TRANSCRIPTIONAL ACTIVATION OF EUKARYOTIC PROMOTERS VIA ZINC FINGER PROTEIN 143

A Dissertation

by

LAURA HUNING

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mary Bryk
Committee Members,	Gary Kunkel
	Arne Lekven
	Hays Rye
Head of Department,	Joshua Wand

August 2020

Major Subject: Biochemistry

ABSTRACT

Zinc Finger Protein 143 (ZNF143) is a ubiquitous C2H2 zinc-finger transcriptional activator in vertebrates. Predicted to regulate over 2000 promoters, ZNF143 activates transcription at both protein coding genes and small RNA genes transcribed by either RNA polymerase II or RNA polymerase III. Recently, an auxiliary function for ZNF143 as an architectural factor localizing at chromatin interactions has emerged. Promoters targeted for activation by ZNF143 are functionally diverse, involving many cellular processes such as cell cycle regulation, cancer, and development. The zebrafish genome contains a previously uncharacterized paralog of *znf143b*, *znf143a,* that contributes to the overall function as an activator in zebrafish. To identify the contribution of activator function of ZNF143a to the overall activation by ZNF143, this dissertation involves an expression analysis and targeted knockdown of each paralog in zebrafish. This dissertation demonstrates ZNF143a as a highly similar activator protein, however differentially expressed in the early development of zebrafish. Furthermore, the use of paralog specific knockdowns of each paralog presents developmental defects in the brain of 24hour embryos. Additionally, the molecular mechanism by which ZNF143 activates small RNA and mRNA encoding target genes is unknown. To investigate the mechanism of ZNF143, transgenic zebrafish lines were constructed to determine any potential architectural features that distinguish promoters regulated by ZNF143. Preliminary results suggest that positive histone modifications are present at promoters containing SPH sites and reduced at promoter regions containing a mutated SPH site.

DEDICATION

To my family and all the wonderful friends that I've made along the way.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Gary Kunkel, for being an exemplary mentor and friend. I would also like to thank Arne Lekven, Mary Bryk, and Hays Rye for their guidance and support as committee members. Lastly, I would like to thank Nicola Ayres for being a great teaching mentor and colleague.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This dissertation was supervised by a committee composed of Dr. Kunkel, Dr. Bryk, and Dr. Rye of the Department of Biochemistry and Biophysics at Texas A&M University, and Dr. Lekven of the Department of Biology and Biochemistry of the University of Houston.

Cell work, transfection assays, preliminary αZNF143 and αCHD8 ChIP-seq, and preliminary ChIP-PCR binding in Ch.II and Ch.III were performed by Dr. Gary Kunkel of the Department of Biochemistry and Biophysics. Additional work for this thesis was completed by the student independently.

Funding sources

Preparation of some of the materials and early experiments in this study were initiated with funding from American Heart Association, Southwest Affiliate (14GRNT20460146). Most of the work was funded by the Department of Biochemistry and Biophysics, Texas A&M University.

NOMENCLATURE

ACH3K9	ACETYLATED HISTONE H3 ON LYSINE 9
AD	ACTIVATING DOMAIN
ARE	ADENYLATE-URIDYLATE-RICH ELEMENT
ATIC	5-AMINOIMIDAZOLE-4-CARBOXAMIDE RIBONUCLEOTIDE FORMYLTRANSFERASE/IMP CYCLOHYDROLASE
C/EBPA	CCAAT-ENHANCER-BINDING PROTEIN A
CHD8	CHROMODOMAIN HELICASE DNA-BINDING PROTEIN 8
CHIP	CHROMATIN IMMUNOPRECIPITATION
CHIA-PET	CHROMATIN INTERACTION ANALYSIS BY PAIRED-END TAG SEQUENCING
CRISPRI	CRISPR INTERFERENCE
CTCF	CCCTC-BINDING FACTOR
C2H2	CYS2-HIS2
C-TERMINAL	CARBOXY-TERMINAL
DBD	DNA BINDING DOMAIN
DCAS9	DEACTIVATED CAS9
DIG	DIGOXIGENIN
DMR	DIFFERENTIALLY METHYLATED REGION
DSE	DISTAL SEQUENCE ELEMENT
EGFP	ENHANCED GREEN FLUORESCENT PROTEIN
F0	FOUNDER GENERATION
F1	FIRST FAMILIAL GENERATION
GABP	GA-BINDING PROTEIN

GSK3AB	GLYCOGEN SYNTHASE KINASE 3 ALPHA B
G1	GAP 1 PHASE
G2	GAP 2 PHASE
GRNA	GUIDE RNA
HEK293	HUMAN EMBRYONIC KIDNEY CELL LINE
HPF	HOURS POST FERTILIZATION
H3K4ME3	HSTONE H3 TRIMETHYLATION ON LYSINE 4
H3K9ME3	HISTONE H3 TRIMETHYLATION ON LYSINE 9
H4AC	ACETYLATION ON HISTONE H4
KB	KILOBASE
M	MITOSIS
MDIG	MINERAL DIST-INDUCED GENE
MRNA	MESSENGER RNA
OCT	OCTAMER
ORF	OPEN READING FRAME
PSE	PROXIMAL SEQUENCE ELEMENT
QRT-PCR	QUANTITATIVE REVERSE TRANSCRIPTASE POLYMERASE CHAIN REACTION
RNAP	RNA POLYMERASE
SBF	SPH BINDING FACTOR
SBS	STAF BINDING SITE
SEQ	SEQUENCING
SGRNA	SINGLE GUIDE RNA

SNRNA	SMALL NUCLEAR RNA
SPH	SPH1 POSTOCTAMER HOMOLOGY
STAF	SELENOCYSTEINE TRANSCRIPTION ACTIVATING FACTOR
TSS	TRANSCRIPTION START SITE
UTR	UNTRANSLATED REGION
U4	U4 SMALL NUCLEAR RNA
ZNF143	ZINC FINGER PROTEIN 143

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

CHAPTER I : INTRODUCTION

Regulation of eukaryotic gene expression is a complex interplay between coordinated actions of transcription factors and higher order structure composing the genome. A highly coordinated and controlled regulation of gene expression by transcription factors and their cofactors is essential to the growth and development of an organism. Although a broad understanding of the roles of transcription factors in the regulation of expression of their gene targets have been established, the mechanisms regarding many specific transcriptional regulators are still largely unknown. Among these, is a highly pervasive transcriptional activator protein, called Zinc Finger Protein 143 (ZNF143). A unique transcriptional activator protein, ZNF143 activates genes transcribed by RNAPII and RNAPIII.

Regulation of gene expression is essential in the maintenance of cellular integrity. The process by which gene expression is controlled can occur at many different stages from a template DNA all the way through the modification and transport of a mature protein. This regulation begins at transcription: the process by which DNA is copied into an RNA transcript by RNA polymerase (RNAP). While bacteria and archaea contain one RNAP, eukaryotes use three distinct polymerases for different classes of RNA[1].

The area of DNA surrounding the transcription start site (TSS), generally 50 nucleotides upstream or downstream, is the core promoter[2]. These regions support the assembly of a pre-initiation complex consisting of the RNA polymerase and its

associated general transcription factors (GTFs) [2]. Core promoters are able to direct transcription, albeit generally at a low level[2]. One limiting step in the expression of a gene is the recruitment and assembly of a pre-initiation complex at the core promoter. At RNAPII transcribed genes, current data suggest a sequential model of pre-initiation complex assembly where core-promoter elements are recognized and bound in a particular order. This sequence involves the binding of general transcription factors (GTFs) in the following sequence: 1. TFIID, 2. TFIIA and TFIIB binding, 3. RNAPII-TFIIF complex, then lastly, TFIIE and TFIIH[3] (Fig. 1). These are required for the formation of a pre-initiation complex. Following the formation of a pre-initiation complex, the DNA duplex is melted, and the pre-initiation complex assumes an open complex formation where synthesis is able to begin of the first few nucleotides of the nascent transcript[3]. After synthesis begins, RNAPII is released from the promoter, and ultimately enters into a productive elongation[3].

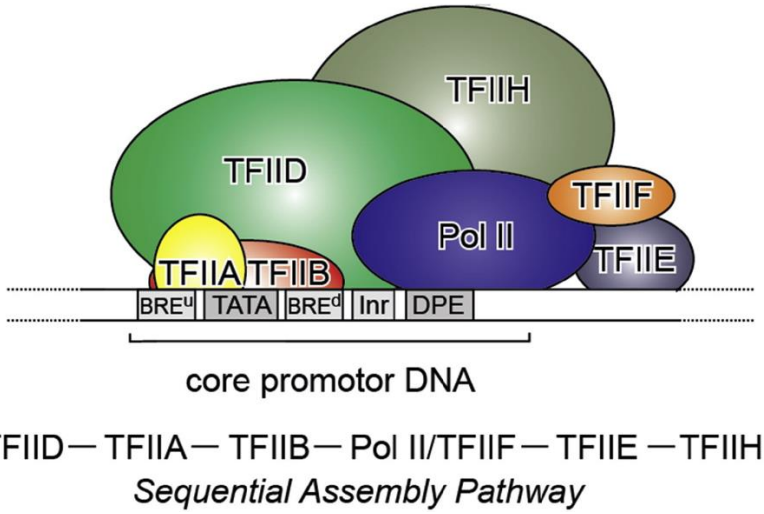


Figure 1. Preinitiation complex formation for transcription at mRNA encoding genes. The pre-initiation complex (PIC) bound to the core promoter is shown (Reprinted from Ref [4]). The PIC contains the GTFs TFIIA, B, D, E, F, and H, and RNA Pol II bound to core promoter elements located in DNA. The assembly of the PIC is highly regulated and occurs in a sequential fashion starting with TFIID. TATA, TATA-box DNA; BREu, B recognition element upstream; BREd, B recognition element downstream; Inr, Initiator; DPE, Down-stream promoter element.

RNAPIII is specific for transcribing short and abundant non-coding transcripts. These include transfer RNAs (tRNAs), small rRNAs, small nucleolar RNAs, the RNA subunit of RNase P, and RNA elements of the signal recognition particle[1]. As opposed to RNAPII, RNAPIII requires two GTFs: TFIIIB, and TFIIIC[1]. A third transcription factor, TFIIIA, is required only for 5S RNA production[1]. For 5S RNA synthesis, TFIIIA binds to the DNA first, followed by TFIIIC and TFIIIB[1]. For the RNA synthesis of tRNA genes, TFIIIC binds to the promoter elements and recruits TFIIIB upstream of the TSS[1].

There are two other groups of proteins that are not required for basal transcription but can influence the amount of transcription that occurs. These groups are

activator proteins and co-activator proteins. Transcriptional activator proteins recognize consensus elements located within DNA and are composed of both a DNA binding domain and an activation domain. A classic hypothesis of activator protein function is that they have the ability to direct GTFs, including RNAP, to facilitate PIC assembly through an activation domain[5]. Alternatively, it has been suggested that targets of activation domains may include the mediator complex to activate transcription[6]. Mediator is considered to act as a bridge between activator proteins and RNAPII[6]. More recently activator proteins have been linked to recruitment of chromatin remodeling and histone modifying activities[7]. However, activator proteins remain poorly understood. There are still many unresolved questions regarding the mechanism of how activator proteins regulate gene expression, for example the molecular targets of activation domains, the specificity of activation domains to their target, and the extent of essentiality of activator proteins all remain poorly defined.

Secondly, co-activator proteins can be required for activator directed transcription, but are not required for basal transcription. Co-activator proteins are becoming increasingly important in the activation of specific genes, but how they facilitate transcription alongside activator proteins is not well understood. Co-activator proteins, unlike activators, do not bind a specific consensus sequence in the DNA. It has been suggested that they might act through three broad mechanisms including: (1) modifying histones to allow proteins access to DNA, (2) bridging transcription factors to other transcription factors or transcriptional machinery, and (3) chromatin remodeling

through ATPase activity[8]. However, specific mechanisms of both activator and co-activator proteins are not well understood.

