY are mostly described as transcriptional activators, I

focused on the downregulated genes when analyzing the RNA-seq data, as those effects are more likely to be direct attributable to the loss of TALE or NFY function. For transcriptional activators, upregulation of genes is more likely caused by indirect effects; for instance, a transcriptional activator could drive transcription of a gene that represses the expression of another, and thus disrupting the activator results in indirect upregulation of this final gene. However, there are reports that both the TALE factors and NFY possess repressor activity. In the case of the TALE factors, Pbx4 can recruit HDACs and N-CoR/SMRT to remove acetyl groups from histones and create a less permissive chromatin environment for transcription (Saleh et al., 2000). In *C. elegans*, NFY can also create restrictive chromatin environments, interacting with the Polycomb repressive complex which deposits the repressive H3K27me3 mark. Furthermore, in *C. elegans*, the loss of NFY leads to ectopic expression of the *egl-5* gene, suggesting that NFY is responsible for repressing *egl-5* in areas it is not supposed to be expressed (Deng et al., 2007).

Since the TALE factors and NFY can possess repressor activity, it is possible that some of the genes upregulated by the dominant negatives are due to direct effects. Fitting with this possibility, many of the GO terms of the upregulated genes fit into the same categories as the downregulated genes, particularly transcription related ontologies (Figure 3.8 B-D). It is possible to dismiss some of these GO terms as indirect; for instance, upregulation of the p53 pathway is often seen with MOs and is considered an off-target effect due more

to the injection of foreign material than any specific effects. There are other GO

terms, however, which are interesting enough to bear mentioning and perhaps

further research. In the genes upregulated by both PBCAB and Nfya-DN, the GO

terms CCAAT/enhancer binding and CCAAT/enhancer binding protein appear

(Figure 3.8 D). The genes in these terms are *cebpa*, *cebpb*, and *cebpd*, which

stand for CCAAT/enhancer binding proteins A, B, and D, respectively. These

proteins are not the same as NFY; however, CEBPA interacts with NF-YA in

humans. Despite their name, the members of the CEBP family bind the DNA

consensus sequence RTTGCGYAAY and do not compete with NFY for CCAAT

box binding. Nonetheless, the category is still an interesting one as the CEBP

genes are transcriptional activators that play roles in animal organ development

and hematopoietic stem cell migration. Finally, Nfya-DN upregulates genes

involved in histone-related GO terms. This is interesting since Nfyb and Nfyc

contact DNA via histone-fold domains, and many researchers have posited that

this facilitates the pioneer activity of NFY. If due to a direct effect, this data

suggests that NFY represses histone expression, perhaps further facilitating the

ability of itself and other transcription factors to compete with nucleosomes for

DNA binding. Indeed, saturation with histones appears to be a strategy in

zebrafish and *Xenopus* to prevent premature ZGA (Joseph et al., 2017; Pálfy et

al., 2017); thus, it is possible that NFY represses histones as an additional

mechanism to ensure efficient ZGA.

The TALE factors drive transcription, development, and homeobox genes

The fact that genes disrupted by PBCAB are predominantly involved in transcription, embryonic development, and homeobox genes agrees with previous data (Figure 3.7 A, B) (reviewed in (Ladam & Sagerström, 2014)). Much of the research into the TALE factors relates to their roles as cofactors for tissue-specific transcription factors such Hox and Engrailed. These proteins, as well as many other TALE cofactors, are master transcription factors. For example, Pbx4 and Prep1 bind the *hoxb1a* locus in zebrafish to activate its transcription in cooperation with Hoxb1b (Figure 1.3) (Choe et al., 2014). *hoxb1a* fits all three major categories of TALE-regulated genes: it is a homeodomain-containing transcription factor that plays a role in central nervous system development, particularly that of r4. In addition, GO terms of the genes downregulated by only PBCAB (and not Nfya-DN; Figure 3.9 A) include many of the same GO terms as the whole set of genes downregulated by PBCAB, indicating that their unique functions are not very different from their coordinated functions with NFY. Taken together, this data adds evidence that the TALE factors play vital roles in transcription regulation and development, particularly of homeobox genes.

NFY drives transcription, development, and cilia genes

The GO terms of genes downregulated by Nfya-DN reveal new insight into the role of NFY in zebrafish. Cilia-related genes compose a large part of the most highly downregulated genes by Nfya-DN (Figure 3.7 C, D). Cilia are small projections contiguous with the plasma membrane of cells which can have motile and sensory functions. They are significant signaling centers, coordinating many cellular signaling pathways such as *sonic hedgehog* (*shh*). Cilia play a significant role in development, particularly left-right body and neural patterning, but also in the cell cycle and differentiation (reviewed in (Drummond, 2012; Sreekumar & Norris, 2019)). In addition, cilia play significant roles in the eye, where the outer segment of photoreceptor cells is a specialized cilium (Wolfrum & Schmitt, 2000). The regulation of cilia-related genes by NFY could explain many of its proposed functions. The role of NFY in the cell cycle is well studied thanks to the presence of CCAAT boxes in many cell cycle genes but the notion that NFY plays an additional part in this process by regulating cilia, which sense extracellular signals related to the cell cycle, provides new insight into NFY's role in this process. Interestingly, *yap1*, one of the TALE/NFY enhancers capable of driving GFP expression (Figure 3.14 J), plays a role in cilia assembly.

Many studies implicate NFY in embryonic development but its role in this process is less well understood than its role in the cell cycle. The notion that NFY activates cilia genes may be a link to its role in development. For example, many

embryos injected with Nfya-DN displayed left-right asymmetry, and cilia are implicated in left-right body patterning (reviewed in (Drummond, 2012)). Furthermore, the smaller eyes of the Nfya-DN-injected embryos could be related to the fact that outer segments of photoreceptor cells are modified cilia in that downregulation of cilia genes could alter eye morphology (Wolfrum & Schmitt, 2000). Finally, the role of NFY in activating cilia genes could disrupt *shh* signaling. Knockout mice for various components of the *shh* signaling pathway display, among other phenotypes, improper brain and skeletal development. As cilia are an important part of *shh* signaling, their loss could explain some of the more severe anterior phenotypes observed. One limitation of Nfya-DN was that many of the most severely affected embryos died too early in development for inclusion in the Alcian blue stain, which could not take place until cartilage formed at 5 dpf. Additionally, the normal patterning posterior to the midbrain-hindbrain boundary suggests that any central nervous system abnormalities must occur anterior to that region. Better characterization of the role of NFY in early anterior development will necessitate additional *in situ* experiments using early anterior markers such as *pax6a*.

In addition to cilia genes, Nfya-DN downregulates many genes involved in transcription regulation and development (Figure 3.7 C, D). All of the cilia-related GO terms disappear in the set of genes downregulated by both PBCAB and Nfya-DN (Figure 3.7 E, F), suggesting that they are independently regulated by NFY but not the TALE factors. Further supporting this conclusion is the set of GO

terms for genes downregulated by Nfya-DN alone, which include many cilia-related ontologies (Figure 3.9 B). The genes downregulated by both PBCAB and Nfya-DN are predominantly related to transcription regulation and development, although there are also some homeobox genes (Figure 3.7 E, F). Interestingly, there are also many cytoskeleton-related GO terms among the genes downregulated by both factors. The cytoskeleton plays a prominent role in cilia formation, providing the organelle's structure and a means of transporting signals between the cilium and cell (reviewed in (Drummond, 2012; Sreekumar & Norris, 2019)). Although these GO terms are too broad to attribute to cilia with any certainty, they could be related to that function.

These findings are significant as they shed new light on the roles of both the TALE factors and NFY during early embryonic development. While the TALE factors are widely understood to play roles in transcription regulation, embryonic development, and regulation of homeobox genes, the role for NFY in cilia-related processes is far less well characterized. Furthermore, these results demonstrate that, although the TALE factors and NFY only cooperate to regulate a subset of their total target genes, the genes that they do cooperate to regulate are involved in transcription regulation and development. Thus, this data supports a model in which the TALE factors and NFY are versatile transcription factors which regulate many genes on their own but also work together during ZGA, where they are critical to the activation of many of the earliest genes which the embryo needs to continue its development and drive new sets of genes.

The TALE factors and NFY bind near differentially expressed transcription and development genes

Having identified genes that are differentially expressed following disruption of the TALE factors or NFY, I next performed ChIP-seq for Pbx4 and Nfya at 3.5 hpf. Together with previously collected ChIP-seq data, I intended to use this data to establish which genes affected by the dominant negatives were also near a TALE or NFY binding site, since this would suggest that these factors directly regulated these genes. In general, about 50%-60% of the genes downregulated by a dominant negative have the corresponding factor's peak nearby (Figure 3.12 A; Table 3.8). This is significant because it provides further evidence that the gene is directly regulated by either TALE, NFY, or both. At 3.5 hpf, genes with TALE and NFY peaks nearby and downregulated by PBCAB and Nfya-DN are involved with transcription regulation and development (Figure 3.12 D, E). This makes sense since genes involved with these processes are a major part of the first wave of zygotic gene expression. Genes downregulated by PBCAB or Nfya-DN only with a ChIP peak of the corresponding factor nearby fit into the same sets of GO terms as the whole sets of genes downregulated in the RNA-seq (Figure 3.12 B, C). Genes downregulated by PBCAB and near TALE peaks (defined as overlapping Pbx4/Prep1 peaks) regulate transcription, development, and homeoboxes (Figure 3.12 B), while genes downregulated by Nfya-DN and near Nfya peaks regulate transcription, development, and cilia

(Figure 3.12 C). Interestingly, several homeobox GO terms also appeared among the genes downregulated by Nfya-DN and near Nfya peaks. This suggests that the TALE factors and NFY begin binding their target genes by 3.5 hpf. This makes sense for the TALE factors based on the aforementioned example at *hoxb1a*, where the TALE factors bind as early as 2 hpf despite transcription not beginning until around 6 hpf (Figure 1.3). It appears that NFY exhibits similar binding behavior although further study would be necessary to determine whether transcription at these genes also begins later.

There are several potential explanations as to why the remaining downregulated genes lack a nearby peak. The most conspicuous is the discrepancy between the stages of the embryos in the RNA-seq and ChIP-seq. Ideally, the samples from both the RNA-seq and ChIP-seq data would come from the same timepoints, but this is technically difficult. Performing RNA-seq at 3.5 hpf would only reveal maternally deposited transcripts, as ZGA has yet to occur, and the levels of these transcripts would be unchanged as their expression does not depend on the dominant negatives. Additionally, an earlier RNA-seq at 6 hpf found very few differentially expressed genes. Therefore, we are not able to assess changes in expression of genes with TALE of Nfya peaks at 3.5 hpf.

Previous work with Prep1 indicates that many new Prep1 binding sites appear between 3.5 hpf and 12 hpf (Ladam et al., 2018), and the same may be true for Pbx4 and Nfya. Thus, a gene downregulated by a dominant negative at

12 hpf may have a TALE or NFY peak, but that peak is simply not present at 3.5 hpf. To try to determine whether this could be the case, I used the 6 hpf RNA-seq data to examine whether downregulated genes near ChIP-seq peaks were expressed more highly at 6 hpf than downregulated genes without ChIP-seq peaks. Although this appears to be the case for neither the TALE factors nor NFY individually (data not shown), genes downregulated by PBCAB and Nfya-DN that were also near both TALE and NFY peaks do appear to be more highly expressed at 6 hpf than genes downregulated by PBCAB and Nfya-DN without those peaks nearby (Figure 3.12 F). This suggests that the discrepancy in timepoints between the RNA-seq and ChIP-seq data causes some genes that would otherwise have a corresponding ChIP-seq peak nearby at 12 hpf to appear not to have one, perhaps because those peaks appear between 3.5 hpf and 12 hpf. To shed further light on this possibility, it would be best to collect ChIP-seq data at 12 hpf for Pbx4 and Nfya and repeat the comparison to see whether a larger subset of genes downregulated by PBCAB and Nfya-DN are near TALE and Nfya ChIP-seq peaks.

Another potential reason for the apparent disparity between the RNA-seq data and ChIP-seq data is that some of the genes downregulated by the dominant negatives are not directly regulated by the TALE factors or NFY. A good indication that a gene is directly regulated by a factor is that it is both downregulated upon disruption of the factor and is near a corresponding peak, whereas a gene downregulated upon disruption of the factor but lacking a

corresponding peak may be indirectly regulated. Alternatively, as with the RNA-seq data, the ChIP-seq data was conducted in whole embryos, which may have diluted some peaks below the fold-enrichment cutoff. Although this effect would not be as pronounced as in the RNA-seq data since the embryos are far less differentiated at 3.5 hpf, there is bound to be some variability in precise stage of development among the embryos collected and this could account for the apparent absence of some peaks near a directly regulated gene.

These findings are significant in that they demonstrate a direct role for the TALE factors and NFY in driving expression of transcription- and development-related genes early in embryogenesis. This research identifies many genes which are near TALE and NFY binding sites and differentially expressed in response to disruption of the TALE factors and NFY. In support of the well-characterized functions of the TALE factors and NFY in driving gene expression, genes downregulated by PBCAB and near TALE peaks or downregulated by Nfya-DN and near Nfya peaks regulate genes involved in transcription regulation and development. Independently, the former also regulate homeobox genes and the latter cilia-related genes. The role of NFY in regulation of cilia-related genes is a new finding that warrants further study. Finally, genes downregulated and bound by both the TALE factors and NFY are involved predominantly in transcription regulation and development, consistent with the hypothesis that these factors cooperate to activate genes involved with these processes at ZGA.

The TALE factors and NFY bind near transcription and development genes early in development

In addition to their relationship with the differentially expressed genes, the ChIP-seq data reveals a great deal of other insight into the behavior of the TALE factors and NFY at 3.5 hpf. One particularly interesting observation is that nearly all (93.8%) of the Pbx4 peaks overlap a Prep1 peak while only 36.8% of Prep1 peaks overlap a Pbx4 peak (Table 3.7; Figure 3.11 D). Several possibilities could account for this observation. One potential explanation is that the Prep1 antiserum is more sensitive than the Pbx4 antiserum, resulting in more Prep1 peaks meeting the 10-fold enrichment cutoff. Alternatively, it is possible that Prep1 binds more regions of the genome at 3.5 hpf than Pbx4. Pbx4 may bind these regions later in development, or it may never bind at these regions. Collecting Pbx4 ChIP-seq data at 12 hpf could help shed light on the nature of the TALE overlap.