In the work described in this dissertation, zebrafish are utilized as an animal model to study *znf143* gene function *in vivo*. Advantages to using zebrafish include their small size, large numbers of eggs, transparency of embryos, external fertilization, and genome similarity to humans[9]. Pre-established genetic methods in combination with the ease of embryo imaging during development, made possible by their transparency, allow for an easy determination of tissue-specific expression patterns through *in-situ* hybridizations, fluorescence expression for transgenesis, and developmental phenotypes resulting from knockdown or knockout studies.

Discovered in the *U1* small nuclear RNA (snRNA) gene of chicken embryos, ZNF143 was initially identified as a binding factor that recognizes a sequence element in the enhancer-like region, referred to as an SPH site, with the protein known as SPH-binding factor (SBF) [10]. Shortly thereafter, a *Xenopus* ortholog of ZNF143 was found to be essential for expression of the selenocysteine tRNA ($tRNA^{Sec}$) gene, and called selenocysteine tRNA gene transcription activating factor (Staf) [11]. Currently, ZNF143 is regarded as a pervasive regulator protein conserved throughout metazoans. Soon after discovery, it was found to be widespread in gene transcription as it is able to activate most snRNA and snRNA-type genes transcribed by either RNA polymerase II (RNAPII) or RNA polymerase III (RNAPIII), and a model protein-coding gene transcribed by RNAPII [12]. Evidence for global regulation by ZNF143 grew with the identification of the first bona fide mRNA promoter regulated by ZNF143, the chaperonin t-complex

polypeptide 1 (*CCT*) gene [13]. Now it is apparent that over 2000 mammalian promoters are regulated by ZNF143 [14]. Many cell-cycle regulators are target genes of ZNF143, as cell cycle regulation is becoming increasingly important in cancer.

An underappreciated characteristic of ZNF143 is that it contains separable activation domains with distinct specificities for either snRNA promoters (transcribed by either RNAPII or RNAPIII) or protein-coding gene promoters (RNAPII)[15]. Further understanding of biochemical mechanisms that direct transcriptional stimulation at these specific promoter types will be important for understanding broad mechanisms of transcriptional activation.

Architectural proteins play critical roles in organizing the genome and regulating gene expression. In recent years, the advent of 3C derived assays has uncovered an additional architectural role for the canonical activator protein ZNF143. Major evidence for involvement in genome organization by ZNF143 includes both its localization at topological domains along with other prominent architectural protein factors[16], and mediating chromatin interactions through being a positive driver of chromatin loops[17, 18]. With another mechanistic component of how ZNF143 might regulate gene expression, several new questions arise. How does the newly uncovered architectural role of ZNF143 converge with its regulatory role as a transcriptional activator? To what extent is ZNF143 involved in the establishment of topological domains? Determining whether ZNF143 plays an architectural role will remain challenging so long as the mechanism of activation by ZNF143 is elusive. This introduction reviews evidence for potential mechanisms of activation by ZNF143 at snRNA and protein coding genes, the

role of ZNF143 in chromatin looping, and the overall importance of ZNF143 in regulating eukaryotic gene expression.

ZNF143 binds to a degenerate primary consensus sequence of 18 base pairs in length in proximal promoter regions [19], referred to as a SPH motif or Staf-binding site (SBS) (Fig.2). These sites are generally located within 200bp of the transcription start sites of gene targets [20]. The core of this SPH element is a conserved CCCA sequence (Fig. 2); binding is completely abrogated with mutations located within and surrounding this core sequence [21]. Substantial variability exists in the 5' and 3' ends of the SPH consensus, indicating a degree of flexibility of ZNF143 binding to its sequence element. Additionally, a secondary ACTACA accessory sequence is often found upstream and in conjunction with primary SPH sites [14]. The inclusion of an accessory ACTACA motif in some SPH elements may have a potential negative regulation by the competing protein THAP11 on ZNF143 target promoters ([14]; discussed later). ZNF143 target promoters including the accessory element may have a distinct context specific regulation when compared to target promoters lacking this element.

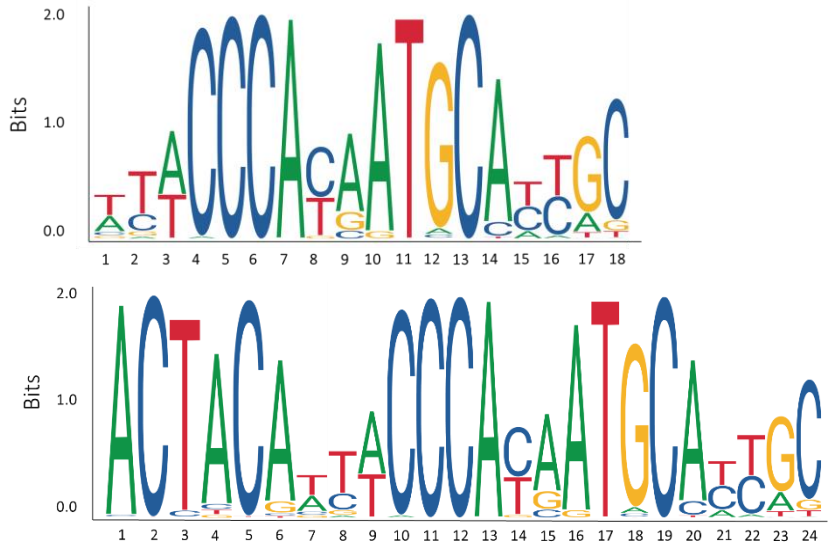


Figure 2. ZNF143 binding site consensus sequence. The consensus DNA binding site for human ZNF143 can exist as a primary SPH motif (top) that consists of 18bp, or an extended SPH motif (bottom) consisting of 24 bp with the 6bp ACTACA sequence attached to the 5' end. Consensus binding site was obtained from Jaspar and expanded further using binding data from Transfac database.

The human *znf143* gene encodes for a 638 amino acid transcriptional activator protein consisting of a well-characterized tandem zinc finger DNA binding domain and two separable activation domains [15] (Fig. 3). The DNA-binding domain consists of seven C2-H2 type zinc fingers located in the center of the protein from amino acids 236 to 445 [15]. However, reflecting the variability of binding site sequence, not every target promoter requires all 7 zinc fingers for high-affinity ZNF143 binding [22]. The human *U6* snRNA gene binding site only requires zinc fingers 2-7, with the first zinc finger being flexible [22]. Mutations performed within the first zinc finger reduced ZNF143 binding and transcription of the *tRNA*^{Sec} promoter [22], indicating that zinc finger binding requirements may be promoter dependent. The DNA binding domain of

ZNF143 makes extensive contacts with the major groove of DNA, and likely associates more closely with the non-template strand as observed in footprinting assays [22, 23].

Figure 3. Primary Structure of ZNF143.
Schematic of the protein domains in human ZNF143. The primary isoform of ZNF143 is composed of 638 amino acids with two separable activation domains located between amino acids 51-236. AD1, mRNA activation domain; AD2, small RNA activation domain; DBD, DNA binding domain.

Unlike other characterized zinc finger DNA-binding domains, activation domains of transcriptional regulators remain poorly understood. Like many activator proteins, the activation domains of ZNF143 are bipartite, but unique in having separable domains for distinct promoter types. In this case, ZNF143 contains separate snRNA and protein-coding (mRNA) activation domains. The mRNA gene activating region is novel to ZNF143 and its paralogs, located from amino acids 51 to 148 and consisting of four 15 amino acid repeats [15]. Each repeat has a consensus of QAVQLEDG(S/T)TAYI(Q/H)H spaced with 10-12 amino acid distance between them [15]. The snRNA gene activating domain is smaller, located within the region from amino acids 148 to 236, and contains a conserved subdomain sequence, ALEQYAAKV, within the center of this region [15]. Mutation of the leucine residue in this segment to an alanine abrogated the activation capability[15]. No atomic structures of ZNF143 activation domains are currently known. This is likely due to intrinsically disordered regions known that frequently exist within

activation domains of many transcription factors [24]. Some activation domains even have the ability to form phase separated condensates along with co-activator proteins at super-enhancer regions *in vivo* [25]. This inherent disorder and low complexity sequence of the ZNF143 mRNA gene activation domain could be why structural information remains elusive. Lastly, the C-terminal region of ZNF143 has an unknown function and displays much reduced sequence conservation.

Highly conserved orthologs of ZNF143 are found throughout vertebrate species, and have been studied in humans, mice, Xenopus and zebrafish. Amino acid identity of 84% exists between human and Xenopus proteins (Staf)[26], and 71% between human and zebrafish proteins[27]. In humans, there is one gene paralog to ZNF143, Zinc Finger Protein 76 (ZNF76). ZNF76 contains an overall 78% amino acid content similarity when compared to ZNF143 [28]. Both of these proteins bind SPH sites with similar affinities, and are able to activate snRNA and mRNA promoter regions transcribed by RNAPII and RNAPIII [26]. RNA expression analyses of these proteins reveal that both ZNF143 and ZNF76 are detectable in most human tissues, with a somewhat higher abundance in the testis and ovary [26]. Although exhibiting similar DNA binding and transcriptional activation abilities, there are a few differences that separate ZNF76 and ZNF143. ZNF143 plays a large role in cancer and tumorigenesis, while the role of ZNF76 is still unclear. While ZNF143 expression is increased in response to treatment with cisplatin, ZNF76 levels are reduced [29]. ZNF143 is also induced in response to DNA damage, while ZNF76 is not [29]. In conclusion, although having distinct functional roles in

cancer and tumorigenesis, the activation and DNA binding properties of ZNF143 and ZNF76 overlap greatly.

Due to an evolutionary duplication of much of the zebrafish genome, it was possible to investigate different roles of two ZNF143 paralogs, called ZNF143a and ZNF143b[28]. These zebrafish paralogs are highly conserved, with almost identical DNA-binding domains, and demonstrate comparable transcriptional activation potentials with a synthetic mRNA gene promoter. However, expression levels from the paralogous genes are dissimilar during zebrafish development, and their phenotypic effects after partial selective knockdown are not completely compensatory[28].

Furthermore, other ZNF143 variants can be synthesized as a result of alternative RNA processing. Inspection of human *znf143* gene tracks on the UCSC genome server [30] shows at least two additional transcripts lacking a coding exon, presumably as a consequence of alternative pre-mRNA splicing (results not shown). At present, the actual existence and phenotypic consequences of these alternative transcripts and variant proteins are unknown and await future investigation.

ZNF143 gene targets include a multitude of snRNA and protein-coding promoters

Binding sites for ZNF143 were originally discovered in chicken *U1* and *U4* snRNA gene enhancers [10]. With increased examination of promoter regions, ZNF143 binding sites can be found in at least 70% of vertebrate snRNA and snRNA-like promoter regions, including many of the U class of snRNAs transcribed by RNA Pol II or RNA Pol III [12]. These genes include the human *U1, U4C, U6, Y4, H1 RNA, 7SK* [11, 12, 31-36]. Vertebrate snRNA promoters are generally composed of a proximal

sequence element (PSE) within 100 base pairs of the transcriptional start site (TSS), and a distal enhancer-like sequence element (DSE) generally located close to 200bp upstream of the TSS. The PSE is essential for proper transcription initiation of snRNA genes, while the DSE stimulates transcription typically by stabilizing the initiation complex [37]. SPH sites are commonly located within the DSE [12, 26, 38]. However, in some instances the SPH site is located adjacent to the PSE [31, 39].

ZNF143 binding sites at snRNA DSE regions commonly occur near octamer elements[12, 26, 38]. Oct-1 is a highly characterized transcriptional activator protein that binds to the octamer motif. Although Oct-1 and SPH binding sites are often in close proximity, these elements appear to function independently, with maximal gene expression influenced by both [10, 20, 40-43]. The further activation of expression from a snRNA template with the addition of both ZNF143 and Oct-1, and a lack of cooperative binding between these two factors [22], suggests that the modes of activation between these two factors could be distinct.

The ability of ZNF143, or ZNF76, to stimulate transcription from a synthetic protein coding promoter was discovered in a co-transfection assay in Drosophila SL2 cells [26]. Similar results were obtained in *Xenopus* oocyte injection experiments with a synthetic mRNA promoter CAT reporter plasmid[26]. The first bona fide mRNA promoter discovered to be regulated by ZNF143 was the chaperonin containing t-complex polypeptide 1 (*Ccta*) gene identified from a yeast one-hybrid screen [13]. The DNA binding domains of ZNF143 and its paralog, ZNF76, were found to recognize and bind two SPH elements within the *ccta* promoter in an electrophoretic mobility shift

assay [13]. Following these early experiments, a multitude of endogenous genes were discovered containing SPH elements with transcription stimulated by ZNF143 [44-49]. Whole-genome assays demonstrated that ZNF143 potentially regulates a large number of protein coding genes, possibly 2000-3000 [14, 19]. Whereas SPH sites in small RNA gene promoters are the canonical 18bp motif, ZNF143 binding sites in protein-coding promoters contain a combination of those with the canonical motif or the expanded motif with the additional ACTACA sequence[14]. Interestingly, an abundance of bidirectional promoters contain ZNF143 binding sites [50].

ZNF143 has been identified as an essential regulator of cell cycle progression genes. Understanding transcriptional regulation of cell cycle genes has been somewhat elusive. However, ZNF143 has emerged as a positive regulator of cell cycle genes as well as being significantly expressed in cancer cells and tumors [51-58]. One of the first identified cell cycle genes regulated by ZNF143 is the *BUB1B* gene, encoding for the BubR1 protein involved in mediating spindle-checkpoint activation [59]. Proper regulation of the spindle checkpoint is important to prevent premature entry into anaphase during cell cycle. ZNF143 was found to bind two different SPH cognate elements within the *BUB1B* promoter, and mutations within these elements led to a sharp decrease in promoter activity [59]. A loss of BubR1 synthesis leads to chromosomal instability and aneuploidy [60]. In addition to *BUB1B*, other genes encoding predominant kinases involved in cell cycle, such as *PLK1*, *AURKB*, *WEE1*, and *CDK2* are targets of regulation by ZNF143 [61].

Additional research supports ZNF143 as a vital regulator of cell cycle progression. ZNF143 siRNA knockdown analysis in human prostate PC3 cells led to 27% of the resulting downregulated genes having involvement in cell cycle and DNA replication [61]. These downregulated genes contained at least one SPH site within 1kb of the TSS, evidence of likely regulation by ZNF143 [61]. Additionally, knockdown of ZNF143 strongly induces G2/M arrest in prostate cancer cells, and likely acts distinctly from the E2F family of transcription factors that are known to regulate cell proliferation [61]. As expected from ZNF143 knockdown studies, prostate cancer cells subjected to forced overexpression of a tagged ZNF143 caused decreased cell proliferation and an upregulation of *PLK1*, *AURKB*, and several MCM encoded proteins involved in cell cycle [62]. Transcriptional regulation of cell cycle related genes by ZNF143 is essential for proper cell division.

More recently, an analysis of ZNF143 binding in human A549 cells provides evidence that ZNF143 may play a role as a mitotic bookmarking factor to help maintain cellular fate and identity[63]. During mitosis, some ZNF143 binding sites retain ZNF143 when compared to the ZNF143 binding patterns in interphase[63]. It is possible that ZNF143 aids in transcription initiation once cells re-enter into G1 phase[63].

Dysregulation of ZNF143 in Cancer, DNA Repair, Developmental disorders

Mutations in transcription factor genes have long been associated with cancer and tumorigenesis [64]. ZNF143 is involved in a large array of cancers including ovarian [55], gastric [58], lung [65], breast [57, 66], colon [52, 53], glioma [65], and esophageal [67]. The cellular mechanism of ZNF143 in cancer is not completely elucidated.

However it is involved in carcinogenesis through enhancing cell motility [52, 57], tumor progression [53], and cancer migration [58]. Additionally, ZNF143 promotes the epithelial–mesenchymal transition (EMT) through PI3K/AKT signaling that confers metastatic properties upon cancer cells [58].

ZNF143 was originally demonstrated to have an important role in cancer as a transcription factor that can bind to cisplatin-modified DNA [29, 68-70]. Cisplatin is a widely used chemotherapy drug that is used to treat solid tumors through its ability to form various adducts on DNA that prevent replication or transcription. Cisplatin resistance is commonly a problem with the treatment of solid tumors. ZNF143 is likely to be involved in cisplatin resistance through its involvement in DNA repair[29].

Many diseases and disorders are associated with key mutations in genomic regulatory regions bound by transcription factors or cofactors, chromatin regulators, or noncoding RNAs [64]. ZNF143 is involved in hematopoietic development through the regulation of the *CEBPA* gene encoding for CCAAT-enhancer-binding protein (C/EBPa)[21]. ZNF143 is implicated in polycystic kidney disease associated with a promoter mutation in *PMM2* (phosphomannomutase 2) that causes congenital disorder of glycosylation 1a (CDG1A)[71].

ZNF143 plays a role in corneal epithelialization and choroideremia. ZNF143 promotes mesenchymal to epithelial transition and upregulates expression of epithelization genes in human corneal endothelial cells [72]. Patients with chorioretinal disease contain ZNF143 binding site mutations in the *CHM* gene[73]. ZNF143 participates in embryonic stem cell proliferation by regulating transcription of the *nanog*

gene[74]. Not surprisingly because of its widespread promoter occupancy, ZNF143 participates in many developmental pathways. Knockdown studies of ZNF143 in zebrafish embryos causes brain and heart phenotypes [27]. Furthermore, analyses using haploid mammalian cell lines demonstrate that *znf143* is an essential gene[75].

Possible mechanisms by which ZNF143 activates gene transcription

As described previously, the two ZNF143 activation domains are quite specific for either mRNA or snRNA gene promoters[15]. The possibility of distinct mechanisms used by ZNF143 at these different promoters is alluring. However, we do not yet know enough to support this hypothesis. The following discussion describes some evidence of direct and indirect transcriptional activation mechanisms but illustrates large gaps in our knowledge.

Most small RNA gene promoters contain both octamer (OCT) and SPH elements in their distal enhancer-like regions. Binding of Oct-1 or ZNF143 to their respective sites stimulates the formation of stable pre-initiation complexes [41]. The Oct-1 specific mechanism involves the recruitment of the SNAPc protein complex, a multiprotein transcription factor complex required for basal transcription of snRNA genes that binds to the PSE [76, 77]. The snRNA gene activating domain of ZNF143 contains sequence similarities to the POU domain of the Oct-1 transcription factor, and it could bind a similar target in the SNAPc complex either directly or indirectly to activate transcription of snRNA genes [15]. Furthermore, ZNF143 is able to activate small RNA genes that are not co-regulated by Oct-1, for example, the Xenopus $tRNA^{sec}$ gene [78]. The relationship

between ZNF143, Oct-1, and the SNAPc complex, and the complete mechanism by which ZNF143 activates small RNA promoter regions is unclear.

Another protein, GABP, is enriched near Oct-1 and ZNF143 at many human small RNA gene promoters. In ChIP-seq gene subsets enriched for basal snRNA transcription machinery, most snRNA-like genes bound the POU2F1 domain of Oct-1 (93%), ZNF143 (81%), and GABP (45%) activator proteins [79]. Combinatory binding of each of these three activators was common for proper gene expression of snRNA like genes, and very few contained enrichment of solely one activator protein [79]. To further analyze the role of ZNF143 as an activator at the *U1* snRNA promoter during the mid-G1 phase of mitosis, endogenous levels of ZNF143 were depleted using siRNAs. Resulting ChIP-enrichments of transcription factors Oct-1, GABPA, SNAPc, GTF2B, and RNAPII were reduced by 40-70%. This suggests ZNF143 may activate snRNA genes through affecting the levels of other common snRNA regulators or RNAPII (Fig. 4). However, ample information remains to be learned regarding the interrelationship between ZNF143 and GABP, and other potential transcription factors at small RNA gene promoters.

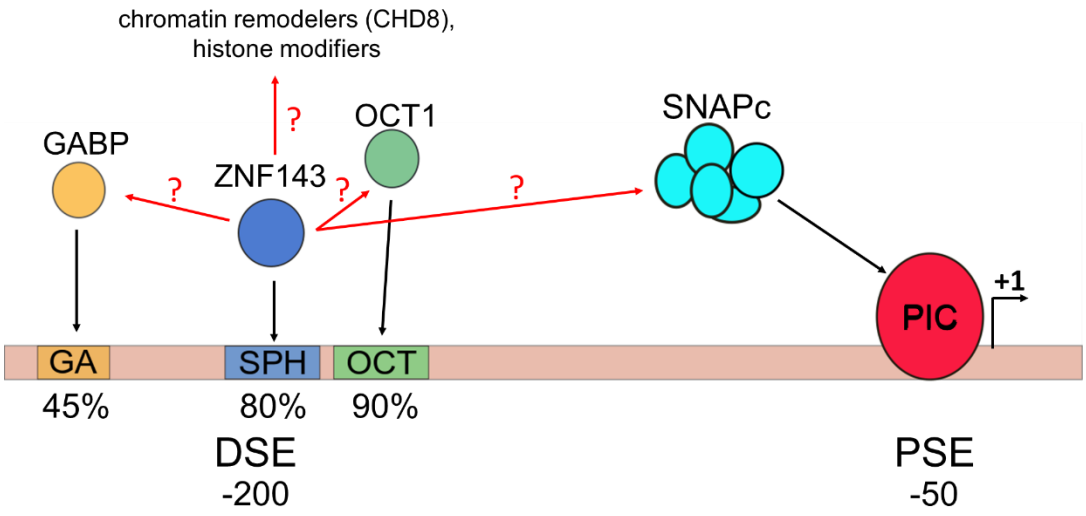


Figure 4. Potential mechanisms of ZNF143 activation at small RNA genes. SPH sites that bind ZNF143 commonly co-occur alongside GA and OCT elements within the DSE of small RNA promoter regions. Binding of ZNF143 may: 1. Recruit activator proteins such as GABP and/or OCT1; 2. Promote a transcriptionally permissive chromatin architecture through recruitment of chromatin remodelers, such as CHD8, or other histone modifying proteins; 3. Directly or indirectly recruit the SNAPc complex to the PSE to facilitate PIC formation. DSE, distal sequence element; PSE, proximal sequence element; PIC, pre-initiation complex.

Additional evidence points to a role for chromatin remodeling in the activation by ZNF143 at snRNA gene promoters. When added *in vitro* to a chromatinized *U6* promoter, ZNF143 was able to bind to the DNA template as observed in foot printing assays and activate transcription [80]. In addition, ZNF143 colocalized with the CHD8 protein on the *U6* promoter, as well as co-immunoprecipitating in a HeLa cell nuclear extract [80]. CHD8 is a member of a large family of ATP-dependent chromatin remodeling proteins [47]. As demonstrated in transient transfection assays, CHD8 further coactivates ZNF143-stimulated transcription from a synthetic mRNA promoter[81]. Surprisingly, the mechanism of CHD8 co-activation is not dependent solely on chromatin remodeling. A CHD8short isoform lacking the ATPase domain,

required for remodeling activity, can function as a stronger coactivator than the full length CHD8 isoform in transfection assays [81].

Besides a role in chromatin remodeling illustrated by the ZNF143-CHD8 interaction, evidence exists for other mechanisms by which ZNF143 affects RNAPII transcriptional regulation. These include competitive binding with other transcription factors at overlapping sites, stimulation of preinitiation complex formation, alterations in epigenetic marks at target promoters, and indirect effects caused by ZNF143 activation of genes encoding other transcriptional control proteins (Fig. 5). Another potential mechanism, although lacking any current evidence, is that ZNF143 could regulate transcriptional elongation through recruitment of other proteins.