TALE and NFY co-bind a subset of genomic loci

Similar to the Pbx4/Prep1 binding observation, there are fewer Nfya peaks at 3.5 hpf than TALE peaks. Once again, this could be due to a less sensitive antiserum in the case of Nfya, or it could be because at 3.5 hpf Nfya binds fewer genomic loci than the TALE factors. The overlap of TALE/Nfya peaks is not huge,

consisting of just 22% of the total Nfya peaks and 16.7% of the total TALE peaks

(Table 3.7; Figure 3.11 J). It is worth noting that about 30% of DECA sites in the

zebrafish genome overlap with a CCAAT box (Ladam et al., 2018). This is not too

different from the totals observed in the ChIP-seq data, especially considering

that the observation that 30% of DECA sites are near a CCAAT box came from

Prep1 ChIP-seq data which contained many more peaks than either the Pbx4 or

Nfya ChIP-seq data. However, if the discrepancy holds true it suggests that even

at 3.5 hpf the TALE factors and NFY could have independent functions. Further

research into the non-overlapping TALE and NFY regions could reveal deeper

insights into other mechanisms for the TALE factors and NFY during ZGA.

     Although the TALE factors and NFY could have independent functions at

3.5 hpf based on the exclusive genomic regions they bind, the functions of these

genes appear to fall within the same ontologies of transcription and development.

For Pbx4 alone and overlapping TALE peaks (Figure 3.11 B, E), this is not

terribly surprising on account of the large percentage of Pbx4 peaks that overlap

with Prep1 peaks. It is worth noting that the homeobox GO terms disappear,

suggesting that either the TALE factors bind these regions later in development

or regulation of these genes happens from more distal enhancers. A similar

pattern holds true for Nfya, which binds near genes involved with transcription

and development but lacks any cilia- or homeobox-related functions, perhaps for

the same reasons (Figure 3.11 H). Based on these GO terms, it is not surprising

that the genes near overlapping TALE and NFY peaks are involved in

transcription and development (Figure 3.11 K). This, along with the RNA-seq data (Figure 3.12 D, E), strengthens the notion that the TALE factors and NFY cooperate to drive genes in these functions at the ZGA.

TALE and NFY bind DECA/CCAAT sites at 3.5 hpf

Not surprisingly, sequence motif analysis with the MEME tool revealed that the top sequence motif at 3.5 hpf for both Pbx4 alone, overlapping Pbx4/Prep1 sites, and sites bound by TALE but not NFY was the DECA motif (Figure 3.11 C, F, M). This agrees with Prep1 only data at 3.5 hpf, where the DECA site was also one of the top sequence motifs (Ladam et al., 2018). In addition, the CCAAT box was the top sequence motif at 3.5 hpf for sites bound by Nfya and sites bound by Nfya but not the TALE factors (Figure 3.11 I, N). Strikingly, MEME detected a well-defined sequence motif at overlapping TALE/Nfya sites consisting of a DECA site and CCAAT box separated by nine base pairs (Figure 3.11 L). The nine base pairs between the DECA site and CCAAT box do not appear to have any sort of consensus sequence identity, but the spacing agrees with the 3.5 hpf Prep1 data in which we originally identified the association of the CCAAT box with the DECA motif.

In our previous publication, Prep1 ChIP-seq data at 3.5 hpf and 12 hpf shows that Prep1 binds many new sites between 3.5 hpf and 12 hpf. Furthermore, many of the Prep1 sites bound at 3.5 hpf are at the DECA motif,

and Prep1 binds very few new DECA sites between 3.5 hpf and 12 hpf (Ladam et al., 2018). Instead, many of these sites are in the more classical composition of a HEXA site near a composite Pbx/tissue-specific transcription factor site, such as Hox (Figure 1.3). While we cannot yet be sure that Pbx will mirror such a pattern, the Prep1 ChIP-seq data suggests it may. The *hoxb1a* genomic locus, for example, is a more classical arrangement of Pbx4 and Prep1 binding sites. Here, Pbx4 and Prep1 bind very early in development, detected as early as 2 hpf (Figure 1.3). Transcription at this locus, however, does not begin until Hoxb1b binds around 6 hpf (Choe et al., 2014). Thus, Pbx4 may bind many of its sites at these more classical enhancer regions just like Prep1.

In addition to the expected sequences, MEME and DREME consistently identify GC boxes near both TALE and NFYA peaks at 3.5 hpf (Figure A.1). This observation is not new, as previous research has noted a strong positional bias of GC boxes relative to CCAAT boxes (Dolfini et al., 2009). Further research into this relationship will be challenging as myriad factors of the Sp and Klf families bind GC boxes. One potential way to identify protein interactions here would be to use a mass spectroscopy technique such as MudPIT. Here, one could use immunoprecipitate Prep1, Pbx4, or Nfya, which would bring any interacting proteins with them. One could then use mass spectroscopy to identify the proteins in the sample and perhaps elucidate which GC box binding factors are present in the complex. In addition, this technique would identify other interacting proteins, lending insight into the composition of these complexes.
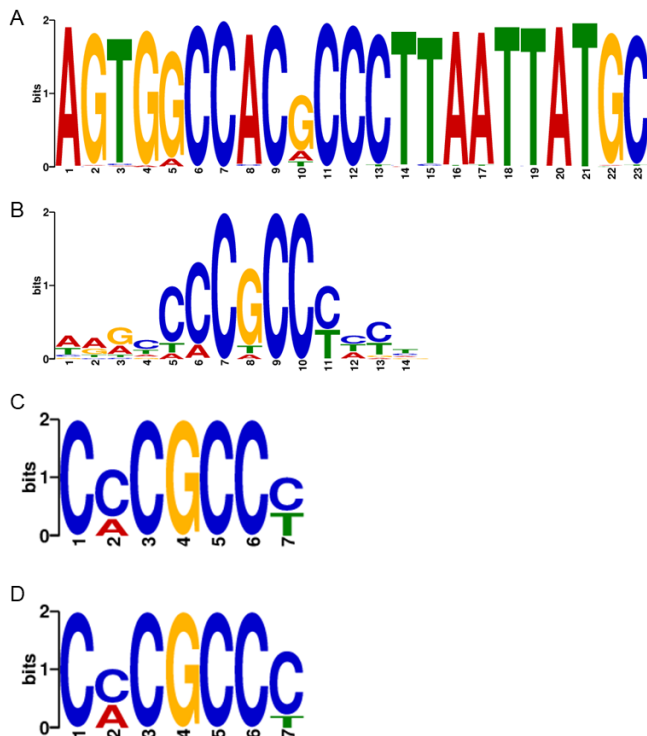
Figure A.1: GC boxes appears near DECA and CCAAT motifs. (A-B) MEME identifies the GC box motif near DECA sites co-bound by Pbx4/Prep1 (A) and near DECA/CCAAT motifs bound by Pbx4/Prep1/Nfya (B). (C-D) DREME identifies the GC box near CCAAT boxes bound by Nfya (C) and DECA motifs bound by Pbx4 (D).

TALE and NFY bind near transcription start sites

Co-bound TALE/NFY regions are generally close to transcription start sites, with about 55% occurring within 5 kB of one and more than 70% occurring within 30 kB of one (Figure 3.10 C). These number are higher than those for any of the individual factors, but of the individual factors Nfya alone is closest. This agrees with previous research, which has identified many CCAAT boxes close to promoters (Dolfini et al., 2009). Despite their original description, new insights

into CCAAT boxes are revealing that they exist beyond promoter regions and in enhancers (Oldfield et al., 2014). This appears to be the case for co-bound TALE/NFY regions based on chromatin marks. At 4.5 hpf, co-bound TALE/NFY regions are enriched for H3K27ac, which marks active enhancers (Figure 3.13 A). This remains consistent through 9 hpf (Figure 3.13 B). H3K4me1 levels, which mark active enhancers, are low at 4.5 hpf but increase substantially by 9 hpf (Figure 3.13 C, D). H3K4me1 levels decline back to baseline levels at 24 hpf (data not shown). This is consistent with a "window of opportunity" associated with H3K4me1 in which an enhancer initially opens for binding before stabilizing. Finally, H3K4me3 levels are low at 4.5 hpf but very high by 9 hpf (Figure 3.13 E, F). H3K4me3 marks active transcription, suggesting that by 9 hpf these regions are transcriptionally active. The presence of these chromatin marks suggests that they are enhancers and furthermore make sense in the context of chromatin modifying enzymes associated with each factor. Pbx4 recruits CBP (Choe et al., 2009), which deposits H3K27ac. NFY recruits the MLL complex, which deposits H3K4me3. NFY also recruits KDM1/coREST, which demethylates H3K4me2 into H3K4me1 (Fossati et al., 2011; Qureshi et al., 2010). This also provides an explanation for why these marks are higher at co-bound TALE/NFY regions than those bound by either TALE or NFY alone.

These observations are significant in several ways. It expands previous research to a genome-wide scale, demonstrating that at 3.5 hpf Pbx4 binds together with Prep1 at enhancers near genes involved in transcription regulation

and embryonic development (Figure 3.11 B, E). As transcription and

development genes are among the earliest genes expressed following ZGA, this

data implies that the TALE factors play a role in activating such genes at the

ZGA. Further, this data shows that NFY, which is generally accepted to bind

promoter regions associated with housekeeping functions, also binds enhancer

regions associated with transcription regulation and early development and plays

a role in activating these genes at ZGA (Figure 3.13 H). Finally, this data reveals

a DNA motif containing a DECA motif adjacent to a CCAAT box (Figure 3.13 L),

showing that at co-regulated genes the TALE factors and NFY bind together.

Enhancers bound by the TALE factors and NFY drive transcription

The ChIP-seq and RNA-seq data both support the notion that the TALE

factors and NFY cooperate to drive transcription from co-bound enhancers. To

test this theory, I made stable transgenic zebrafish lines in which TALE/NFY co-

bound enhancers would drive the expression of GFP. Although most of these

putative enhancers were indeed able to drive GFP expression (Figure 3.14), not

all of them could. This is likely due in part to the selection process for these DNA

elements. The putative enhancer selection happened prior to the collection of

Pbx4 or Nfya ChIP-seq data, relying instead on Prep1 ChIP-seq binding profiles

at 3.5 hpf, composition of binding sites, and the expression pattern of the nearest

gene. Selected putative enhancers contained strong Prep1 peaks, a DECA motif,

a CCAAT box, and an anterior expression pattern for the nearest gene. Furthermore, they were all roughly 500 bp in length and centered on the DECA motif with no regard for other potential binding sites nearby. Of the three DNA elements that were not able to drive GFP expression, two (*her6* and *prdm14*) lacked any Nfya peak and displayed only weak Pbx4 peaks (Figure 3.15 B, C). The inability of these DNA elements to drive gene expression seems to be because they lacked all three factors (Prep1, Pbx4, and Nfya) that appear to be necessary to activate transcription. In the case of the third DNA element, *pax5*, all three factors appeared to be present (Figure 3.15 A); however, this element lacked the ability to drive GFP expression. A potential explanation for this is that the element lacked other essential sequence information required to assemble a functional transcription complex. As the element was cut off at 500 bp, it is possible that the DNA recognition motif of another essential transcription factor normally in the endogenous enhancer was cut out of the DNA element tested, resulting in an incomplete, and therefore non-functional, transcription complex. This fits with our proposed mechanism of cooperative TALE/NFY action, in which the TALE and NFY factors use PTF activity to access their binding sites in compact chromatin and recruit chromatin modifying enzymes to form a more permissive chromatin environment and allowing other transcription factors to access their binding sites and begin to build transcription complexes (Figure 4.1). It's worth mentioning once again here that NFY has been implicated as one of the essential factors in ZGA in mice (Lu et al., 2016).

Enhancer activity correlates with ChIP peak strength

All five of the enhancer elements able to drive GFP expression (Figure 3.14 B, D, F, H, J) displayed robust Pbx4 and Prep1 peaks (Figure 3.14 A, C, E, G, I). While they all also showed Nfya peaks, several were small. In particular, the TALE and Nfya ChIP signal for *fgf8a* and *yap1* were weakest (Figure 3.14 G, I), but both factors could still drive GFP expression (Figure 3.14 H, J). It could be that these peaks grow larger at later timepoints; the ChIP tracks are from the 3.5 hpf ChIP-seq data, whereas the photos are taken at 24 hpf. As previously mentioned, Prep1 ChIP-seq data shows that many new Prep1 peaks arise between 3.5 hpf and 12 hpf (Ladam et al., 2018). Although few of these are DECA motifs, it is possible that the peaks at *fgf8a* and *yap1* grow stronger as development progresses.

The tissue-specific nature of the expression patterns among the five enhancers capable of driving GFP expression lends credence to the hypothesis that the TALE factors and NFY, which are ubiquitously expressed, are playing a more general role. If the TALE factors and NFY were driving expression themselves with no other input, the expectation would be that any enhancers capable of driving GFP expression would express GFP ubiquitously since the TALE factors and NFY would be present throughout the organism at this stage of development to bind the enhancers and activate transcription. Since all five enhancers demonstrate tissue-restricted expression, it is more likely that they

play a more generic role, opening the chromatin and exposing other tissue-specific transcription factor binding sites (Figure 4.1). In tissues where these transcription factors are expressed, the enhancer can drive GFP expression. This data supports our proposed role for the TALE factors and NFY at ZGA, where the factors utilize PTF activity to identify their binding sites in the zygotic genome and begin establishing permissive chromatin profiles around essential genes to be expressed in the first wave of zygotic gene expression, particularly genes involved in transcription regulation and embryonic development.