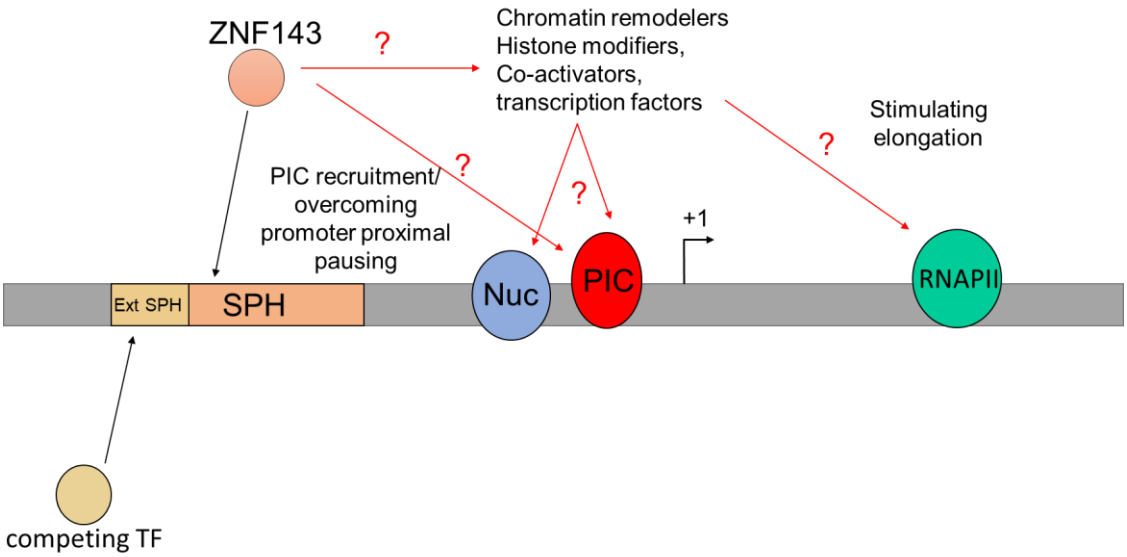


Figure 5. Potential mechanisms of ZNF143 activation at protein coding genes. SPH sites commonly occur within a few hundred base pairs upstream of the transcription start site. These sites may contain the extended 5' ACTACA SPH sequence causing transcription factors that recognize similar consensus sequences to exclude binding of ZNF143. ZNF143 may stimulate activation at protein coding genes by: 1. Recruiting chromatin remodelers, histone modifiers, co-activator proteins, or other transcription factors to the promoter; 2. Facilitating PIC complex formation or overcoming polymerase proximal promoter pausing; 3. Stimulating transcriptional elongation. TF, transcription factor; Nuc, nucleosome; PIC, pre-initiation complex; extSPH, extended SPH sequence.

Delineating transcription factor regulation within the genome has been made easier with the advent of ChIP-seq technology. A combinatorial binding of transcription factors integrate cellular signals, which leads to a context specific regulation of gene expression. The elongated binding site for ZNF143 partially overlaps with binding elements of other transcription factors in the cell. Mutually exclusive binding events have been discovered between ZNF143 and two other non-related transcription factors, the intracellular portion of Notch1 (ICN1) [14, 82] and THAP11 [14]. The binding site of THAP11 always coincides with SPH elements, overlapping the accessory ACTACA motif that exists in expanded SPH elements [14]. The Notch1 consensus element

overlaps with the TTCCCA sequence contained in the primary SPH site [14]. Both proteins can bind SPH elements *in vitro* and *in vivo* [14, 82]. Alternatively, ZNF143 is also able to bind about 40% of Notch1 sites [82]. While Notch1 is mostly a positive regulator, THAP11 acts generally as a repressive protein, and out-competes ZNF143 for binding [14]. As a response, RNAPII occupancy decreases at target promoters when THAP11 is overexpressed and ZNF143 is reduced [14]. Additionally, promoters containing the expanded ZNF143 ACTACA-containing binding sequence have lower ChIP-seq enrichments of RNAPII and active H3K4/H3K36me3 markings [14]. This supports a competition model where transcription factors with partial overlapping consensus sequences can bind exclusively to target gene sets, and thereby compete against each other for regulation.

ZNF143 likely activates transcription at the level of pre-initiation complex formation. At protein coding genes, depletion of the ZNF143 protein has been reported to cause a decrease in RNAPII density [83]. Transcriptional profiling revealed that ZNF143 regulates the amount of RNAPII density within the gene body after the TSS [83]. This decrease in RNAPII in response to ZNF143 knockdown could be attributed to a decrease in transcription initiation or prevention of pause release of RNAPII [83]. Future experiments need to be performed to delineate between these possibilities.

ZNF143 may contribute to transcriptional activation via alterations in epigenetic marks contained at target promoter regions. Mutations within the ACTACA accessory sequence of the expanded ZNF143 binding site has implications for the amount of RNAPII recruitment and active histone markings contained at ZNF143 target promoter

regions. Vinckevicius et. al. performed a CRISPR-Cas9 endonuclease aided homology directed repair to introduce transverse mutations into ACTACA binding motifs contained in two candidate endogenous promoters bound by ZNF143/THAP11/HCFC1 in SW620 human cells [84]. Resulting ChIP analyses using allele specific primers revealed that the mutant ACATACA clones had a reduced amount of ZNF143/THAP11/HCFC1 binding at the *OPHN1* promoter with a concurrent reduction in RNAPII, trimethyl-H3K4, and acetyl-H3K9 levels [84]. This study suggests that permissive histone modifications might be involved in the activation of the promoter regions containing ZNF143 binding sites including the 5' ACTACA motif bound by the ZNF143/THAP11/HCFC1 complex. Additional protein cofactors likely provide a bridge between the binding of ZNF143 and changes in chromatin modifications. Furthermore, specific cell types with distinct chromatin architectures contain some differential ZNF143 binding events [17]. In a comparison of ZNF143 binding in GM12878 K562 and HeLaS3 cells, it was revealed that ZNF143 binding may depend on mono- and di-methyl H3K4 marks [17]. These permissive histone modifications could be involved in recruiting ZNF143 or be a result of ZNF143 binding, but the relationship between permissive chromatin modifications and ZNF143 is still unclear.

ZNF143 may play a global, indirect role in modulating histone modifications. In hepatocellular carcinoma (HCC) cells, an increased level of ZNF143 activates transcription of the *mdig* demethylase gene[85]. The corresponding reduction in H3K9me3 enrichment can support increased expression from the CDC6 promoter and fits a model with HCC cell proliferation and tumor growth [85]. Together, these results

suggest ZNF143 as potentiating the modulation of epigenetic marks at specific target promoter regions through a cofactor protein. It will not be surprising if similar indirect mechanisms are discovered whereby ZNF143 regulates other important transcriptional regulatory protein genes. For example, the two major promoters for the human *chd8* gene contain possible ZNF143 binding sites (unpublished results).

Prominent regulator proteins in development commonly have their own autoregulatory feedback loops as a mechanism to control their own expression level [86]. With a pronounced role in development and expressed in most tissues [26], ZNF143 is no exception. However, not all tissues and cell types express a consistent level of ZNF143. ZNF143 has a high expression in various in human tumors[61], yet decreases in expression as malignancy level increases in breast cancer tumors[57]. A tight control of ZNF143 is required for varying expression levels in distinct cell types. Regulation of *znf143* encompasses both a transcriptional auto-regulation feedback loop and a post-translational control of different RNA isoforms through alternative polyadenylation [87, 88] that fine-tunes its expression in a tissue specific manner.

Ngondo et al. discovered that major regulation of ZNF143 involved a transcriptional autoregulatory feedback loop that depends on two different TSSs with efficiencies controlled by ZNF143 binding sites of varying affinities [87]. The resulting two RNA isoforms contain the same open reading frame with differing 5' UTR lengths but have different translational efficiencies. A balance is determined through the potential use of lower-affinity SPH sites at higher ZNF143 levels that restricts transcription of the longer transcript in favor of a usual weakly expressed shorter

transcript. In conditions where ZNF143 is less abundant, a higher affinity SPH site drives the production of the longer *znf143* transcript[87].

In addition to an auto-transcriptional regulatory switch that regulates the amount of mRNA depending on the abundance of ZNF143, expression is regulated at the translational level. The presence of three different *znf143* transcripts with different 3'UTR lengths were discovered in human HEK293 and HeLa cells [88]. The different 3'UTRs vary with regard to the inclusion of an ARE element, commonly associated with RNA instability, and a mir-590-3p micro RNA target site, resulting in differential stabilities [88]. miR-590-3p translational repression was found to be associated with a decrease in ZNF143 protein in retinoic acid induced teratocarcinoma cells [88]. To summarize, ZNF143 levels are subject to regulation both through an auto-regulatory transcriptional feedback loop and through alternative polyadenylation site usage.

One important facet of gene regulation that influences activation, repression, and recombination, is chromatin topology. There is an increasing amount of evidence that distal regulatory elements exist in such a manner that they can come into close proximity with target promoter regions through chromatin loops[17]. Established looping factors, such as CCCTC-binding factor (CTCF) and the Cohesin complex, are key insulator proteins that participate in greater than 80% of chromatin interactions [89]. It is hypothesized that CTCF acts through an array of different transcription factors that are recruited to a specific locus in a tissue-specific manner [90]. ZNF143 is likely to be a participating cofactor in CTCF-mediated chromatin interactions due to its colocalization

at CTCF-enriched clusters, positive influence on chromatin domains, and insulating properties.

ZNF143 was first revealed to be involved in long-range chromatin interactions in human leukemia cell lines through ChIA-PET experiments [91]. This evidence corroborated a previous study showing co-localization of CTCF, Cohesin, and ZNF143 [16]. Since then, many groups have identified ZNF143 as a factor significantly enriched at chromatin anchors [17, 92, 93], and likely cooperates with CTCF and Cohesin to establish chromatin structure [93]. However, ZNF143 binding motifs are not always present at locations with enrichment of both CTCF and ZNF143. Additionally, the correlation of ZNF143 at CTCF-Cohesin co-localization sites tends to involve weaker ZNF143 binding sites [17]. These binding site enrichments along with motif analyses likely suggest ZNF143 is acting as a cofactor participating in CTCF mediated chromatin interactions (Fig. 6).

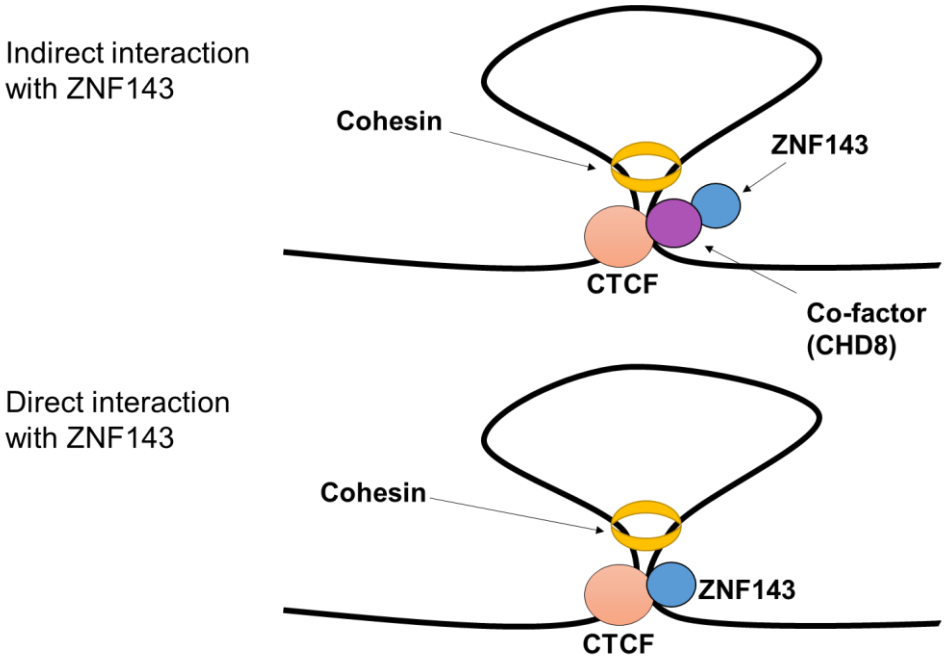


Figure 6. ZNF143 engages in CTCF mediated chromatin interactions. ZNF143 is enriched at chromatin loops along with CTCF and the Cohesin protein complex. SPH sites are not always present in ZNF143 enriched chromatin interactions, suggesting a possible indirect involvement of ZNF143. An additional co-factor protein that makes direct interactions with CTCF and ZNF143 may recruit ZNF143 to CTCF bound chromatin loops. Through an unknown mechanism, ZNF143 acts as a positive driver of chromatin loops, and has insulating properties like established architectural proteins.

It remains unclear how ZNF143 participates in chromatin loops. ZNF143 does however have a positive effect on the number of chromatin loops established [34,69]. In addition to influencing the number of chromatin loops, ZNF143 likely has insulating properties similar to other predominant architectural proteins [94]. Insulator proteins act as physical barriers within chromatin, blocking interactions between enhancers and promoters [95]. Therefore, ZNF143 has an architectural role in the genome in addition to its activating role at specific promoters.

ZNF143 and CTCF share a mutual direct protein interaction with the chromatin remodeler, chromodomain-helicase-DNA-binding protein 8 (CHD8). This finding could be significant because few direct protein binding partners have been discovered for both CTCF and ZNF143. CTCF has been reported to associate with limited transcription factors [96], topoisomerases [97], and chromatin remodelers [98, 99]. A direct CTCF-CHD8 protein interaction has been described in mouse whereby CHD8 interacts with CTCF through its BRK domains, and is essential for insulator activity of H19 DMR insulator [99]. It is hypothesized that CHD8 contributes to insulating activity of CTCF through the establishment of nucleosome free regions that may prevent the spread of certain epigenetic modifications [99]. As described earlier, CHD8 is involved in transcriptional activation through its interaction with ZNF143 [31]. While still unexplored, a potential mechanism might exist whereby CHD8, with its chromatin modifying activity, could be recruited by a combination of CTCF and ZNF143 interactions at chromatin anchors in order to establish nucleosome free regions.

ZNF143 is a unique dual-natured eukaryotic transcription factor containing the ability to activate small RNA and mRNA encoding genes. Ubiquitous in vertebrates, ZNF143 regulates 2000-3000 promoter regions[14]. Not limited solely to transcriptional activation, ZNF143 also is important for establishing proper chromatin topology within the cell. A multitude of different cellular and biological processes is influenced by ZNF143 including cell cycle-regulation, cell division, cancer, DNA repair, development, and genetic disorders.

Due to expansions in genomics databases[100], a previously unidentified paralogous gene for *znf143* was identified in zebrafish. Previous studies of *znf143* in zebrafish have only included analysis of the b form, *znf143b*. Differences in protein domains between paralogous genes can lead to deviations in gene function. A well-known example of divergence in protein function is within the *Hox* gene family. Transgenic analyses have demonstrated differences functional roles between *hox* genes in a range of embryonic tissues [101]. However, paralogy does not always imply divergence of gene function. Paralogous genes can also contribute to a gene dosage effect of gene expression [102]. Before a functional analysis can be performed of ZNF143, it is critical to know how much *znf143a* is contributing to the overall gene function of *znf143*.

Therefore, a characterization of *znf143a* and *znf143b* has been performed in Chapter II. This study includes comparisons of sequence between paralogous proteins, quantitative levels of mRNA expression across development, tissue specific expression via *in-situ* hybridization analyses, transcriptional activation assays, and CRISPRi knockdowns of each paralog in zebrafish.

While a determination of quantitative levels of mRNA expression between each *znf143* paralog may not tell us if a gene dosage effect is responsive for the evolutionary maintenance of two *znf143* genes, it can provide an idea of how much of a contribution the *znf143a* paralog is making to overall expression of *znf143*. Additionally, paralogous genes generally tend to be more tissue-specific than orthologous genes [103]. The *znf143a* paralog could drive expression in a separate tissue distinct from *znf143b*.

Therefore, a mRNA hybridization assay is important in determining a tissue-specific expression of *znf143* mRNA.

In order to assess if the activation capability of each ZNF143 protein is divergent, transient transfection luciferase assays have been performed. This is critical to determine if ZNF143a maintains potent activation activity like ZNF143b. Lastly, the functional roles between each *znf143* paralog gene within zebrafish have been assessed using CRISPR interference as a genetic knockdown approach. Knocking down each paralog independently can also aid the determination of the protective redundancy of the *znf143* proteins. Before becoming functionally distinct, paralogous genes can veil a loss of function of the other gene paralog [104].

Despite the large involvement in varying biological processes, the mechanistic action of ZNF143 to activate target genes remains enigmatic. A deeper understanding of genetic regulation is critical for elucidating the underlying mechanisms of human disease. However, understanding how a ubiquitous activator protein broadly activates transcription at a variety of tissue-specific genes is a large undertaking and likely context specific. Research described in Chapter III employs a candidate promoter approach to learn more about a context-specific mechanism of ZNF143 at target genes, with a broader goal of furthering the understanding of site-specific activator proteins in eukaryotes.

To investigate promoter specific regulation by ZNF143, a transgenesis approach in zebrafish has been employed. As established, ZNF143 binds to SPH binding sites located in promoter regions of target genes. Two parallel promoter constructs have been

created to study ZNF143 function *in vivo*. This includes a comparison of promoters containing the SPH site for ZNF143, and the same promoter regions with key residues of the SPH site mutagenized in order to eliminate binding. DNA-protein interactions can be probed for between the parallel lines for the same promoter region through using chromatin immunoprecipitation analyses in zebrafish embryos. It is expected that the distortion of an SPH element *in vivo* will have a downstream effect on the amount of RNA polymerase that gets recruited or remains bound to the transgene. Although currently unknown, a mutagenized SPH element could cause differing enrichments of epigenetic markers, cofactors, histones, or even distort higher order chromatin contacts at target promoter regions. This dissertation seeks to identify architectural differences contained within target promoters regulated by ZNF143.

CHAPTER II : TWO PARALOGOUS *ZNF143* GENES IN ZEBRAFISH ENCODE TRANSCRIPTIONAL ACTIVATOR PROTEINS WITH SIMILAR FUNCTIONS BUT EXPRESSED AT DIFFERENT LEVELS DURING EARLY DEVELOPMENT*

*Cited from: Huning, L., Kunkel, G.R. Two paralogous znf143 genes in zebrafish encode transcriptional activator proteins with similar functions but expressed at different levels during early development. BMC Mol and Cell Biol 21, 3 (2020).

Background

Zinc Finger Protein 143 (ZNF143) is a sequence-specific transcriptional activator protein involved in stimulation of transcription from over 2000 mammalian promoters [14, 19]. ZNF143 binds to so-called SphI Postoctamer Homology (SPH) motifs or Staf (Selenocysteine transcription activating factor) Binding Sites (SBSs) that are located typically within a couple hundred base pairs of the transcriptional start site of gene promoters. ZNF143 was characterized initially for its transcriptional activation activity and more recently for its occupancy at the boundaries of topologically associated domains in chromatin along with the CCCTC-Binding Factor (CTCF) protein[15, 17, 19, 40, 92, 105-107]. ZNF143 is able to regulate transcription from both small nuclear RNA (snRNA) promoters and mRNA promoters by RNA polymerase II or III [15].

Additionally, ZNF143 has been implicated in regulation of cell cycle progression and tumor growth[14, 52, 59, 61, 108, 109], and is an important regulator involved in zebrafish development[27]. However, past functional studies of ZNF143 in zebrafish

have not considered a previously uncharacterized paralogous gene (*znf143a*) also expressed during early development.

Gene duplication is thought to bestow a long-term evolutionary advantage due to a decreased number of constraints present on one of the functional copies of the gene [110]. Zebrafish, among other teleost fish, have an increased copy number of genes due to a third teleost specific whole genome duplication event occurring 100 million years ago[111].Therefore, an additional copy of the gene could incorporate mutations that may confer novel function or expression. The previously studied gene, *znf143b*, has an uncharacterized paralog located on chromosome 7. Here we investigate this paralog, *znf143a*, including its quantitative and spatial expression patterns, its ability to activate mRNA promoters, and comparative phenotypic knockdown analysis through the use of CRISPR interference (CRISPRi).

Results

We characterized the novel gene *znf143a*, a paralog of the previously studied *znf143b* (protein also named Staf or SPH Binding Factor (SBF)). Quantitative expression levels, spatial expression patterns, transcriptional activation potentials of the encoded proteins, and phenotypic analyses following knockdown of both genes are reported. The two ZNF143 proteins exhibit a large sequence homology containing 65% amino acid identity, with most of this conservation in the DNA binding domain (DBD) residing in the seven C2H2 zinc fingers located between amino acids 229-438 (Fig. 7). ZNF143a encodes a slightly shorter protein containing 613 amino acids instead of the 623 amino

acid ZNF143b protein. Deviations in sequence identity between the two proteins primarily exist within the amino-terminal region, containing the mRNA activation domain from residues 51-149 (ZNF143b) and the snRNA activation domain from 150-228 (ZNF143b), and the carboxy-terminal region of unknown function (Fig. 7).

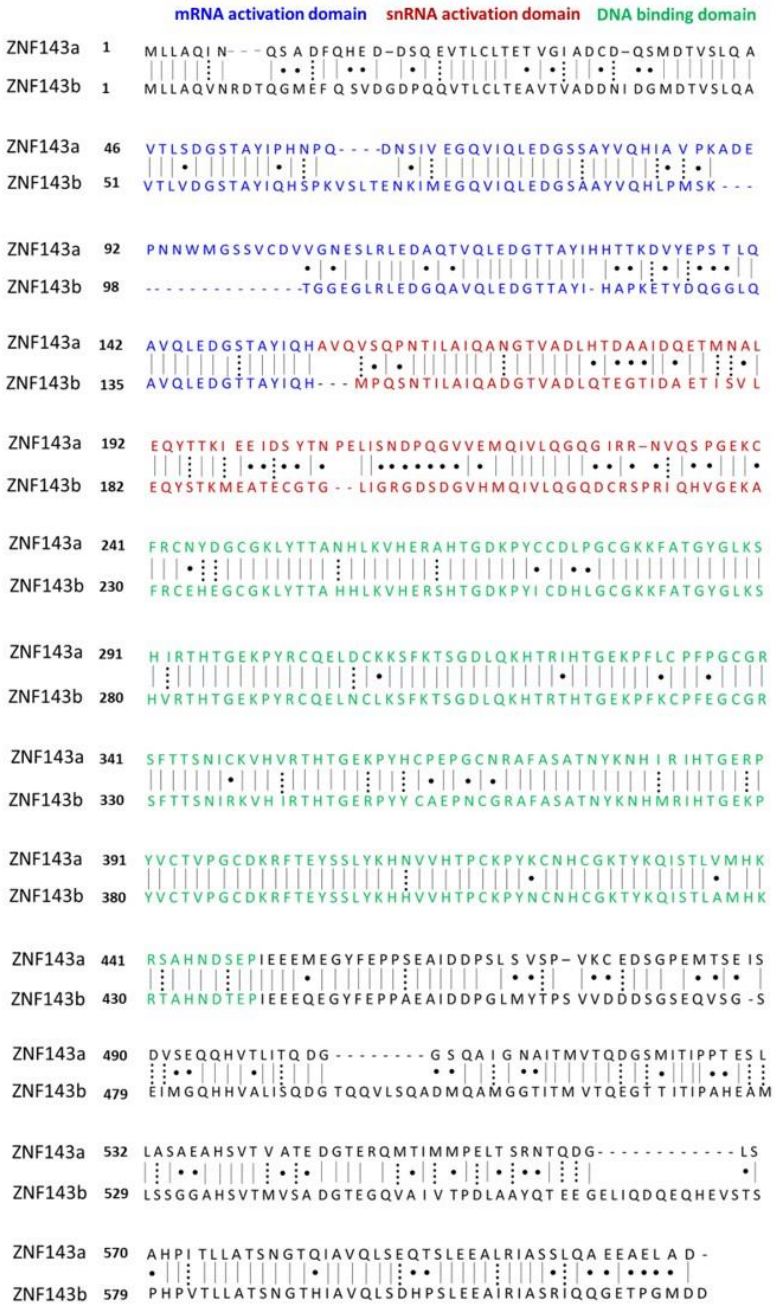


Figure 7. Amino acid sequence alignment of zebrafish ZNF143a vs. ZNF143b. Protein sequences of ZNF143a and ZNF143b were aligned with the lalign sequence analysis program using the algorithm of Huang and Miller[112]. Identical amino acids are indicated by a straight line, similar amino acids are indicated by a dotted line, and dissimilar amino acids are indicated by a single dot. Amino acids within the DNA binding domain are colored green, amino acids within the mRNA activation region are colored blue, and amino acids within the snRNA gene activation region are colored red.