Enhancers are inactive in human cells

Interestingly, none of the five enhancers capable of driving GFP expression in zebrafish were capable of driving gene expression in HEK293T cells (unpublished data). This suggests that the TALE factors and NFY themselves are not solely responsible for the activity of the enhancers that they bind. Rather, it seems likely that they recruit other factors to build functional transcription complexes. These other factors are most probably chromatin modifying enzymes which help establish a permissive chromatin environment for transcription and expose binding sites for other transcription factors (Figure 4.1). In addition, the TALE factors and NFY may themselves aid in the recruitment of other factors essential for transcription activation. Being zebrafish enhancers, these DNA elements can function normally in the context of the zebrafish

proteome. In human cells such as HEK293T cells, however, differences in

binding affinity between human and fish proteins prevent the recruitment and

binding of the proper factors.

Transgenic GFP is maternally deposited

Interestingly, all five of the enhancers able to drive gene expression

appeared to exhibit maternal deposition, as many eggs are GFP-positive prior to

ZGA (Figure 3.17). These eggs invariably grow into GFP-positive embryos,

whereas GFP-negative eggs invariably grow into GFP-negative embryos (Figure

3.16 B). The mechanism behind this unusual observation is unclear; it could be

that the enhancers are active in the maternal germline, leading to deposition of

GFP. Alternatively, the mRNAs transcribed by the enhancers could be lacking an

essential component for maternal transcript repression in oocytes. For example,

the Y-box protein Ybx1 binds maternal mRNAs and represses their translation

(Sun et al., 2018). It is possible that the transcripts driven by the enhancer

elements lack a binding site for Ybx1, resulting in their translation in oocytes and

thus deposition of maternal GFP protein. While I was able to identify multiple

male founders for each of the wild-type pTransgenesis lines, I was not able to

identify any female founders to determine whether this effect persisted in a

different transgenesis system and could be attributed to the enhancers or

another component of the transgenic construct. Further screening with transgenic F1 females could provide insight into this intriguing finding.

These findings are significant in that they reveal that enhancers containing DECA motifs and CCAAT boxes are capable of driving transcription *in vivo.* Prior to this work, the DECA motif had no assigned biological function. The fact that enhancer elements containing a DECA motif can drive transcription suggests a role for the motif in driving gene expression. In addition, the majority of research describes CCAAT box activity at promoter regions (reviewed in (Dolfini et al., 2012)). The presence of CCAAT boxes in these active enhancer regions demonstrates that NFY also has a role in driving transcription in regions besides promoters. Furthermore, expression of the genes near these enhancers is generally early in development and in the anterior embryo. Although the observed patterns do not always mimic that of the nearest gene, this may be due to extenuating circumstances and does not rule out a joint role for the TALE factors and NFY in driving the expression of such early anterior genes. Taken together, this work outlines a new role for the TALE factors and NFY in activating transcription at early anterior genes.

The TALE factors and NFY are required for enhancer activity

With the knowledge that the DECA/CCAAT enhancers were sufficient to drive gene expression, I set out to determine whether they were also necessary

for it. Indeed, both the dominant negatives and the mutation of the DECA site and CCAAT box led to a loss of activity for the *tcf3a* and *tle3a* enhancers (Figure 3.18 E, F; Figure 3.19 C, G). This suggests an essential role for the TALE factors and NFY in the ability of these enhancers to drive transcription. In the case of the dominant negatives, GFP signal was not completely lost (Figure 3.18, B-F). There are many potential reasons for this, many of which also applied to the phenotype characterization and RNA-seq. Chief among these is that the TALE factors and NFY are both maternally deposited, giving endogenous Pbx4 and Nfya something of a head start in forming functional TALE and NFY complexes during the lag period where the dominant negatives are being translated. As GFP is a very stable protein, it is likely that most of the detected GFP signal originates with these functional complexes earlier in development and lingers until 18 hpf, when I scored the embryos. Another potential reason for the inconsistent results, as with the phenotype characterization (Table 3.1; Table 3.2), is that I scored these embryos individually and injections are not always completely consistent. It is conceivable that some embryos may receive less injected material than others. The needles used for microinjection have very small bores and frequently clog, dispensing less material per injection as a result. The sharpness of the needles also varies, affecting how easily they pierce the cell membrane. Duller needles cause tears, which can allow material to leak out. The RT-qPCR pooling helps to correct for these individual discrepancies, but this assay still detects GFP transcripts which likely originate from one of the aforementioned reasons.

pTransgenesis and E1b transgenic lines show similar expression patterns

In the pTransgenesis system, the wild-type enhancers for *tcf3a* and *tle3a* were able to drive transcription of RFP in similar patterns to what was observed in the E1b-GFP lines (Figure 3.19 A, E). This, along with the multiple E1b-GFP lines for *tcf3a* and *tle3a* (Table 3.9), suggests that the enhancers' activity is specific and not due to extenuating factors such as genome integration site. The mutant DECA site was previously shown to not bind the TALE factors (Vlachakis et al., 2000), whereas NFY has only ever been characterized as binding the sequence CCAAT. Nonetheless, one could perform ChIP-qPCR on the transgenic mutant embryos to confirm a loss of both TALE and NFY binding in the mutant enhancers. The loss of RFP in these mutant embryos (Figure 3.19 C, G) suggests that the TALE factors and NFY are essential to the enhancers' ability to drive transcription and are indispensable for the activity of many genes.

Not all transgenic enhancer genes are present in ChIP-seq and RNA-seq data

Comparing the enhancers capable of driving GFP expression to the ChIP-seq, I find that *tcf3a*, *tle3a* (*gro2*), *dachb*, and *fgf8a* all appear as genes near TALE and NFY peaks. *yap1* is absent from this list; however, it could be due to the weaker ChIP peaks observed for that enhancer failing to meet either a fold enrichment or statistical cutoff for one of the factors. As previously indicated, the

peaks could grow stronger as development progresses, explaining the ability of

the *yap1* enhancer to drive GFP expression. In the RNA-seq data, only *dachb*,

with a 0.59-fold change, is downregulated enough to meet the cutoff in the

PBCAB data set. As for the other genes near the enhancers in this data set, *tcf3a*

shows a 1.49-fold change, *tle3a* shows a 1.07-fold change, *fgf8a* shows a 1.16-

fold change, and *yap1* shows a 1.24-fold change. In the Nfya-DN data set, *tcf3a*

shows a 1.09-fold change, *tle3a* shows a 0.88-fold change, *dachb* shows a 0.87-

fold change, *fgf8a* shows a 0.99-fold change, and *yap1* shows a 1.01-fold

change. This presents a discrepancy with the transgenic data (Figure 3.14),

which shows that these enhancers are dependent on TALE and NFY function,

yet all but *dachb* with PBCAB fail to meet the cutoffs. There are some possible

explanations for this. As explained earlier, there is a time discrepancy between

the two sets of experiments, with the RNA-seq analysis happening at 12 hpf and

the observation for the dominant negative-injected E1b-GFP and pTransgenesis

embryos happening at 18 hpf and 32 hpf, respectively. Although all of the genes

associated with these enhancers are expressed very early, higher expression

may not begin for them until sometime between 12 hpf and 18 hpf. At this time,

the loss of the reporter gene may be more noticeable compared to a control.

Alternatively, it is possible that the endogenous enhancers contain elements

which allow for some sort of redundancy in activation which the cloned

enhancers lack. As the enhancer selection did not regard any other potential

transcription factor binding sites, it is possible that a key binding site for another

factor is missing. This would make the cloned enhancers less robust and explain

their loss of activity despite the apparent indifference of the endogenous

enhancers. In this same vein, it is possible that the dominant negatives disrupt

the larger gene network. For example, if the dominant negatives activate a

repressor that represses a second repressor that downregulates *tcf3a* expression

then this could balance out the loss of activator activity from the TALE factors

and NFY, explaining the apparent lack of change in tcf3a expression despite the

dominant negatives. Finally, it is possible that the cloned enhancers do not

directly regulate the genes nearest to them for which they are named. Further

study of additional transcription factor binding sites in these enhancers will be

necessary to elucidate the exact reasoning for this disagreement.

This data is significant in that it demonstrates that the TALE factors and

NFY are essential for the activity of enhancers they occupy. Although each factor

individually is essential at many genes, this data is the first to demonstrate that

they are essential at co-bound enhancers. Given that the TALE factors and NFY

are broadly expressed early in development but many target genes are spatially

restricted, this data suggests a generic role such as opening chromatin to expose

binding sites of tissue-specific transcription factors (Figure 4.1). The fact that

individual dominant negatives can disrupt enhancer activity suggests that both

factors are required for this activity. Furthermore, it supports a hypothesis in

which the TALE factors and NFY cooperate to open specific genomic loci at ZGA

to activate expression of transcription- and development-related genes.

Future Directions

The results from this work establish a role for the TALE factors and NFY during MZT. They also set the stage for many additional avenues of work.

Does NFY play a role in anterior central nervous system development?

Although the phenotypes caused by PBCAB and Nfya-DN were very similar, they were not identical. Specifically, while the TALE factors affected hindbrain segmentation and patterning, the regions examined posterior to the midbrain-hindbrain boundary appeared normal in the embryos injected with Nfya-DN. Previous research implicates NFY in central nervous system (CNS) development, but this data suggests that it does not have a role in CNS development posterior to the midbrain-hindbrain boundary. The anterior nature of the abnormalities in the Nfya-DN-injected animals supports the notion that CNS defects may exist and that they may be anterior to the midbrain-hindbrain boundary. Further research using more anterior markers such as *pax6a* could help further establish a role for NFY in CNS development. Alternatively, specific cell types in which the TALE factors and NFY may cooperate, such as neural crest cells, could be isolated and studied in single-cell high-throughput assays such as RNA-seq or ChIP-seq. To do this, one could generate transgenic zebrafish containing GFP driven by a promoter specific to neural crest cells, then

dissociate the embryos and isolate the neural crest cells using FACS. The latter experiment also alleviates a potential complication with the RNA-seq and ChIP-seq experiments performed in this project, in which specific signal could have been diluted due to the use of whole embryos.

What is the role of NFY in cilia-related genes?

The GO terms of genes downregulated by Nfya-DN revealed a strong enrichment for cilia-related genes. Currently, no role for NFY in cilia-related gene expression has been reported. Analysis of Nfya ChIP peaks near the cilia-related genes downregulated by Nfya-DN could reveal whether the regulation of these genes by NFY occurs at enhancers or promoters, and whether there is a preference for motile or sensory cilia-related genes. The latter could also inform experiments regarding the specific abnormalities in these embryos and perhaps tie some of the observed phenotypes in Nfya-DN-injected embryos to cilia-related deficiencies. For example, leftward fluid flow in the Kupfer's vesicle provides a developmental cue in left-right symmetry establishment during development. This fluid flow is generated by motile cilia, disruption of which could explain the left-right asymmetry observed in Nfya-DN-injected fish. Examination of this potential link between the Nfya-DN phenotype and cilia could establish yet another role for NFY in embryonic development.

Do Pbx4 and Nfya bind additional sites by 12 hpf?

By 12 hpf, Prep1 binds many additional sites that it had not bound at 3.5 hpf. These sites predominantly contained HEXA motifs whereas the sites bound at 3.5 hpf predominantly contained DECA motifs, suggesting unique roles for each motif at each stage of development (Ladam et al., 2018). The future directions of this work will almost certainly involve the collection of NFYA ChIP-seq data at 12 hpf. Comparing these results at 12 hpf to the results at 3.5 hpf will allow for better understanding of the distinct roles for the TALE factors and NFY at these different developmental stages.

How do chromatin marks at TALE/NFY enhancers change?

The TALE factors and NFY appear to have a broad role at the enhancers they bind, such as changing the surrounding chromatin profile to be more permissive for transcription factor binding and transcription. Further observation of the changes that take place in response to TALE and NFY binding could help better understand whether this is indeed the role of these factors, establish PTF activity, and identify which chromatin modifying enzymes they are recruiting to these enhancers. This would require ChIP-seq experiments on newly fertilized embryos for a variety of chromatin marks such as H3K27me3 and H3K27ac. In

addition, ChIP-seq experiments at the same time for Prep1, Pbx4, and Nfya could provide insight into just how early these factors bind.

Do the TALE factors and NFY stabilize one another's binding?

Although the TALE factors and NFY both appear to be required for enhancer activity, it is not clear whether the factors possess equal PTF capabilities or whether one factor may stabilize the binding of another. To determine whether one factor stabilizes the binding of the other, it would be possible to use ChIP-qPCR at regions where the factors bind near one another. In this experiment, embryos could be injected with one of the dominant negatives, followed by ChIP for the other and qPCR at the co-bound loci. In theory, if one factor is essential for stabilizing the other, then that factor will not be present at the co-bound locus. Alternatively, it would be possible to design an *in vitro* pull-down experiment, wherein biotinylated oligos containing various mutated versions of the DECA motif and CCAAT box could be incubated in cell lysate then pulled out of solution with binding factors attached. The factors can be eluted from the oligo, followed by a western blot to assess binding.

What other factors associate at TALE/NFY enhancers?

Although the TALE factors and NFY appear to be essential to the ability of these enhancers to drive transcription, there are likely many other factors that play roles as well. For example, GC boxes tightly associate with CCAAT boxes and DECA sites, suggesting roles for one of the many factors that bind this motif. Using a technique such as MudPIT, one could immunoprecipitate either the TALE factors or NFY and identify other factors within the complex. This would help elucidate any other essential factors as well as tissue-specific transcription factors that may bind at the opening enhancers and could also be used in conjunction with an *in vitro* binding experiment as described above, perhaps using a biotinylated version of the *tcf3a* enhancer as the bait oligo (Figure A.2).
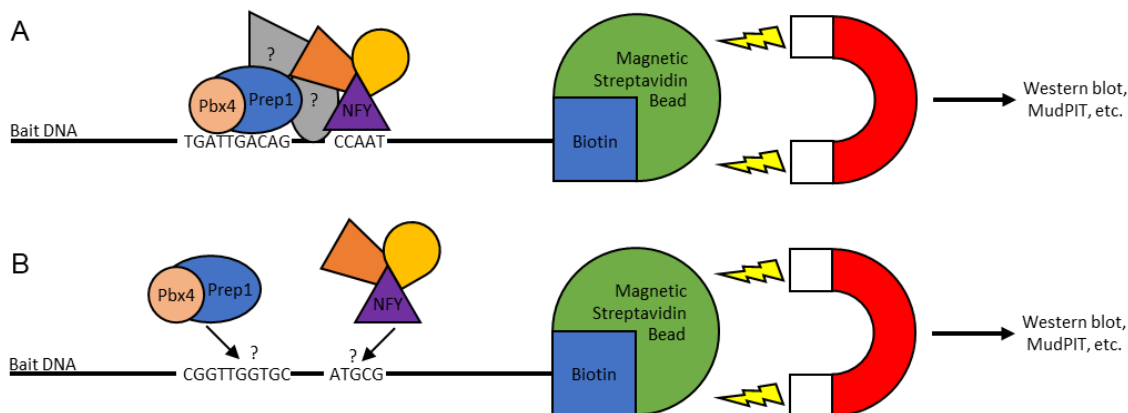
Figure A.2: Using a bait oligo to assess TALE and NFY complex assembly. Schematic of a proposed experiment using a bait oligo to assess properties of TALE and NFY complex assembly. (A-B) Biotinylated bait oligos containing wild-type (A) or mutated (B) DECA/CCAAT motifs are incubated in 3.5 hpf zebrafish cell lysate then pulled out of solution using magnetic streptavidin-coated beads. Proteins bound to the bait oligo are then eluted for downstream applications, such as western blots to assess the presence of specific factors or MudPIT to identify other factors.

Concluding Remarks

This work provides evidence that the TALE factors and NFY work together at enhancers to drive the transcription of genes involved in transcription regulation and embryonic development at the ZGA. Dominant negative versions of each factor lead to similar phenotypes in zebrafish embryos and reveal common as well as independent sets of genes that each factor regulates. ChIP-seq data helps narrow these differentially expressed genes to those most likely to be regulated directly by each factor and shows how the chromatin profile surrounding these sites resembles enhancers as they become active. Enhancers bound by the TALE factors and NFY can drive transcription, and loss of either factor results in loss of activator activity. Taken together, these results suggest a pioneer role for the TALE factors and NFY at ZGA, where the factors begin opening the chromatin profile at their binding sites to initiate transcription of key transcription and development genes.

APPENDIX B: COMPARISON OF DIFFERENTIALLY EXPRESSED GENES IN ZEBRAFISH EMBRYOS TREATED WITH DOMINANT NEGATIVE PBX4 OR TALE MORPHOLINOS

Background

Antisense Morpholino oligos (MOs) are a standard method for disrupting gene expression in reverse genetics. Rather than the standard nucleic acid backbone of ribose sugars and phosphates, MO backbones consist of methylenemorpholine rings linked through phosphorodiamidate groups. The methylenemorpholine ring provides an attachment point for DNA bases, which can form standard Watson-Crick base pairs with RNA or single-stranded DNA. By designing a MO with the complementary sequence to a key region in a target mRNA, the MO can hybridize at the target site of the target mRNA and block the interaction of that region with key factors. In this way, MOs can block certain steps in gene expression, such as splicing or translation, and prevent the creation of the target protein. Blocking the expression of the target protein, or knocking it down, often results in abnormal phenotypes, lending insight into the function of the target gene. Thus, morphant phenotypes have become a staple in reverse genetic studies (reviewed in (Heasman, 2002)).

Due to their simple premise, researchers use MOs to disrupt gene expression in a variety of organisms. MOs have even gained approval from the US Food and Drug Administration in the treatment of Duchenne muscular dystrophy as the drug eteplirsen. In the laboratory, MOs are commonly used to disrupt gene function in zebrafish. However, as advancements in in gene editing technology, such as CRISPR/Cas9, facilitate the generation of germline mutants,

the reliability of MOs in gene disruption has come into question. One study compared the morphant phenotype to the germline mutant phenotype of 20 genes in zebrafish. Of the 20 morphant phenotypes, only ten accurately recapitulated the germline mutant phenotype. The same group also showed that, of the germline mutants from the Sanger Zebrafish Mutation Project, only 20% aligned with corresponding morphant phenotypes (Kok et al., 2015). While the specific reasons for this incongruity remain a hot topic of debate, the fact remains that the accuracy of morphant phenotypes is no longer a sure thing.

Previously, our lab used a cocktail of MOs targeting *pbx4*, *pbx2*, and *prep1.1* to disrupt TALE function in zebrafish embryos. We observed that the TALE MOs caused the embryos to have smaller heads, smaller eyes, and swollen pericardium/cardiac edema by 24 hpf (Ladam et al., 2018). This phenotype accurately mirrored that of existing *pbx4* and *prep1.1* germline mutants, indicating that the morphant phenotype was specific (Deflorian et al., 2004; Pöpperl et al., 2000; Waskiewicz et al., 2002). Using the TALE MOs, we generated RNA-seq libraries from zebrafish embryos that had been injected with either TALE MOs or a control at 12 hpf. DAVID analysis of genes differentially expressed between the TALE MO and control conditions showed downregulation of genes involved with transcription, embryonic development, and homeobox genes. These results agreed with published literature regarding the function of the TALE factors (reviewed in (Ladam & Sagerström, 2014)).

Aside from MOs and germline mutants, dominant negatives can disrupt TALE function. One such dominant negative is PBCAB, which is a truncated version of *pbx4* lacking the C-terminal domain. PBCAB can still form heterodimers with Prep1; however, it lacks the ability to cross the nuclear membrane. When expressed in saturating amounts, PBCAB can outcompete endogenous Pbx4 for Prep1 binding and sequester it in the cytoplasm, resulting in a loss of TALE function (Choe et al., 2002). Zebrafish embryos injected with PBCAB display the same phenotype as embryos injected with TALE MOs, suggesting that their effect is consistent with previously published data regarding TALE disruption (Figure 3.2 C; Figure 3.3 E, F; Figure 3.4 E, F; Table 3.1; Table 3.2; Table 3.3). In addition, our lab generated RNA-seq libraries from zebrafish embryos injected with either PBCAB or GFP (control) mRNA at 12 hpf. Analysis of the downregulated genes using DAVID revealed disruption of genes involved with transcription regulation, embryonic development, and homeobox genes (Figure 3.7 A, B). However, comparison of the specific downregulated genes reveals limited overlap of the genes. Here, I present a comparison of genes downregulated by PBCAB to genes downregulated by TALE MOs.

Methods

The methodology used in this Appendix is presented in Chapter II: Materials & Methods.

Results

To assess which genes' expression PBCAB and the TALE MOs disrupted, I used DESeq2 in Dolphin (Anders & Huber, 2010). For further analysis, I only considered genes with a fold-change of 1.5 or greater and an adjusted p-value of 0.01 or less. For PBCAB, this yielded 646 downregulated and 854 upregulated genes. For the TALE MOs, it yielded 526 downregulated genes and 145 upregulated genes (Figure B.1 A). Comparing the genes downregulated by PBCAB to those downregulated by TALE MOs, I found 149 genes downregulated in both conditions (Figure B.1 B). GO term analysis using the DAVID tool revealed that these 149 genes downregulated by both PBCAB and TALE MOs functioned primarily in roles related to transcription, embryonic development, and homeobox genes (Figure B.1 C). Comparing the genes upregulated by PBCAB to the genes upregulated by TALE MOs, I identified 51 genes common to the two conditions (Figure B.1 D). The GO terms of these 51 genes were also enriched for processes related to transcription, development, and homeobox genes (Figure B.1 E).
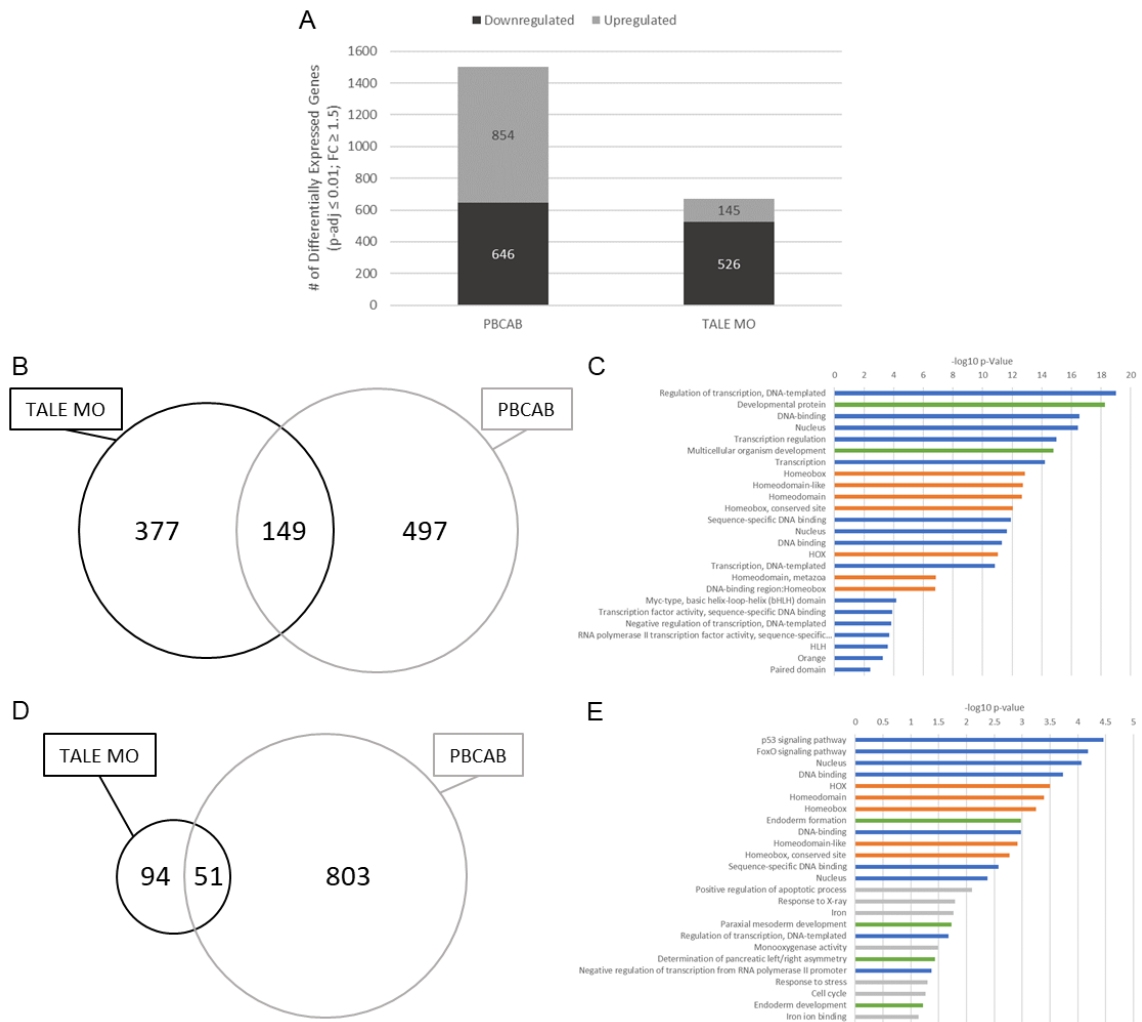
Figure B.1: PBCAB and TALE MOs alter the expression of distinct sets of genes. (A-B) Venn diagrams showing genes downregulated (A) or upregulated (B) in embryos injected with PBCAB or TALE MOs. (C-D) Top GO terms returned by DAVID for genes downregulated (C) or upregulated (D) by both PBCAB and TALE MOs. Blue bars correspond to transcription-related, green to embryogenesis-related, orange to homeodomain-related, and gray bars to other ontologies.

Discussion

Dominant negatives and antisense Morpholino oligos (MOs) are different methodologies to disrupt gene expression. Here, I generated RNA-seq libraries from 12 hpf zebrafish embryos injected with PBCAB, a dominant negative Pbx4 (Choe et al., 2002), or with GFP as a control. I performed differential expression analysis using DESeq2 to identify which genes' expression changed between the PBCAB and GFP conditions (Anders & Huber, 2010), finding 646 downregulated and 854 upregulated genes in the PBCAB condition. Using previously published RNA-seq data from 12 hpf zebrafish embryos injected with either TALE MOs or a control, I performed the same differential expression analysis and identified 526 downregulated and 145 upregulated genes (Figure B.1 A). Comparing the two lists, I found 149 common downregulated genes and 51 common upregulated genes (Figure B.1 B, D). DAVID revealed that both the 149 common downregulated and 51 common upregulated genes were enriched for GO terms involved with transcription, embryonic development, and homeobox genes (Figure B.1 C, E).

The bulk of the literature describes activating activity for the TALE factors, which would suggest that downregulated genes are most likely to be due to direct effects (reviewed in (Ladam & Sagerström, 2014). Although PBCAB and the TALE MOs utilize different mechanisms to attenuate TALE function, the set of genes they disrupt should be relatively similar, as the majority should be

attributed to the loss of TALE function. However, I find only 149 genes with at least a 1.5-fold change (23.1% (149/646) of PBCAB, 28.3% (149/526 of TALE MO) in both conditions (Figure B.1 B). Outwardly, PBCAB and the TALE MOs result in very similar phenotypes, causing smaller heads, smaller eyes, and swollen pericardium/cardiac edema in zebrafish embryos by 24 hpf (Figure 3.2 C; (Ladam et al., 2018)). However, based on these results, it appears that PBCAB and the TALE MOs disrupt the expression of distinct sets of genes.