Due to some sequence differences in the activation domain regions of the N-terminal region of the protein, we hypothesized that the transcriptional activation properties between the two proteins may differ. It has been demonstrated that zebrafish ZNF143b can bind and activate mRNA promoter regions such as the *pax2a* promoter[27]. To assess the capacity of ZNF143a to act as a transcriptional activator protein at mRNA promoters, transient transfection assays were performed using cultured ZF4 zebrafish cells. A synthetic promoter containing five binding sites for ZNF143 was used to drive a luciferase reporter. The addition of *znf143a* expression vector plasmid at 5ng and 10ng led to a significant increase of transcription, 12-15 13-19 times the amount of luciferase activity relative to the empty expression vector control (Fig. 8). Furthermore, we did not detect any synergistic effect after adding 5ng of each *znf143* expression plasmid. We note that previous work demonstrated a direct effect of ZNF143b in such experiments, dependent on the presence of SPH sites in the promoter and the activation domains in ZNF143b[27]. Although there was no a significant difference in activation levels between ZNF143a and ZNF143b (Fig. 8), Although both ZNF143a and ZNF143b demonstrate strong transcriptional activation activities, their relative potencies were unable to be compared quantitatively due to possible differential protein expression, as noticed in human embryonic kidney (HEK293) cells (results not shown). Similar to earlier work [27], we were unable to detect myc-tagged ZNF143 in transfected zebrafish cells. Nevertheless, ZNF143a can act as a strong transcriptional activator protein.

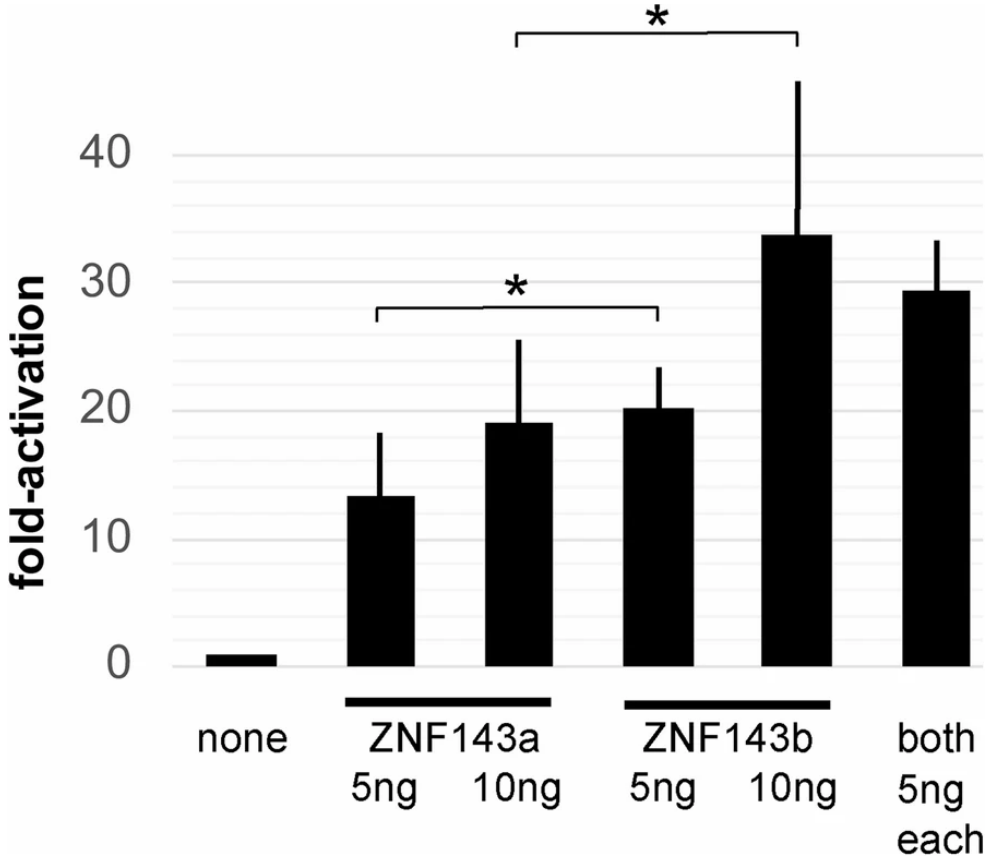


Figure 8. ZNF143a and ZNF143b exhibit similar transcriptional activation potential.
Zebrafish ZF4 cells were transfected with pGL3-SPH5 firefly luciferase reporter gene plasmid, plus pRL-SV40 renilla luciferase reporter plasmid, and as noted, pCI-myczznf143a or pCI-myczznf143b expression vector plasmid. Relative luciferase expression was determined by comparing the firefly/renilla luciferase ratio for each sample to that ratio for the sample with addition of expression vector plasmid containing no gene. Bar height shows the mean value from independently transfected wells, and error bars report the standard deviation from the mean. The single asterisk denotes a significant difference ($p < 0.05$). As noted in a previous publication [27], we were unable to detect expression of the myc-tagged ZNF143 in ZF4 cells. Expression of ZNF143b in human HEK293 cells was somewhat higher than ZNF143a, and if representative of relative synthesis in ZF4 cells, could explain the greater transcriptional activation by ZNF143b in transient transfection experiments.

Gene duplications in vertebrates can lead to a divergence in tissue specific expression that may be suggestive of varying contribution of paralogous genes to specific organ functions [110]. We hypothesized that *znf143a* may be expressed in a

tissue-type distinct from *znf143b*. Tissue specific expression patterns of both znf143 mRNAs were determined by *in situ* hybridization analyses using digoxigenin (DIG)-labeled antisense riboprobes. Due to a high sequence similarity of the two coding sequences, antisense probes were designed to target primarily the 3'-untranslated region (3'UTR) sequences of each *znf143* paralog. Both riboprobes for the paralog genes demonstrated strong expression in the brain in 24-hour post fertilization (hpf) zebrafish embryos including the forebrain, midbrain, and hindbrain regions (Fig. 9). As evident from similar hybridization patterns, expression of *znf143a* and *znf143b* genes are not spatially distinct.

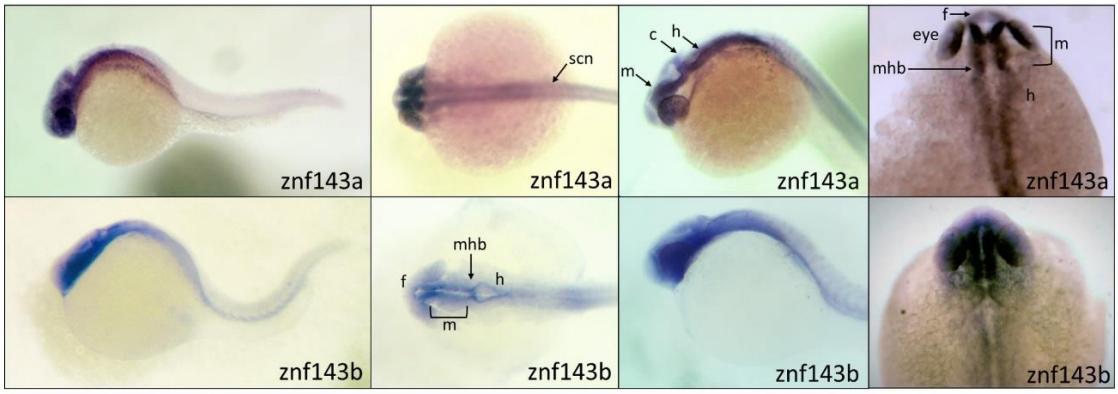


Figure 9. Similar spatial expression of *znf143a* and *znf143b* in 24hpf embryos. Antisense digoxigenin riboprobes against *znf143a* and *znf143b* were used for whole-mount *in situ* hybridization assays. The probe for *znf143b* targeted the last exon of the coding region (approximately 186 nt corresponding to the last 61 amino acids) and the 3'UTR of the gene, while the probe for *znf143a* was designed solely against the 3'UTR. Embryos used were fixed and stained at 24hpf. The panels illustrate views from four different embryos for each probe, with the right-most panels showing an enlarged dorsal view of the head. We note that in all experiments, the *znf143b* probe produced fainter staining, most likely because of a smaller than optimal riboprobe that was necessary to ensure gene specificity. Specific head, brain and neural structures are identified with the following abbreviations: f, forebrain; m, midbrain; h, hindbrain; mhb, midbrain hindbrain boundary; c, cerebellum; scn, spinal cord neurons.

Another possible outcome with two different *znf143* genes is that they may be expressed at a different developmental time periods during zebrafish embryogenesis. To explore this hypothesis, total RNA from embryos was isolated at various developmental time points including shield (6hpf), bud (10hpf), 17-somite (16hpf), and 24hpf, and converted to complementary DNA (cDNA). Quantitative PCR analyses showed that *znf143a* is expressed significantly higher in early development especially at the shield and bud stages, while *znf143b* is expressed at similar expression levels in the first 24 hours of development (Fig. 10). Both *znf143a* and *znf143b* are expressed at a similar level by 24hpf. The expression difference between bud stage and 24hpf embryos was verified by using additional primer pairs for each gene (results not shown). In these latter qRT-PCR experiments the expression level of *znf143a* decreased 2.5-fold between bud and 24hpf, while the level of *znf143b* remained constant.

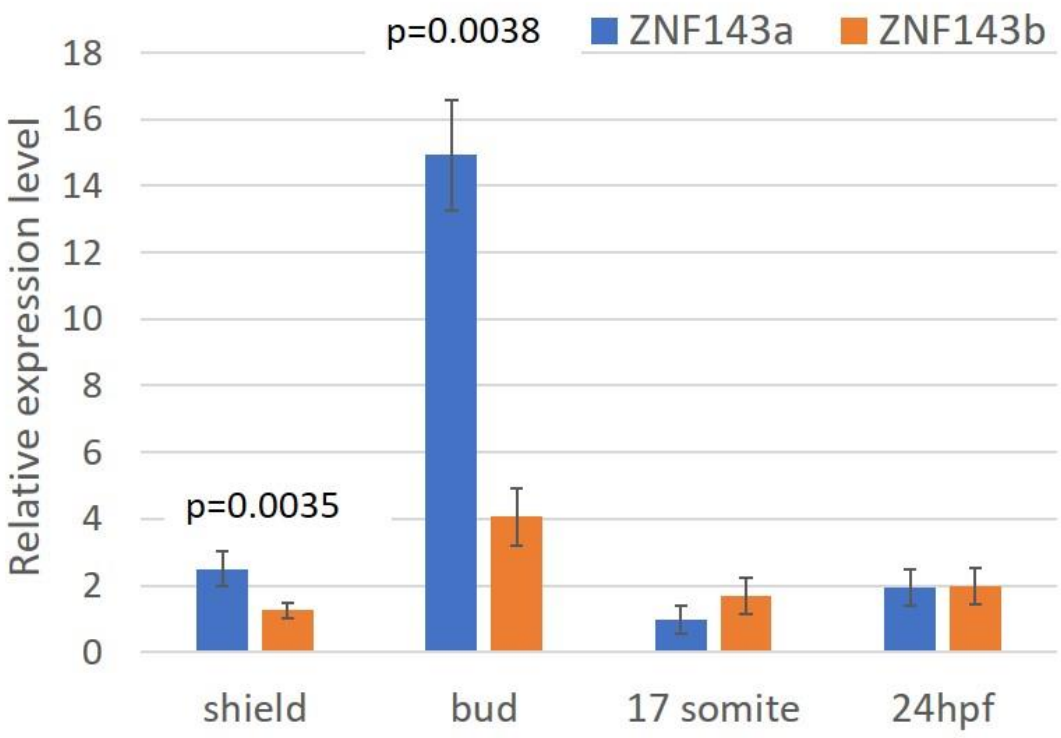


Figure 10. *znf143* paralog genes are differentially expressed in early development in zebrafish.
cDNA was obtained from total RNA isolated from zebrafish embryos at either the shield, bud, 17-somite, or 24hpf stages. qPCR was performed for both *znf143a* and *znf143b*. Numbers on the y-axis represent the mean values for relative expression of each gene after normalization to the geometric mean of two housekeeping gene controls (*ef1α* and *rpl13α*), and comparison to the lowest value (*znf143a* at 17-somite stage). Error bars represent standard deviation from the mean. Statistically significant differences in expression levels are signified by the inclusion of p-values< 0.05, determined by Student's t test.

The advent of CRISPR technologies has added a new method to knock down gene function. Repression through CRISPR/Cas9 uses a guide RNA-directed deactivated Cas9 (dCas9) protein as a roadblock for transcriptional initiation or elongation, known as CRISPRi [113, 114]. This method has been applied to study transient gene knockdown in both *C. elegans* and zebrafish that resulted in mild morphological phenotypes [115]. In this earlier report the injection of multiple strand specific guide RNAs improved gene

repression using CRISPRi [115]. We used this CRISPRi method to study individual gene function of either *znf143a* or *znf143b*. Three single guide RNAs (sgRNAs) were designed to target the early transcribed regions of the *znf143a* or *znf143b* genes (Fig. 11A). Embryos displayed similar phenotypes when knockdown was elicited with three sgRNAs targeting *znf143a* or *znf143b* (Fig. 11B,C). Phenotypic defects included a loss of midbrain/hindbrain formation, hindbrain enlargement, and an overall loss of brain organization. Knockdown of either paralog resulted in a complete loss of brain formation in about 22% of injected embryos (Fig. 11C). These fish are referred to as a "smoothened" phenotype due to a lack of brain structures causing a uniform unwrinkled head. Despite such massive defects in the brain, the "smoothened" fish still underwent somitogenesis and displayed a developed axis. A low percentage of injected knockdown embryos (less than 11%) displayed severe defects including a shortened axis (Fig. 11C). Embryos injected with either set of gRNAs, but no dCas9 protein, did not display any phenotypic effects (results not shown). A small proportion of embryos injected only with dCas9 protein displayed abnormal phenotypes (Fig. 11C). Therefore, knocking down a single paralog, either *znf143a* or *znf143b*, was sufficient to exhibit brain defects.

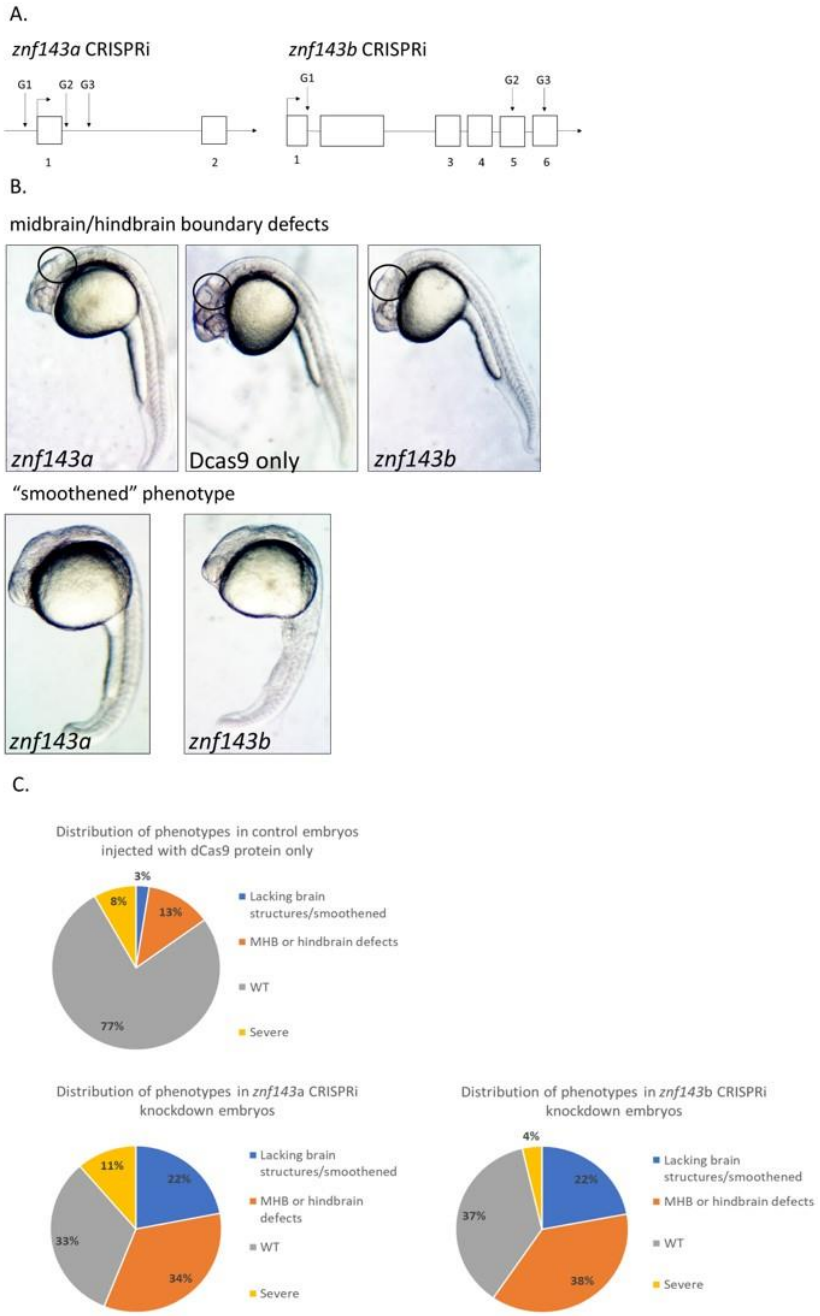


Figure 11. CRISPRi knockdown of either znf143a or znf143b induce brain developmental defects. A. Diagram showing targets for sgRNAs. Numbered boxes indicate positions of exons. B. One-cell zebrafish embryos were injected with sgRNA/dCas9 protein complexes to knock down *znf143a*, *znf143b*, or in control experiments with dCas9 protein only (−sgRNA) and observed at 24hpf. Representative photographs of abnormal brain phenotypes are shown, with circled regions showing the areas of most interest. C. Consistent phenotypes are grouped into classes. Severe phenotypes lacked axial development. 159 embryos were counted for the control injections lacking any sgRNAs, 132 embryos were counted for the znf143a CRISPRi injections, and 136 embryos were counted for the znf143b CRISPRi injections.

Although these genes have redundant functions, both are essential to zebrafish development. The specificity of gene-specific knockdown was investigated in quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) experiments that demonstrated modest, but significant reduction in mRNA levels (25-40%) (Fig. 12). The moderate reductions in mRNA levels observed are similar in magnitude to other reported multiple guide CRISPRi knockdowns in zebrafish [115]. Interestingly, knockdown of znf143a caused a 1.5-fold increase in total znf143b mRNA, an effect that was not apparent with znf143b knockdown (Fig. 12).

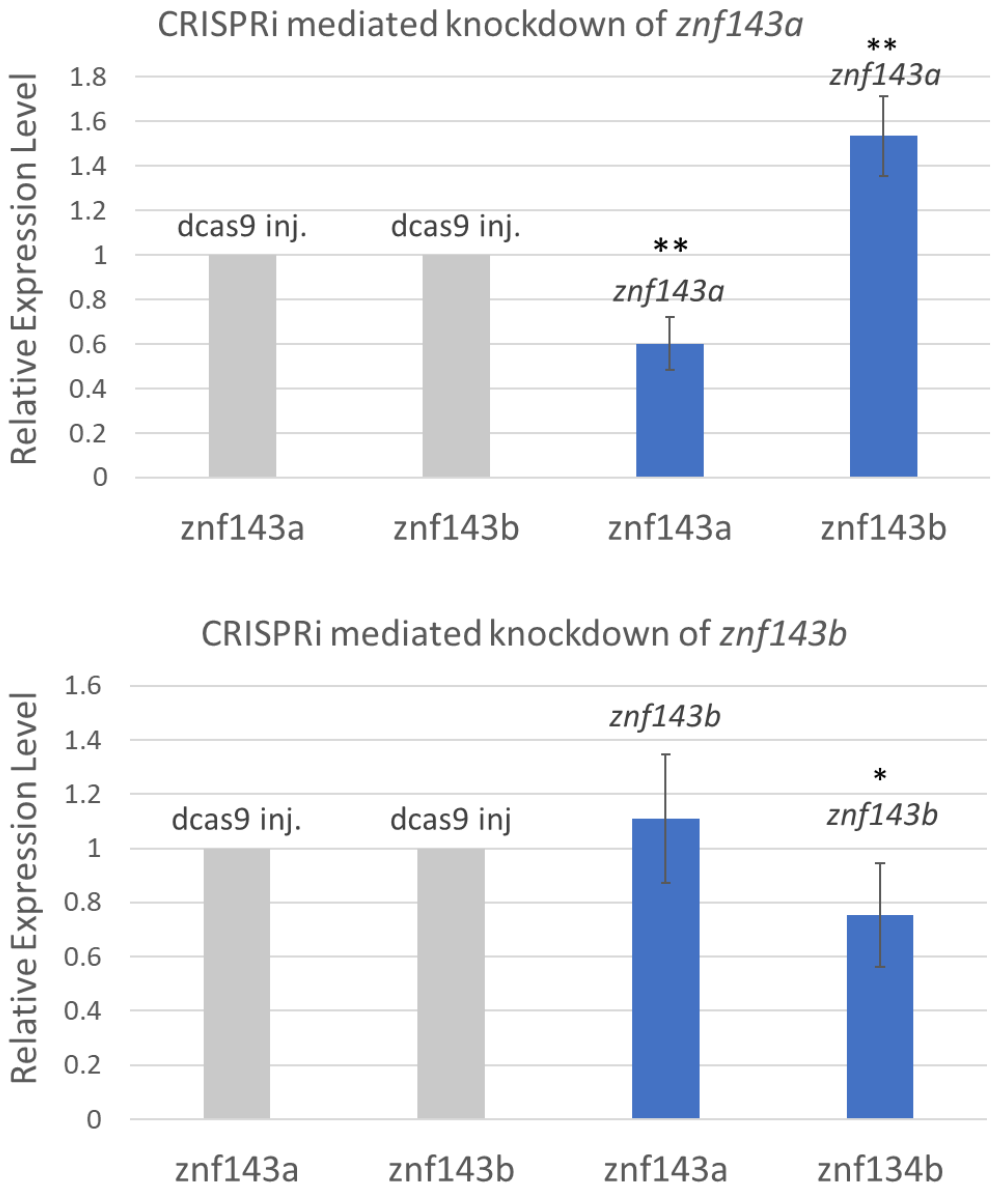


Figure 12. Analysis of specific gene knockdown in CRISPRi experiments.

Quantitative RT-PCR was used to analyze relative levels of *znf143a* and *znf143b* transcripts. The amounts of each znf143 cDNA were determined relative to those in the dCas9 injection control after normalization to the geometric mean of the widely-expressed transcripts (*ef1α* and *rpl13α*). The height of each column represents the mean of 4 or 5 independent injection experiments, and error bars represent standard deviation from the mean. A single asterisk signifies a p-value < 0.05 relative to the control sample lacking any sgRNAs, while a double asterisk signifies a p-value < 0.01.

Discussion

In this study, we show that two versions of zebrafish ZNF143 expressed from paralogous genes have similar functions, including strong transcriptional activation potential, comparable spatial expression during embryogenesis, and identical phenotypic outcomes after partial knockdown. The only notable difference that we detected was relatively higher expression of *znf143a* during early zebrafish development. Although global roles for both ZNF143a and ZNF143b are suggested by phenotypic effects following individual knockdown, future RNAseq experiments could illustrate the scope of paralog specific regulation.