Although most research implicates the TALE factors as transcriptional activators, there are several reports of the TALE factors possessing repressive activity as well. For example, Pbx4 can recruit histone deacetylases, which can remodel chromatin to be less permissive to transcription (Choe et al., 2009; Saleh et al., 2000). This repressive activity suggests that some of the genes upregulated by PBCAB or the TALE MOs may also be due to direct activity of the TALE factors. DESeq2 analysis revealed 854 genes upregulated by PBCAB and only 145 upregulated by the TALE MOs, which in and of itself is a very large discrepancy (Figure B.1 A). Of these genes, only 51 are upregulated by both PBCAB and TALE MOs (6% of total PBCAB and 35.2% of total TALE MO) (Figure B.1 D). Although many of these may be due to indirect effects, and therefore could be attributable to PBCAB or the TALE MOs themselves, the fact remains that overlap between the two methodologies is far smaller than expected.

Despite the relatively small overlap among differentially expressed genes, the GO terms enriched within those overlapping differentially expressed gene sets are generally consistent with the GO terms enriched within the entire PBCAB and TALE MO sets. This is not terribly surprising since both the PBCAB and TALE MO sets themselves affect the same classes of genes. Genes downregulated by PBCAB and the TALE MOs consist of genes involved with transcription regulation, embryonic development, and homeodomains (Figure B.1 C), which are the same categories as the genes downregulated by PBCAB alone and the TALE MOs alone (Figure 3.7 A, B; (Ladam et al., 2018)). The distribution of the categories within each condition is slightly different, with genes downregulated by PBCAB generally more highly enriched for general transcription regulation and having a greater number of uncategorized terms (Figure 3.7 A, B). The genes downregulated by the TALE MOs alone are more enriched for homeobox genes (Ladam et al., 2018). The GO terms downregulated by both PBCAB and the TALE MOs appear closer to that of PBCAB alone, but retain no uncategorized terms, suggesting that the uncategorized terms are predominantly exclusive to the different methodologies.

Although the categorical distribution for upregulated genes is more consistent between those upregulated by PBCAB and those upregulated by both PBCAB and TALE MOs, some of the categories are different. For example, the top result for genes upregulated by both factors is the p53 signaling pathway. Although related to transcription, p53 activation is cited as a potential off-target

effect of MOs, calling the specificity of these results into question (Gerety &

Wilkinson, 2011). Furthermore, as previously mentioned, upregulated genes are

more likely to be indirect effects based on the fact that the TALE factors are

generally described as transcriptional activators (Ladam & Sagerström, 2014).

In conclusion, PBCAB and the TALE MOs cause very similar phenotypes

in zebrafish embryos by 24 hpf. These phenotypes closely mirror those of the

pbx4 germline mutants (Pöpperl et al., 2000), suggesting that they are specific

and accurate. However, RNA-seq analysis at 12 hpf of the genes differentially

expressed in zebrafish injected with PBCAB or the TALE MOs reveals limited set

of differentially expressed genes. While it appears safe to assume that both

PBCAB and TALE MOs accurately display phenotypes associated with disruption

of each factor, the difference in GO terms between each leaves some room for

uncertainty.

REFERENCES

Aanes, H., Winata, C. L., Lin, C. H., Chen, J. P., Srinivasan, K. G., Lee, S. G. P., … Mathavan, S. (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Research*, *21*(8), 1328–1338. https://doi.org/10.1101/gr.116012.110

Abu-Shaar, M., Ryoo, H. D., & Mann, R. S. (1999). Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. *Genes & Development*, *13*(8), 935–945. https://doi.org/10.1101/gad.13.8.935

Amin, S., Donaldson, I. J., Zannino, D. A., Hensman, J., Rattray, M., Losa, M., … Bobola, N. (2015). Hoxa2 selectively enhances Meis binding to change a branchial arch ground state. *Developmental Cell*, *32*(3), 265–277. https://doi.org/10.1016/j.devcel.2014.12.024

Amodeo, A. A., Jukam, D., Straight, A. F., & Skotheim, J. M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), E1086-95. https://doi.org/10.1073/pnas.1413990112

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Asahara, H., Dutta, S., Kao, H. Y., Evans, R. M., & Montminy, M. (1999). Pbx-Hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner. *Molecular and Cellular Biology*, *19*(12), 8219–8225. https://doi.org/10.1128/mcb.19.12.8219

Aspland, S. E., & White, R. A. (1997). Nucleocytoplasmic localisation of extradenticle protein is spatially regulated throughout development in Drosophila. *Development (Cambridge, England)*, *124*(3), 741–747. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9043089

Azcoitia, V., Aracil, M., Martínez-A, C., & Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Developmental Biology*, *280*(2), 307–320. https://doi.org/10.1016/j.ydbio.2005.01.004

Bailey, J. S., Rave-Harel, N., McGillivray, S. M., Coss, D., & Mellon, P. L. (2004). Activin regulation of the follicle-stimulating hormone beta-subunit gene involves Smads and the TALE homeodomain proteins Pbx1 and Prep1. *Molecular Endocrinology (Baltimore, Md.)*, *18*(5), 1158–1170. https://doi.org/10.1210/me.2003-0442

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202-8. https://doi.org/10.1093/nar/gkp335

Bashirullah, A., Cooperstock, R. L., & Lipshitz, H. D. (2001). Spatial and temporal control of RNA stability. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(13), 7025–7028. https://doi.org/10.1073/pnas.111145698

Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L., & Hunter, C. P. (2003). Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. *Development (Cambridge, England)*, *130*(5), 889–900. https://doi.org/10.1242/dev.00302

Baxevanis, A. D., Arents, G., Moudrianakis, E. N., & Landsman, D. (1995). A variety of DNA-

binding and multimeric proteins contain the histone fold motif. *Nucleic Acids Research*, *23*(14), 2685–2691. https://doi.org/10.1093/nar/23.14.2685

Bazzini, A. A., Lee, M. T., & Giraldez, A. J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science (New York, N.Y.)*, *336*(6078), 233–237. https://doi.org/10.1126/science.1215704

Berkes, C. A., Bergstrom, D. A., Penn, B. H., Seaver, K. J., Knoepfler, P. S., & Tapscott, S. J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Molecular Cell*, *14*(4), 465–477. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15149596

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., & Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. *Genes & Development*, *13*(8), 946–953. https://doi.org/10.1101/gad.13.8.946

Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F., & Blasi, F. (1998). The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. *The EMBO Journal*, *17*(5), 1434–1445. https://doi.org/10.1093/emboj/17.5.1434

Berthelsen, J., Zappavigna, V., Mavilio, F., & Blasi, F. (1998). Prep1, a novel functional partner of Pbx proteins. *The EMBO Journal*, *17*(5), 1423–1433. https://doi.org/10.1093/emboj/17.5.1423

Bertolino, E., Reimund, B., Wildt-Perinic, D., & Clerc, R. G. (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *The Journal of Biological Chemistry*, *270*(52), 31178–31188. https://doi.org/10.1074/jbc.270.52.31178

Bhattacharya, A., Deng, J. M., Zhang, Z., Behringer, R., de Crombrugghe, B., & Maity, S. N. (2003). The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. *Cancer Research*, *63*(23), 8167–8172. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14678971

Bogdanović, O., Fernandez-Miñán, A., Tena, J. J., De La Calle-Mustienes, E., Hidalgo, C., Van Kruysbergen, I., … Gómez-Skarmeta, J. L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*, *22*(10), 2043–2053. https://doi.org/10.1101/gr.134833.111

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bošković, A., Eid, A., Pontabry, J., Ishiuchi, T., Spiegelhalter, C., Raghu Ram, E. V. S., … Torres-Padilla, M.-E. (2014). Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes & Development*, *28*(10), 1042–1047. https://doi.org/10.1101/gad.238881.114

Brendolan, A., Ferretti, E., Salsi, V., Moses, K., Quaggin, S., Blasi, F., … Selleri, L. (2005). A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. *Development (Cambridge, England)*, *132*(13), 3113–3126. https://doi.org/10.1242/dev.01884

Budry, L., Balsalobre, A., Gauthier, Y., Khetchoumian, K., L'honoré, A., Vallette, S., … Drouin, J. (2012). The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. *Genes & Development*, *26*(20), 2299–2310. https://doi.org/10.1101/gad.200436.112

Bürglin, T R, & Ruvkun, G. (1992). New motif in PBX genes. *Nature Genetics*, *1*(5), 319–320. https://doi.org/10.1038/ng0892-319

Bürglin, Thomas R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Research*, *25*(21), 4173–4180. https://doi.org/10.1093/nar/25.21.4173

Buschlen, S., Amillet, J.-M., Guiard, B., Fournier, A., Marcireau, C., & Bolotin-Fukuhara, M. (2003). The S. Cerevisiae HAP complex, a key regulator of mitochondrial function, coordinates nuclear and mitochondrial gene expression. *Comparative and Functional Genomics*, *4*(1), 37–46. https://doi.org/10.1002/cfg.254

Cai, M., Langer, E. M., Gill, J. G., Satpathy, A. T., Albring, J. C., KC, W., … Murphy, K. M. (2012). Dual actions of Meis1 inhibit erythroid progenitor development and sustain general hematopoietic cell proliferation. *Blood*, *120*(2), 335–346. https://doi.org/10.1182/blood-2012-01-403139

Capellini, T. D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D., … Selleri, L. (2006). Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development (Cambridge, England)*, *133*(11), 2263–2273. https://doi.org/10.1242/dev.02395

Capellini, T. D., Handschuh, K., Quintana, L., Ferretti, E., Di Giacomo, G., Fantini, S., … Selleri, L. (2011). Control of pelvic girdle development by genes of the Pbx family and Emx2. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *240*(5), 1173–1189. https://doi.org/10.1002/dvdy.22617

Capellini, T. D., Vaccari, G., Ferretti, E., Fantini, S., He, M., Pellegrini, M., … Zappavigna, V. (2010). Scapula development is governed by genetic interactions of Pbx1 with its family members and with Emx2 via their cooperative control of Alx1. *Development (Cambridge, England)*, *137*(15), 2559–2569. https://doi.org/10.1242/dev.048819

Capellini, T. D., Zewdu, R., Di Giacomo, G., Asciutti, S., Kugler, J. E., Di Gregorio, A., & Selleri, L. (2008). Pbx1/Pbx2 govern axial skeletal development by controlling Polycomb and Hox in mesoderm and Pax1/Pax9 in sclerotome. *Developmental Biology*, *321*(2), 500–514. https://doi.org/10.1016/j.ydbio.2008.04.005

Ceribelli, M., Benatti, P., Imbriano, C., & Mantovani, R. (2009). NF-YC complexity is generated by dual promoters and alternative splicing. *The Journal of Biological Chemistry*, *284*(49), 34189–34200. https://doi.org/10.1074/jbc.M109.008417

Chang, C. P., Jacobs, Y., Nakamura, T., Jenkins, N. A., Copeland, N. G., & Cleary, M. L. (1997). Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Molecular and Cellular Biology*, *17*(10), 5679–5687. https://doi.org/10.1128/mcb.17.10.5679

Chang, H., Yeo, J., Kim, J.-G., Kim, H., Lim, J., Lee, M., … Kim, V. N. (2018). Terminal Uridylyltransferases Execute Programmed Clearance of Maternal Transcriptome in Vertebrate Embryos. *Molecular Cell*, *70*(1), 72-82.e7.

https://doi.org/10.1016/j.molcel.2018.03.004

Chen, F., Ogawa, K., Liu, X., Stringfield, T. M., & Chen, Y. (2002). Repression of Smad2 and Smad3 transactivating activity by association with a novel splice variant of CCAAT-binding factor C subunit. *The Biochemical Journal*, *364*(Pt 2), 571–577. https://doi.org/10.1042/BJ20011703

Chen, Y. H., Lin, Y. T., & Lee, G. H. (2009). Novel and unexpected functions of zebrafish CCAAT box binding transcription factor (NF-Y) B subunit during cartilages development. *Bone*, *44*(5), 777–784. https://doi.org/10.1016/j.bone.2009.01.374

Chen, Z., & Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nature Genetics*, *51*(6), 947–951. https://doi.org/10.1038/s41588-019-0418-7

Choe, S.-K., Ladam, F., & Sagerström, C. G. (2014). TALE factors poise promoters for activation by Hox proteins. *Developmental Cell*, *28*(2), 203–211. https://doi.org/10.1016/j.devcel.2013.12.011

Choe, S.-K., Lu, P., Nakamura, M., Lee, J., & Sagerström, C. G. (2009). Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis. *Developmental Cell*, *17*(4), 561–567. https://doi.org/10.1016/j.devcel.2009.08.007

Choe, S.-K., Vlachakis, N., & Sagerström, C. G. (2002). Meis family proteins are required for hindbrain development in the zebrafish. *Development (Cambridge, England)*, *129*(3), 585–595. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11830560

Cirillo, L. A., & Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Molecular Cell*, *4*(6), 961–969. Retrieved from https://pdf.sciencedirectassets.com/272198/1-s2.0-S1097276500X00329/1-s2.0-S1097276500802257/main.pdf?x-amz-security-token=AgoJb3JpZ2luX2VjEC8aCXVzLWVhc3QtMSJHMEUCICXNZTZiVIJN%2BCJtLCqhJx2v8%2Bv11Bzds6TylwtbIRTnAiEA487TIeiC8F6fC1JDCANZUV24Cy%2Btx4h21ByCZB

D'Agostino, I., Merritt, C., Chen, P.-L., Seydoux, G., & Subramaniam, K. (2006). Translational repression restricts expression of the C. elegans Nanos homolog NOS-2 to the embryonic germline. *Developmental Biology*, *292*(1), 244–252. https://doi.org/10.1016/j.ydbio.2005.11.046

Dalle Nogare, D. E., Pauerstein, P. T., & Lane, M. E. (2009). G2 acquisition by transcription-independent mechanism at the zebrafish midblastula transition. *Developmental Biology*, *326*(1), 131–142. https://doi.org/10.1016/j.ydbio.2008.11.002

De Iaco, A., Coudray, A., Duc, J., & Trono, D. (2019). DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Reports*, *20*(5), e47382. https://doi.org/10.15252/embr.201847382

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nature Genetics*, *49*(6), 941–945. https://doi.org/10.1038/ng.3858

De Kumar, B., Parker, H. J., Paulson, A., Parrish, M. E., Pushel, I., Singh, N. P., … Krumlauf, R. (2017). HOXA1 and TALE proteins display cross-regulatory interactions and form a combinatorial binding code on HOXA1 targets. *Genome Research*, *27*(9), 1501–1512.

https://doi.org/10.1101/gr.219386.116

De Renzis, S., Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biology*, *5*(5), e117. https://doi.org/10.1371/journal.pbio.0050117

Deflorian, G., Tiso, N., Ferretti, E., Meyer, D., Blasi, F., Bortolussi, M., & Argenton, F. (2004). Prep1.1 has essential genetic functions in hindbrain development and cranial neural crest cell differentiation. *Development (Cambridge, England)*, *131*(3), 613–627. https://doi.org/10.1242/dev.00948

Deng, H., Sun, Y., Zhang, Y., Luo, X., Hou, W., Yan, L., … Zhang, H. (2007). Transcription factor NFY globally represses the expression of the C. elegans Hox gene Abdominal-B homolog egl-5. *Developmental Biology*, *308*(2), 583–592. https://doi.org/10.1016/j.ydbio.2007.05.021

Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzel, L., … Neugebauer, K. M. (2017). Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Research*, *27*(7), 1184–1194. https://doi.org/10.1101/gr.215954.116

DiMartino, J. F., Selleri, L., Traver, D., Firpo, M. T., Rhee, J., Warnke, R., … Cleary, M. L. (2001). The Hox cofactor and proto-oncogene Pbx1 is required for maintenance of definitive hematopoiesis in the fetal liver. *Blood*, *98*(3), 618–626. https://doi.org/10.1182/blood.v98.3.618

Dolfini, D., Gatta, R., & Mantovani, R. (2012). NF-Y and the transcriptional activation of CCAAT promoters. *Critical Reviews in Biochemistry and Molecular Biology*, *47*(1), 29–49. https://doi.org/10.3109/10409238.2011.628970

Dolfini, D., Zambelli, F., Pavesi, G., & Mantovani, R. (2009). A perspective of promoter architecture from the CCAAT box. *Cell Cycle (Georgetown, Tex.)*, *8*(24), 4127–4137. https://doi.org/10.4161/cc.8.24.10240

Donati, G., Gatta, R., Dolfini, D., Fossati, A., Ceribelli, M., & Mantovani, R. (2008). An NF-Y-dependent switch of positive and negative histone methyl marks on CCAAT promoters. *PloS One, 3*(4), e2066. https://doi.org/10.1371/journal.pone.0002066

Drummond, I. A. (2012). Cilia functions in development. *Current Opinion in Cell Biology*, *24*(1), 24–30. https://doi.org/10.1016/j.ceb.2011.12.007

Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., & Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes & Development*, *33*(3–4), 194–208. https://doi.org/10.1101/gad.321174.118

Essner, J. J., Vogan, K. J., Wagner, M. K., Tabin, C. J., Yost, H. J., & Brueckner, M. (2002). Conserved function for embryonic nodal cilia. *Nature*, *418*(6893), 37–38. https://doi.org/10.1038/418037a

Ferretti, E, Marshall, H., Pöpperl, H., Maconochie, M., Krumlauf, R., & Blasi, F. (2000). Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. *Development (Cambridge, England)*, *127*(1), 155–166. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10654609

Ferretti, Elisabetta, Cambronero, F., Tümpel, S., Longobardi, E., Wiedemann, L. M., Blasi, F., & Krumlauf, R. (2005). Hoxb1 enhancer and control of rhombomere 4 expression: complex

interplay between PREP1-PBX1-HOXB1 binding sites. *Molecular and Cellular Biology*, *25*(19), 8541–8552. https://doi.org/10.1128/MCB.25.19.8541-8552.2005

Ferretti, Elisabetta, Li, B., Zewdu, R., Wells, V., Hebert, J. M., Karner, C., … Selleri, L. (2011). A conserved Pbx-Wnt-p63-Irf6 regulatory module controls face morphogenesis by promoting epithelial apoptosis. *Developmental Cell*, *21*(4), 627–641. https://doi.org/10.1016/j.devcel.2011.08.005

Ferretti, Elisabetta, Schulz, H., Talarico, D., Blasi, F., & Berthelsen, J. (1999). The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse. *Mechanisms of Development*, *83*(1–2), 53–64. https://doi.org/10.1016/S0925-4773(99)00031-3

Ferretti, Elisabetta, Villaescusa, J. C., Di Rosa, P., Fernandez-Diaz, L. C., Longobardi, E., Mazzieri, R., … Blasi, F. (2006). Hypomorphic mutation of the TALE gene Prep1 (pKnox1) causes a major reduction of Pbx and Meis proteins and a pleiotropic embryonic phenotype. *Molecular and Cellular Biology*, *26*(15), 5650–5662. https://doi.org/10.1128/MCB.00313-06

Forsburg, S. L., & Guarente, L. (1989). Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. *Genes & Development*, *3*(8), 1166–1178. https://doi.org/10.1101/gad.3.8.1166

Fossati, A., Dolfini, D., Donati, G., & Mantovani, R. (2011). NF-Y recruits Ash2L to impart H3K4 trimethylation on CCAAT promoters. *PloS One*, *6*(3), e17220. https://doi.org/10.1371/journal.pone.0017220

Frontini, M., Imbriano, C., Manni, I., & Mantovani, R. (2004). Cell cycle regulation of NF-YC nuclear localization. *Cell Cycle (Georgetown, Tex.)*, *3*(2), 217–222. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14712092

Gagnon, J. A., Obbad, K., & Schier, A. F. (2018). The primary role of zebrafish nanog is in extra-embryonic tissue. *Development (Cambridge, England)*, *145*(1), dev147793. https://doi.org/10.1242/dev.147793

Gallo, C. M., Munro, E., Rasoloson, D., Merritt, C., & Seydoux, G. (2008). Processing bodies and germ granules are distinct RNA granules that interact in C. elegans embryos. *Developmental Biology*, *323*(1), 76–87. https://doi.org/10.1016/j.ydbio.2008.07.008

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., … Liu, J. (2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell*, *173*(1), 248-259.e15. https://doi.org/10.1016/j.cell.2018.02.028

Gardner, K. E., Allis, C. D., & Strahl, B. D. (2011). Operating on chromatin, a colorful language where context matters. *Journal of Molecular Biology*, *409*(1), 36–46. https://doi.org/10.1016/j.jmb.2011.01.040

Gerety, S. S., & Wilkinson, D. G. (2011). Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development. *Developmental Biology*, *350*(2), 279–289. https://doi.org/10.1016/j.ydbio.2010.11.030

Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, *11*(8), R86. https://doi.org/10.1186/gb-2010-11-8-r86

Golonzhka, O., Nord, A., Tang, P. L. F., Lindtner, S., Ypsilanti, A. R., Ferretti, E., … Rubenstein,

J. L. R. (2015). Pbx Regulates Patterning of the Cerebral Cortex in Progenitors and Postmitotic Neurons. *Neuron*, *88*(6), 1192–1207. https://doi.org/10.1016/j.neuron.2015.10.045

Graindorge, A., Thuret, R., Pollet, N., Osborne, H. B., & Audic, Y. (2006). Identification of post-transcriptionally regulated Xenopus tropicalis maternal mRNAs by microarray. *Nucleic Acids Research*, *34*(3), 986–995. https://doi.org/10.1093/nar/gkj492

Grebbin, B. M., Hau, A.-C., Groß, A., Anders-Maurer, M., Schramm, J., Koss, M., … Schulte, D. (2016). Pbx1 is required for adult SVZ neurogenesis. *Development (Cambridge, England)*, dev.128033. https://doi.org/10.1242/dev.128033

Grieder, N. C., Marty, T., Ryoo, H. D., Mann, R. S., & Affolter, M. (1997). Synergistic activation of a Drosophila enhancer by HOM/EXD and DPP signaling. *The EMBO Journal*, *16*(24), 7402–7410. https://doi.org/10.1093/emboj/16.24.7402

Gross, J. M., Perkins, B. D., Amsterdam, A., Egaña, A., Darland, T., Matsui, J. I., … Dowling, J. E. (2005). Identification of zebrafish insertional mutants with defects in visual system development and function. *Genetics*, *170*(1), 245–261. https://doi.org/10.1534/genetics.104.039727

Gurtner, A., Fuschi, P., Magi, F., Colussi, C., Gaetano, C., Dobbelstein, M., … Piaggio, G. (2008). NF-Y dependent epigenetic modifications discriminate between proliferating and postmitotic tissue. *PloS One*, *3*(4), e2047. https://doi.org/10.1371/journal.pone.0002047

Hahn, S., Pinkham, J., Wei, R., Miller, R., & Guarente, L. (1988). The HAP3 regulatory locus of Saccharomyces cerevisiae encodes divergent overlapping transcripts. *Molecular and Cellular Biology*, *8*(2), 655–663. https://doi.org/10.1128/mcb.8.2.655

Hamatani, T., Carter, M. G., Sharov, A. A., & Ko, M. S. H. (2004). Dynamics of global gene expression changes during mouse preimplantation development. *Developmental Cell*, *6*(1), 117–131. Retrieved from www.developmentalcell.com/cgi/content/full/6/1/

Hanley, O., Zewdu, R., Cohen, L. J., Jung, H., Lacombe, J., Philippidou, P., … Dasen, J. S. (2016). Parallel Pbx-Dependent Pathways Govern the Coalescence and Fate of Motor Columns. *Neuron*, *91*(5), 1005–1020. https://doi.org/10.1016/j.neuron.2016.07.043

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R., & Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genetics*, *7*(10), e1002266. https://doi.org/10.1371/journal.pgen.1002266

Harvey, S. A., Sealy, I., Kettleborough, R., Fenyes, F., White, R., Stemple, D., & Smith, J. C. (2013). Identification of the zebrafish maternal and paternal transcriptomes. *Development (Cambridge, England)*, *140*(13), 2703–2710. https://doi.org/10.1242/dev.095091

Heasman, J. (2002). Morpholino oligos: making sense of antisense? *Developmental Biology*, *243*(2), 209–214. https://doi.org/10.1006/dbio.2001.0565

Hendrickson, P. G., Doráis, J. A., Grow, E. J., Whiddon, J. L., Lim, J.-W., Wike, C. L., … Cairns, B. R. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nature Genetics*, *49*(6), 925–934. https://doi.org/10.1038/ng.3844

Hiller, M., Agarwal, S., Notwell, J. H., Parikh, R., Guturu, H., Wenger, A. M., & Bejerano, G. (2013). Computational methods to detect conserved non-genic elements in phylogenetically isolated genomes: Application to zebrafish. *Nucleic Acids Research*, *41*(15). https://doi.org/10.1093/nar/gkt557

Hisa, T., Spence, S. E., Rachel, R. A., Fujita, M., Nakamura, T., Ward, J. M., … Copeland, N. G. (2004). Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *The EMBO Journal*, *23*(2), 450–459. https://doi.org/10.1038/sj.emboj.7600038

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1–13. https://doi.org/10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, S., Ling, J. J., Yang, S., Li, X.-J., & Li, S. (2011). Neuronal expression of TATA box-binding protein containing expanded polyglutamine in knock-in mice reduces chaperone protein response by impairing the function of nuclear factor-Y transcription factor. *Brain : A Journal of Neurology*, *134*(Pt 7), 1943–1958. https://doi.org/10.1093/brain/awr146

Hudry, B., Remacle, S., Delfini, M.-C., Rezsohazy, R., Graba, Y., & Merabet, S. (2012). Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors. *PLoS Biology*, *10*(6), e1001351. https://doi.org/10.1371/journal.pbio.1001351

Hurtado, R., Zewdu, R., Mtui, J., Liang, C., Aho, R., Kurylo, C., … Herzlinger, D. (2015). Pbx1-dependent control of VMC differentiation kinetics underlies gross renal vascular patterning. *Development (Cambridge, England)*, *142*(15), 2653–2664. https://doi.org/10.1242/dev.124776

Imoto, I., Sonoda, I., Yuki, Y., & Inazawa, J. (2001). Identification and characterization of human PKNOX2, a novel homeobox-containing gene. *Biochemical and Biophysical Research Communications*, *287*(1), 270–276. https://doi.org/10.1006/bbrc.2001.5578

Ivanova, I., Much, C., Di Giacomo, M., Azzi, C., Morgan, M., Moreira, P. N., … O'Carroll, D. (2017). The RNA m6A Reader YTHDF2 Is Essential for the Post-transcriptional Regulation of the Maternal Transcriptome and Oocyte Competence. *Molecular Cell*, *67*(6), 1059-1067.e4. https://doi.org/10.1016/j.molcel.2017.08.003

Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., … Zaret, K. S. (2016). The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Molecular Cell*, *62*(1), 79–91. https://doi.org/10.1016/j.molcel.2016.03.001

Jacobs, Y., Schnabel, C. A., & Cleary, M. L. (1999). Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. *Molecular and Cellular Biology*, *19*(7), 5134–5142. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10373562%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC84356

Jevtić, P., & Levy, D. L. (2017). Both Nuclear Size and DNA Amount Contribute to Midblastula Transition Timing in Xenopus laevis. *Scientific Reports*, *7*(1), 7908.

https://doi.org/10.1038/s41598-017-08243-z

Joseph, S. R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., … Vastenhouw, N. L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *ELife*, *6*, 1–31. https://doi.org/10.7554/eLife.23326

Jozwik, K. M., Chernukhin, I., Serandour, A. A., Nagarajan, S., & Carroll, J. S. (2016). FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. *Cell Reports*, *17*(10), 2715–2723. https://doi.org/10.1016/j.celrep.2016.11.028

Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R. G., … Burley, S. K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell*, *106*(1), 71–81. https://doi.org/10.1016/s0092-8674(01)00417-2

Katsuno, M., Adachi, H., Minamiyama, M., Waza, M., Doi, H., Kondo, N., … Sobue, G. (2010). Disrupted transforming growth factor-beta signaling in spinal and bulbar muscular atrophy. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(16), 5702–5712. https://doi.org/10.1523/JNEUROSCI.0388-10.2010

Kilstrup-Nielsen, C., Alessio, M., & Zappavigna, V. (2003). PBX1 nuclear export is regulated independently of PBX-MEINOX interaction by PKA phosphorylation of the PBC-B domain. *The EMBO Journal*, *22*(1), 89–99. https://doi.org/10.1093/emboj/cdg010

Kim, S. K., Selleri, L., Lee, J. S., Zhang, A. Y., Gu, X., Jacobs, Y., & Cleary, M. L. (2002). Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus. *Nature Genetics*, *30*(4), 430–435. https://doi.org/10.1038/ng860

Knoepfler, P. S., Bergstrom, D. A., Uetsuki, T., Dac-Korytko, I., Sun, Y. H., Wright, W. E., … Kamps, M. P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. *Nucleic Acids Research*, *27*(18), 3752–3761. https://doi.org/10.1093/nar/27.18.3752

Knoepfler, P. S., & Kamps, M. P. (1997). The highest affinity DNA element bound by Pbx complexes in t(1;19) leukemic cells fails to mediate cooperative DNA-binding or cooperative transactivation by E2a-Pbx1 and class I Hox proteins - evidence for selective targetting of E2a-Pbx1 to a subset of P. *Oncogene*, *14*(21), 2521–2531. https://doi.org/10.1038/sj.onc.1201097

Kobayashi, M., Fujioka, M., Tolkunova, E. N., Deka, D., Abu-Shaar, M., Mann, R. S., & Jaynes, J. B. (2003). Engrailed cooperates with extradenticle and homothorax to repress target genes in Drosophila. *Development (Cambridge, England)*, *130*(4), 741–751. https://doi.org/10.1242/dev.00289

Kok, F. O., Shin, M., Ni, C.-W., Gupta, A., Grosse, A. S., van Impel, A., … Lawson, N. D. (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Developmental Cell*, *32*(1), 97–108. https://doi.org/10.1016/j.devcel.2014.11.018

Kopito, R. B., & Elbaum, M. (2009). Nucleocytoplasmic transport: a thermodynamic mechanism. *HFSP Journal*, *3*(2), 130–141. https://doi.org/10.2976/1.3080807

Kornberg, R. D. (1977). Structure of chromatin. *Annual Review of Biochemistry*, *46*(4), 931–954. https://doi.org/10.1146/annurev.bi.46.070177.004435

Korsmeyer, S. J. (1992). Chromosomal translocations in lymphoid malignancies reveal novel proto-oncogenes. *Annual Review of Immunology*, *10*, 785–807. https://doi.org/10.1146/annurev.iy.10.040192.004033

Koss, M., Bolze, A., Brendolan, A., Saggese, M., Capellini, T. D., Bojilova, E., … Selleri, L. (2012). Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module. *Developmental Cell*, *22*(5), 913–926. https://doi.org/10.1016/j.devcel.2012.02.009

Kucukural, A., Yukselen, O., Ozata, D. M., Moore, M. J., & Garber, M. (2019). DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genomics*, *20*(1), 6. https://doi.org/10.1186/s12864-018-5362-x

Ladam, F., & Sagerström, C. G. (2014). Hox regulation of transcription: more complex(es). *Developmental Dynamics*, *243*(1), 4–15. https://doi.org/10.1002/dvdy.23997

Ladam, F., Stanney, W., Donaldson, I. J., Yildiz, O., Bobola, N., & Sagerström, C. G. (2018). TALE factors use two distinct functional modes to control an essential zebrafish gene expression program. *ELife*, *7*, 1–33. https://doi.org/10.7554/eLife.36144

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laue, K., Rajshekar, S., Courtney, A. J., Lewis, Z. A., & Goll, M. G. (2019). The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. *Nature Communications*, *10*(1), 1551. https://doi.org/10.1038/s41467-019-09582-3

Laurent, A., Calabrese, M., Warnatz, H.-J., Yaspo, M.-L., Tkachuk, V., Torres, M., … Penkov, D. (2015). ChIP-Seq and RNA-Seq analyses identify components of the Wnt and Fgf signaling pathways as Prep1 target genes in mouse embryonic stem cells. *PloS One*, *10*(4), e0122518. https://doi.org/10.1371/journal.pone.0122518

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., … Krause, H. M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*, *131*(1), 174–187. https://doi.org/10.1016/j.cell.2007.08.003

Lee, M. T., Bonneau, A. R., Takacs, C. M., Bazzini, A. A., DiVito, K. R., Fleming, E. S., & Giraldez, A. J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature*, *503*(7476), 360–364. https://doi.org/10.1038/nature12632

Leichsenring, M., Maes, J., Mössner, R., Driever, W., & Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science (New York, N.Y.)*, *341*(6149), 1005–1009. https://doi.org/10.1126/science.1242527

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323. https://doi.org/10.1186/1471-2105-12-323

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Q., Ritter, D., Yang, N., Dong, Z., Li, H., Chuang, J. H., & Guo, S. (2010). A systematic

approach to identify functional motifs within vertebrate developmental enhancers. *Developmental Biology*, *337*(2), 484–495. https://doi.org/10.1016/j.ydbio.2009.10.019

Li, X. Y., Hooft van Huijsduijnen, R., Mantovani, R., Benoist, C., & Mathis, D. (1992). Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies an activation domain. *The Journal of Biological Chemistry*, *267*(13), 8984–8990. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1577736

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature*, *456*(7220), 400–403. https://doi.org/10.1038/nature07388

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., & Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*, *518*(7540), 560–564. https://doi.org/10.1038/nature14234

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–210. https://doi.org/10.1016/0092-8674(87)90561-7

Losa, M., Risolino, M., Li, B., Hart, J., Quintana, L., Grishina, I., … Selleri, L. (2018). Face morphogenesis is promoted by Pbx-dependent EMT via regulation of Snail1 during frontonasal prominence fusion. *Development (Cambridge, England)*, *145*(5), dev157628. https://doi.org/10.1242/dev.157628

Love, N. R., Thuret, R., Chen, Y., Ishibashi, S., Sabherwal, N., Paredes, R., … Amaya, E. (2011). pTransgenesis: a cross-species, modular transgenesis resource. *Development*, *138*(24), 5451–5458. https://doi.org/10.1242/dev.066498

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., & Zhang, Y. (2016). Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell*, *165*(6), 1375–1388. https://doi.org/10.1016/j.cell.2016.05.050

Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics (Oxford, England)*, *27*(12), 1696–1697. https://doi.org/10.1093/bioinformatics/btr189

Machon, O., Masek, J., Machonova, O., Krauss, S., & Kozmik, Z. (2015). Meis2 is essential for cranial and cardiac neural crest development. *BMC Developmental Biology*, *15*(1), 40. https://doi.org/10.1186/s12861-015-0093-6

Maity, S. N., & de Crombrugghe, B. (1998). Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends in Biochemical Sciences*, *23*(5), 174–178. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9612081

Manley, N. R., Selleri, L., Brendolan, A., Gordon, J., & Cleary, M. L. (2004). Abnormalities of caudal pharyngeal pouch development in Pbx1 knockout mice mimic loss of Hox3 paralogs. *Developmental Biology*, *276*(2), 301–312. https://doi.org/10.1016/j.ydbio.2004.08.030

Mantovani, R., Li, X. Y., Pessara, U., Hooft van Huisjduijnen, R., Benoist, C., & Mathis, D. (1994). Dominant negative analogs of NF-YA. *The Journal of Biological Chemistry*, *269*(32), 20340–20346. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8051128

Mathavan, S., Lee, S. G. P., Mak, A., Miller, L. D., Murthy, K. R. K., Govindarajan, K. R., …

Lufkin, T. (2005). Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS Genetics*, *1*(2), 260–276. https://doi.org/10.1371/journal.pgen.0010029

Mayran, A., & Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. *The Journal of Biological Chemistry*, *293*(36), 13795–13804. https://doi.org/10.1074/jbc.R117.001232

Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., & Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nature Genetics*, *50*(2), 259–269. https://doi.org/10.1038/s41588-017-0035-2

McCulley, D. J., Wienhold, M. D., Hines, E. A., Hacker, T. A., Rogers, A., Pewowaruk, R. J., … Sun, X. (2018). PBX transcription factors drive pulmonary vascular adaptation to birth. *The Journal of Clinical Investigation*, *128*(2), 655–667. https://doi.org/10.1172/JCI93395

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., … Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, *28*(5), 495–501. https://doi.org/10.1038/nbt.1630

McNabb, D S, Xing, Y., & Guarente, L. (1995). Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes & Development*, *9*(1), 47–58. https://doi.org/10.1101/gad.9.1.47

McNabb, David S, & Pinto, I. (2005). Assembly of the Hap2p/Hap3p/Hap4p/Hap5p-DNA complex in Saccharomyces cerevisiae. *Eukaryotic Cell*, *4*(11), 1829–1839. https://doi.org/10.1128/EC.4.11.1829-1839.2005

Merabet, S., & Mann, R. S. (2016). To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends in Genetics*, *32*(6), 334–347. https://doi.org/10.1016/j.tig.2016.03.004

Milech, N., Kees, U. R., & Watt, P. M. (2001). Novel alternative PBX3 isoforms in leukemia cells with distinct interaction specificities. *Genes, Chromosomes & Cancer*, *32*(3), 275–280. https://doi.org/10.1002/gcc.1190

Mishima, Y., & Tomari, Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell*, *61*(6), 874–885. https://doi.org/10.1016/j.molcel.2016.02.027

Monica, K., Galili, N., Nourse, J., Saltman, D., & Cleary, M. L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. *Molecular and Cellular Biology*, *11*(12), 6149–6157. https://doi.org/10.1128/mcb.11.12.6149

Morgan, R., El-Tanani, M., Hunter, K. D., Harrington, K. J., & Pandha, H. S. (2017). Targeting HOX/PBX dimers in cancer. *Oncotarget*, *8*(19), 32322–32331. https://doi.org/10.18632/oncotarget.15971

Moskow, J. J., Bullrich, F., Huebner, K., Daar, I. O., & Buchberg, A. M. (1995). Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. *Molecular and Cellular Biology*, *15*(10), 5434–5443. https://doi.org/10.1128/mcb.15.10.5434

Nakamura, T., Jenkins, N. A., & Copeland, N. G. (1996). Identification of a new family of Pbx-related homeobox genes. *Oncogene*, *13*(10), 2235–2242. Retrieved from https://www.researchgate.net/publication/14256434_Identification_of_a_new_family_of_Pbx

-related_homeobox_genes

Nakamura, T., Largaespada, D. A., Shaughnessy, J. D., Jenkins, N. A., & Copeland, N. G. (1996). Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nature Genetics*, *12*(2), 149–153. https://doi.org/10.1038/ng0296-149

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., … Mantovani, R. (2013). Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell*, *152*(1–2), 132–143. https://doi.org/10.1016/j.cell.2012.11.047

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., … Müller, F. (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Research*, *23*(11), 1938–1950. https://doi.org/10.1101/gr.153692.112

Nien, C.-Y., Liang, H.-L., Butcher, S., Sun, Y., Fu, S., Gocha, T., … Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genetics*, *7*(10), e1002339. https://doi.org/10.1371/journal.pgen.1002339

Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S. D., & Cleary, M. L. (1990). Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor. *Cell*, *60*(4), 535–545. https://doi.org/10.1016/0092-8674(90)90657-z

Oldfield, A. J., Yang, P., Conway, A. E., Cinghu, S., Freudenberg, J. M., Yellaboina, S., & Jothi, R. (2014). Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Molecular Cell*, *55*(5), 708–722. https://doi.org/10.1016/j.molcel.2014.07.005

Pai, C. Y., Kuo, T. S., Jaw, T. J., Kurant, E., Chen, C. T., Bessarab, D. A., … Sun, Y. H. (1998). The Homothorax homeoprotein activates the nuclear localization of another homeoprotein, extradenticle, and suppresses eye development in Drosophila. *Genes & Development*, *12*(3), 435–446. https://doi.org/10.1101/gad.12.3.435

Pálfy, M., Joseph, S. R., & Vastenhouw, N. L. (2017). The timing of zygotic genome activation. *Current Opinion in Genetics & Development*, *43*, 53–60. https://doi.org/10.1016/j.gde.2016.12.001

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., & Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature*, *397*(6721), 714–719. https://doi.org/10.1038/17833

Peers, B., Sharma, S., Johnson, T., Kamps, M., & Montminy, M. (1995). The pancreatic islet factor STF-1 binds cooperatively with Pbx to a regulatory element in the somatostatin promoter: importance of the FPWMK motif and of the homeodomain. *Molecular and Cellular Biology*, *15*(12), 7091–7097. https://doi.org/10.1128/mcb.15.12.7091

Penkov, D., Mateos San Martín, D., Fernandez-Díaz, L. C., Rosselló, C. A., Torroja, C., Sánchez-Cabo, F., … Torres, M. (2013). Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Reports*, *3*(4), 1321–1333. https://doi.org/10.1016/j.celrep.2013.03.029

Pinkham, J. L., Olesen, J. T., & Guarente, L. P. (1987). Sequence and nuclear localization of the Saccharomyces cerevisiae HAP2 protein, a transcriptional activator. *Molecular and Cellular*

Biology, *7*(2), 578–585. https://doi.org/10.1128/mcb.7.2.578

Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., & Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell*, *96*(4), 587–597. https://doi.org/10.1016/s0092-8674(00)80662-5

Poot, R. A., Dellaire, G., Hülsmann, B. B., Grimaldi, M. A., Corona, D. F., Becker, P. B., … Varga-Weisz, P. D. (2000). HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *The EMBO Journal*, *19*(13), 3377–3387. https://doi.org/10.1093/emboj/19.13.3377

Pöpperl, Heike, Rikhof, H., Chang, H., Haffter, P., Kimmel, C. B., & Moens, C. B. (2000). lazarus is a novel pbx gene that globally mediates hox gene function in zebrafish. *Molecular Cell*, *6*(2), 255–267. https://doi.org/10.1016/S1097-2765(00)00027-7

Pöpperl, Helke, Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., … Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell*, *81*(7), 1031–1042. https://doi.org/10.1016/s0092-8674(05)80008-x

Prince, V. E., Joly, L., Ekker, M., & Ho, R. K. (1998). Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. *Development (Cambridge, England)*, *125*(3), 407–420. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9425136

Prince, V. E., Moens, C. B., Kimmel, C. B., & Ho, R. K. (1998). Zebrafish hox genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, valentino. *Development (Cambridge, England)*, *125*(3), 393–406. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9425135

Prioleau, M. N., Huet, J., Sentenac, A., & Méchali, M. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. *Cell*, *77*(3), 439–449. https://doi.org/10.1016/0092-8674(94)90158-9

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Qureshi, I. A., Gokhan, S., & Mehler, M. F. (2010). REST and CoREST are transcriptional and epigenetic regulators of seminal neural fate decisions. *Cell Cycle (Georgetown, Tex.)*, *9*(22), 4477–4486. https://doi.org/10.4161/cc.9.22.13973

Rabani, M., Pieper, L., Chew, G. L., & Schier, A. F. (2017). A Massively Parallel Reporter Assay of 3′ UTR Sequences Identifies In Vivo Rules for mRNA Degradation. *Molecular Cell*, *68*(6), 1083-1094.e5. https://doi.org/10.1016/j.molcel.2017.11.014

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(Web Server issue), W187-91. https://doi.org/10.1093/nar/gku365

Ramos, S. B. V., Stumpo, D. J., Kennington, E. A., Phillips, R. S., Bock, C. B., Ribeiro-Neto, F., & Blackshear, P. J. (2004). The CCCH tandem zinc-finger protein Zfp36l2 is crucial for female fertility and early embryonic development. *Development (Cambridge, England)*, *131*(19), 4883–4893. https://doi.org/10.1242/dev.01336

Rhee, J. W., Arata, A., Selleri, L., Jacobs, Y., Arata, S., Onimaru, H., & Cleary, M. L. (2004). Pbx3 deficiency results in central hypoventilation. *The American Journal of Pathology*, *165*(4), 1343–1350. https://doi.org/10.1016/S0002-9440(10)63392-5

Rieckhof, G. E., Casares, F., Ryoo, H. D., Abu-Shaar, M., & Mann, R. S. (1997). Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. *Cell*, *91*(2), 171–183. https://doi.org/10.1016/s0092-8674(00)80400-6

Romier, C., Cocchiarella, F., Mantovani, R., & Moras, D. (2003). The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. *The Journal of Biological Chemistry*, *278*(2), 1336–1345. https://doi.org/10.1074/jbc.M209635200

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., & Stainier, D. Y. R. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*, *524*(7564), 230–233. https://doi.org/10.1038/nature14580

Ryoo, H. D., & Mann, R. S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes & Development*, *13*(13), 1704–1716. https://doi.org/10.1101/gad.13.13.1704

Saadaoui, M., Merabet, S., Litim-Mecheri, I., Arbeille, E., Sambrani, N., Damen, W., … Graba, Y. (2011). Selection of distinct Hox-Extradenticle interaction modes fine-tunes Hox protein activity. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2276–2281. https://doi.org/10.1073/pnas.1006964108

Saleh, M., Rambaldi, I., Yang, X. J., & Featherstone, M. S. (2000). Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. *Molecular and Cellular Biology*, *20*(22), 8623–8633. https://doi.org/10.1128/MCB.20.22.8623-8633.2000

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., … Fraser, P. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics*, *42*(1), 53–61. https://doi.org/10.1038/ng.496

Selleri, L, Depew, M. J., Jacobs, Y., Chanda, S. K., Tsang, K. Y., Cheah, K. S., … Cleary, M. L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development (Cambridge, England)*, *128*(18), 3543–3557. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11566859

Selleri, Licia, DiMartino, J., van Deursen, J., Brendolan, A., Sanyal, M., Boon, E., … Cleary, M. L. (2004). The TALE homeodomain protein Pbx2 is not essential for development and long-term survival. *Molecular and Cellular Biology*, *24*(12), 5324–5331. https://doi.org/10.1128/MCB.24.12.5324-5331.2004

Selleri, Licia, Zappavigna, V., & Ferretti, E. (2019). "Building a perfect body": control of vertebrate organogenesis by PBX-dependent regulatory networks. *Genes & Development*, *33*(5–6), 258–275. https://doi.org/10.1101/gad.318774.118

Serrano, N., & Maschat, F. (1998). Molecular mechanism of polyhomeotic activation by Engrailed. *The EMBO Journal*, *17*(13), 3704–3713. https://doi.org/10.1093/emboj/17.13.3704

Shen, W. F., Rozenfeld, S., Lawrence, H. J., & Largman, C. (1997). The Abd-B-like Hox

homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. *The Journal of Biological Chemistry*, *272*(13), 8198–8206. https://doi.org/10.1074/jbc.272.13.8198

Shindo, Y., & Amodeo, A. A. (2019). Dynamics of Free and Chromatin-Bound Histone H3 during Early Embryogenesis. *Current Biology : CB*, *29*(2), 359-366.e4. https://doi.org/10.1016/j.cub.2018.12.020

Soufi, A., Donahue, G., & Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, *151*(5), 994–1004. https://doi.org/10.1016/j.cell.2012.09.045

Sreekumar, V., & Norris, D. P. (2019). Cilia and development. *Current Opinion in Genetics & Development*, *56*(Figure 1), 15–21. https://doi.org/10.1016/j.gde.2019.05.002

Stankunas, K., Shang, C., Twu, K. Y., Kao, S.-C., Jenkins, N. A., Copeland, N. G., … Chang, C.-P. (2008). Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circulation Research*, *103*(7), 702–709. https://doi.org/10.1161/CIRCRESAHA.108.175489

Stapel, L. C., Zechner, C., & Vastenhouw, N. L. (2017). Uniform gene expression in embryos is achieved by temporal averaging of transcription noise. *Genes and Development*, *31*(16), 1635–1640. https://doi.org/10.1101/gad.302935.117

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M., & Levine, M. (2002). Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo. *Cell*, *111*(5), 687–701. https://doi.org/10.1016/s0092-8674(02)01087-5

Stoeckius, M., Grün, D., Kirchner, M., Ayoub, S., Torti, F., Piano, F., … Rajewsky, N. (2014). Global characterization of the oocyte-to-embryo transition in Caenorhabditis elegans uncovers a novel mRNA clearance mechanism. *The EMBO Journal*, *33*(16), 1751–1766. https://doi.org/10.15252/embj.201488769

Sun, J., Yan, L., Shen, W., & Meng, A. (2018). Maternal Ybx1 safeguards zebrafish oocyte maturation and maternal-to-zygotic transition by repressing global translation. *Development (Cambridge, England)*, *145*(19), dev166587. https://doi.org/10.1242/dev.166587

Tadros, W., Goldman, A. L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., … Lipshitz, H. D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. *Developmental Cell*, *12*(1), 143–155. https://doi.org/10.1016/j.devcel.2006.10.005

Tropberger, P., & Schneider, R. (2013). Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nature Structural & Molecular Biology*, *20*(6), 657–661. https://doi.org/10.1038/nsmb.2581

Tümpel, S., Cambronero, F., Ferretti, E., Blasi, F., Wiedemann, L. M., & Krumlauf, R. (2007). Expression of Hoxa2 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. *Developmental Biology*, *302*(2), 646–660. https://doi.org/10.1016/j.ydbio.2006.10.029

Tyner, C., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Eisenhart, C., … Kent, W. J. (2017). The UCSC Genome Browser database: 2017 update. *Nucleic Acids Research*, *45*(D1), D626–D634. https://doi.org/10.1093/nar/gkw1134

Van Auken, K., Weaver, D., Robertson, B., Sundaram, M., Saldi, T., Edgar, L., … Wood, W. B. (2002). Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in C. elegans embryogenesis. *Development (Cambridge, England)*, *129*(22), 5255–5268. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12399316

Vastenhouw, N. L., Cao, W. X., & Lipshitz, H. D. (2019). The maternal-to-zygotic transition revisited. *Development (Cambridge, England)*, *146*(11), dev161471. https://doi.org/10.1242/dev.161471

Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., … Schier, A. F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*, *464*(7290), 922–926. https://doi.org/10.1038/nature08866

Villaescusa, J. C., Li, B., Toledo, E. M., Rivetti di Val Cervo, P., Yang, S., Stott, S. R., … Arenas, E. (2016). A PBX1 transcriptional network controls dopaminergic neuron development and is impaired in Parkinson's disease. *The EMBO Journal*, *35*(18), 1963–1978. https://doi.org/10.15252/embj.201593725

Vitobello, A., Ferretti, E., Lampe, X., Vilain, N., Ducret, S., Ori, M., … Rijli, F. M. (2011). Hox and Pbx factors control retinoic acid synthesis during hindbrain segmentation. *Developmental Cell*, *20*(4), 469–482. https://doi.org/10.1016/j.devcel.2011.03.011

Vlachakis, N, Choe, S. K., & Sagerström, C. G. (2001). Meis3 synergizes with Pbx4 and Hoxb1b in promoting hindbrain fates in the zebrafish. *Development (Cambridge, England)*, *128*(8), 1299–1312. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11262231

Vlachakis, Nikolaos, Ellstrom, D. R., & Sagerström, C. G. (2000). A novel pbx family member expressed during early zebrafish embryogenesis forms trimeric complexes with meis3 and hoxb1b. *Developmental Dynamics*, *217*(1), 109–119. https://doi.org/10.1002/(SICI)1097-0177(200001)217:1<109::AID-DVDY10>3.0.CO;2-8

Voeltz, G. K., & Steitz, J. A. (1998). AUUUA sequences direct mRNA deadenylation uncoupled from decay during Xenopus early development. *Molecular and Cellular Biology*, *18*(12), 7537–7545. https://doi.org/10.1128/mcb.18.12.7537

Wagner, K., Mincheva, A., Korn, B., Lichter, P., & Pöpperl, H. (2001). Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mechanisms of Development*, *103*(1–2), 127–131. https://doi.org/10.1016/S0925-4773(01)00349-5

Wang, Q. T., Piotrowska, K., Ciemerych, M. A., Milenkovic, L., Scott, M. P., Davis, R. W., & Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. *Developmental Cell*, *6*(1), 133–144. https://doi.org/10.1016/S1534-5807(03)00404-0

Wang, Y., Stary, J. M., Wilhelm, J. E., & Newmark, P. A. (2010). A functional genomic screen in planarians identifies novel regulators of germ cell development. *Genes & Development*, *24*(18), 2081–2092. https://doi.org/10.1101/gad.1951010

Waskiewicz, A. J., Rikhof, H. a, & Moens, C. B. (2002). Eliminating zebrafish pbx proteins reveals a hindbrain ground state. *Developmental Cell*, *3*(5), 723–733. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12431378

Waskiewicz, J., Rikhof, H., Hernandez, R. E., & Moens, C. B. (2001). Zebrafish Meis functions to stabilize Pbx proteins and regulate hindbrain patterning. *Development*, *128*(21), 4139–4151.

Welsh, I. C., Hart, J., Brown, J. M., Hansen, K., Rocha Marques, M., Aho, R. J., … Selleri, L. (2018). Pbx loss in cranial neural crest, unlike in epithelium, results in cleft palate only and a broader midface. *Journal of Anatomy*, *233*(2), 222–242. https://doi.org/10.1111/joa.12821

Wolfrum, U., & Schmitt, A. (2000). Rhodopsin transport in the membrane of the connecting cilium of mammalian photoreceptor cells. *Cell Motility and the Cytoskeleton*, *46*(2), 95–107. https://doi.org/10.1002/1097-0169(200006)46:2<95::AID-CM2>3.0.CO;2-Q

Yamanaka, T., Miyazaki, H., Oyama, F., Kurosawa, M., Washizu, C., Doi, H., & Nukina, N. (2008). Mutant Huntingtin reduces HSP70 expression through the sequestration of NF-Y transcription factor. *The EMBO Journal*, *27*(6), 827–839. https://doi.org/10.1038/emboj.2008.23

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., … Tang, F. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Structural & Molecular Biology*, *20*(9), 1131–1139. https://doi.org/10.1038/nsmb.2660

Yoshikawa, T., Piao, Y., Zhong, J., Matoba, R., Carter, M. G., Wang, Y., … Ko, M. S. H. (2006). High-throughput screen for genes predominantly expressed in the ICM of mouse blastocysts by whole mount in situ hybridization. *Gene Expression Patterns : GEP*, *6*(2), 213–224. https://doi.org/10.1016/j.modgep.2005.06.003

Yoshioka, Y., Suyari, O., Yamada, M., Ohno, K., Hayashi, Y., & Yamaguchi, M. (2007). Complex interference in the eye developmental pathway by Drosophila NF-YA. *Genesis (New York, N.Y. : 2000)*, *45*(1), 21–31. https://doi.org/10.1002/dvg.20260

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, *25*, 2227–2241. https://doi.org/10.1101/gad.176826.111

Zhang, B., Wu, X., Zhang, W., Shen, W., Sun, Q., Liu, K., … Xie, W. (2018). Widespread Enhancer Dememorization and Promoter Priming during Parental-to-Zygotic Transition. *Molecular Cell*, *72*(4), 673-686.e6. https://doi.org/10.1016/j.molcel.2018.10.017

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhang, Y., Vastenhouw, N. L., Feng, J., Fu, K., Wang, C., Ge, Y., … Liu, X. S. (2014). Canonical nucleosome organization at promoters forms during genome activation. *Genome Research*, *24*(2), 260–266. https://doi.org/10.1101/gr.157750.113

Zhao, B. S., Wang, X., Beadell, A. V., Lu, Z., Shi, H., Kuuspalu, A., … He, C. (2017). m6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature*, *542*(7642), 475–478. https://doi.org/10.1038/nature21355