Genome duplications can lead to gene paralogs that evolve different functions or tissue specificities over time. However, this is not the case with *znf143a* and *znf143b* as both exhibited similar spatial expression patterns and likely perform similar tissue-specific roles. Both proteins act as potent activators in the cell, but we did not observe a synergistic activation in transfection assays. Each paralog gene encodes for a protein able to stimulate transcription from target promoters through SPH sites. Despite conserved tissue specific expression patterns, both paralogs appear to be important for zebrafish development when targeted individually for CRISPRi knockdown analyses. We did not observe any exacerbation of phenotypic effects when expression from both paralogs was reduced simultaneously (data not shown).

In humans, there are also two *znf143* paralog genes, *znf143* and *znf76* [26]. Both genes are highly conserved among mammals. Human ZNF76 contains 57% amino acid sequence identity and 78% similarity to human ZNF143. Both proteins are able to bind

and activate promoters of genes transcribed by RNA polymerase II or RNA polymerase III containing SPH sites, and are highly expressed in most tissues [26]. Zebrafish also contain the conserved *znf76* paralog. The extra *znf143* paralog, *znf143a*, is only retained in teleost fish.

The promoters for both zebrafish *znf143* paralogs contain a putative SPH binding site suggesting a potential autoregulatory feedback loop at both genes. Human ZNF143 and ZNF76, are able to bind to SPH sites with similar affinities and activate the *znf143* promoter [19].When *znf143* or znf76 is over-expressed in cells, this can downregulate the levels of endogenous ZNF143[87]. When ZNF143 is saturated in the cell, it has the ability to bind to non-canonical binding sites, and this produces a transcriptional start site switch mechanism that can result in two different transcripts with one transcript much more efficient at production of a protein [87]. It is possible that ZNF143a, due to its ability to bind and activate SPH sites, also contributes to a similar auto-regulatory feedback loop that regulates the *znf143* gene. Alternatively, ZNF143b may be able to bind and regulate expression of *znf143a* due to the presence of an SPH site in its promoter. This cross-talk between *znf143* and other *znf143* paralogs may contribute to the fitness of the zebrafish and be an important factor when knocking down a single paralog as witnessed in knockdown experiments. The existence of ZNF143a may be another regulating factor crucial for maintaining the correct expression levels of ZNF143b.

There are multiple possibilities as to why both *znf143a* and *znf143b* are critical for the early development of zebrafish. One possibility is that *znf143a* encodes for a

protein that confers a function that is different from *znf143b*. Although zebrafish ZNF143 proteins contain 65% sequence identity, there are a few regions with divergent sequence. One such region is C-terminal to the DBD, and accounts for unknown function in both ZNF143 paralogs. ZNF143a contains two deletions in the C-terminal region of the protein encompassing 8 amino acids and 12 amino acids when compared to ZNF143b (Fig. 6). A second possibility for the requirement of both *znf143* paralogs for healthy development is the temporal difference in quantitative expression of *znf143a* and *znf143b*. Coordinated control of gene expression, both spatially and temporally, is critical for proper development of an organism. Expression of *znf143a* early in development or later expression of *znf143b* could help maintain the specific chromatin environment necessary for the correct gene products to be produced at the correct time during development of the zebrafish embryo. Lastly, paralogous duplicate genes may be retained in evolution to allow more control of gene dosage[116, 117].

Multiple copies of *znf143* that express at different times could contribute to a more finely tuned dosage effect that contributes to the overall fitness of the zebrafish during development. Necessity for a narrow window of *znf143a* or *znf143b* gene dosage could provide a rationale for the strong phenotypic defects observed in our CRISPRi knockdowns, despite only having a modest decrease in overall quantitative gene expression. In addition, paralogous versions of *znf143* can buffer mutations arising in either gene by the mechanism known as transcriptional adaptation[116, 118]. Possibly, the modest increase in *znf143b* mRNA level after CRISPRi knockdown of *znf143a* (Fig.

11) could reflect transcriptional adaptation induced by short transcripts produced by knockdown with guide RNAs binding downstream of the transcriptional start site.

Methods

Plasmid constructions

The pCI-myczznf143b expression plasmid and pGL3-SPH5 firefly luciferase reporter plasmid were described previously [16]. In order to construct the pCI-myczznf143a plasmid, three fragments of znf143a were amplified by PCR from zebrafish cDNA and ligated together using naturally occurring restriction enzymes within the *znf143a* sequence and an engineered MluI site and NotI site added into the 5′-end and 3′-end primers, respectively. Additionally, a single myc tag was engineered at the amino-terminus. The first and second fragments were ligated using a naturally occurring unique EcoRI site, and the second and third fragments joined using a naturally occurring unique BamHI site. The PCR fragments were restricted with the appropriate restriction enzymes, gel-purified, and ligated together into the pCI-neo vector (Promega) that was previously restricted with MluI and NotI. The znf143a sequence was verified by the dideoxy method.

A plasmid was constructed using the pGEM-T vector (Promega) in order to synthesize a znf143a riboprobe for *in situ* hybridization assays. pGEM-znf143a was designed to target the 3'UTR of *znf143a*, as this region of the gene was the most distinct from znf143b. Primers used to amplify the 3'UTR region of znf143a had the following

sequences: 5′- CCACCTTCACCTTGAGAC-3′, and 5′-AATATCACCATCATCAGTTTA-3′.

Plasmids were constructed using the pDR274 vector [28], obtained from AddGene, in order to synthesize single guide RNAs (sgRNAs) for CRISPRi. Guide RNAs were designed stepwise in the sense direction to target the 5′ transcribed regions of either *znf143b* or *znf143a,* and construction of the pDR274 plasmids followed a standard protocol [28]. Guide RNAs targeted the first intronic region of the *znf143a* gene with the following primers: Guide 1 Forward: 5′-TAGGCGATCTGCAGTACGTTACA-3′, Reverse: 5′-AAACTGTAACGTACTGCAGATCG-3′; Guide 2 Forward: 5′-TAGGTGAAACTAGATATCGCTGC-3′, Reverse: 5′-AAACGCAGCGATATCTAGTTTCA-3′; Guide 3 Forward: 5′-TAGGAAACTAACGTTACACGCCT-3′, Reverse: 5′-AAACAGGCGTGTAACGTTAGTTT-3′. Guides targeted the first exon of the znf143b gene with the following primers: Guide 1 Forward: 5′-TAGGTGCATGGTGGTCGAACGA-3′, Reverse: 5′-AAACTCGTTCGACCACCATGCA-3′; Guide 2 Forward: 5′-TAGGGCATGGAGTTTCAGAGTG-3′, Reverse: 5′-AAACCACTCTGAAACTCCATGC-3′; Guide 3 Forward: 5′-TAGGACAAGTGATTCAGCTGG-3′, Reverse: 5′-AAACCCAGCTGAATCACTTGTC-3′. Oligonucleotides were annealed and ligated into the BsaI site of pDR274 [28]. Sequences were verified using the dideoxy method.

Cell culture, transfection, and reporter gene assays

Transient transfection experiments in zebrafish ZF4 cells or human HEK293 cells were performed as described previously [16]. Protocols with HEK293 cells followed BSL-2 guidelines approved by the Texas A&M Institutional Biosafety Committee (Permit IBC2016–047).

RNA isolation, cDNA synthesis, RT-PCR

Total RNA was isolated from wild-type zebrafish embryos at designated developmental stages, converted to cDNA, and used in qRT-PCR experiments as described previously [29]. Gene-specific primers for *znf143a* were: 5′-GCGGTTCCAAAAGCAGATGAGC-3′ and 5′-CTTCCAGCTGAACGGTCTGAGC-3′, and for *znf143b* were: 5′-CTCTACTAAGATGGAGGCCACAG-3′ and 5′-CTGGATTCTAGGTGAACGACAGTC-3′ These primer sets were chosen for binding to nonhomologous regions of the two open reading frames (ORFs).

Zebrafish husbandry, and *in situ* hybridization

AB/TL wild-type strain zebrafish for breeding to produce embryos were obtained from colleagues in the Biology Department, Texas A&M University, and were maintained using standard methods with protocols approved by the Texas A&M University Animal Care and Use Committee (AUP #2016–0102 and AUP #2019–0139). After use for embryo production, adult fish were returned to breeding stock tanks containing 10–20 animals. Embryos used for *in situ* hybridization or injection experiments were chosen

randomly. Whole-mount *in situ* hybridizations were performed according to standard methods [30]. Gene-specific DIG-riboprobes were transcribed from pGEM-znf143a and pCI-myczSBF (zznf143b [16];) plasmids. The *znf143a* probe was generated using T7 RNA polymerase after linearization with NotI and hybridized only to the 3'UTR. The *znf143b* probe was generated using T3 RNA polymerase after linearization of pCI-myczSBF with EarI. Therefore, the *znf143b* probe hybridized to the 3'UTR and a region of the ORF corresponding to the C-terminal 61 amino acids.

CRISPRi

Guide RNAs from DraI-linearized pDR274 templates were synthesized using the MAXIscript T7 in Vitro Transcription kit (Invitrogen) following an established protocol [28]. The concentrations of sgRNAs were determined using a Nanodrop spectrophotometer, and the quality of the RNA was verified by agarose gel electrophoresis. 70 ng of each sgRNA directed against a specific gene, and 3 μg of dCas9 protein (Integrated DNA Technologies; diluted to 1 mg/mL) were incubated at 37C for 10 min to form a mixture of dCas9:sgRNA complexes. Control samples contained dCas9 protein without sgRNAs. All injection samples were adjusted to a total volume of 10 μL using nuclease-free water and contained 0.2% phenol red. Approximately 1 nL of the dCas9:sgRNA mixture or dCas9 control was injected into one-cell zebrafish embryos. 30–60 embryos were injected with a single type of mixture per injection day.

CHAPTER III : MECHANISM OF ACTIVATION VIA ZINC FINGER PROTEIN 143 AT CANDIDATE PROMOTERS

Background

Transcriptional regulation is a highly conserved process that is important for the maintenance of cellular homoeostasis within eukaryotic organisms. Promoter bound activator proteins are generally regarded as being critical for recruiting basal transcription machinery to a specific gene that results in RNAPII entering transcription elongation. However, there are many critical components to this process.

Proper gene expression requires for transcription to initiate not only at the correct time, but also in the correct manner. This can require site specific activator or repressor proteins and their associated co-factor proteins to relay a specific cellular signal. Moreover, chromatin is organized in a way such that nucleosomes, consisting of histone octamer proteins, spatially prohibit transcription from occurring. For transcription factors to bind within the promoter, nucleosomes must be slid or displaced to allow polymerases access to the DNA template. Furthermore, epigenetic modifications to histones or DNA can be inhibitory or stimulatory to transcription. Each of these components contribute to the genetic landscape of a target gene. One facet of proper gene expression not well understood is the mechanism by which activator proteins act at target genes within the cell.

One pervasive eukaryotic transcriptional activator protein conserved in vertebrates is ZNF143. Dual-natured, ZNF143 can regulate both small RNA and protein

coding genes. Since its discovery, the mechanism regarding how ZNF143 activates transcription has been unknown. In this chapter, a simplistic candidate promoter approach has been employed to study activation by ZNF143. This process includes three stages: 1. A preliminary selection process to screen for model gene candidates for regulation by ZNF143, 2. The construction of parallel transgenic lines containing or lacking an SPH binding site for each gene candidate, and 3. A ChIP analysis between the two parallel transgenic lines for each promoter.

To study regulation by ZNF143 at target genes, two genetic constructs were prepared in parallel for each candidate promoter. The first genetic construct contains a native SPH binding site for ZNF143 driving an *egfp* reporter. The second genetic construct contains a SPH binding site with key residues mutated to disallow binding of ZNF143. Site-directed mutagenesis was used to mutate the CCCA residues within the SPH site, within the context of an *egfp* reporter.

My hypothesis is that a mutagenized binding site would affect the amount of ZNF143 that is bound to a gene target, and ultimately lessen the amount of RNAP that is enriched. Differential amounts of ZNF143 binding between native and mutagenized SPH sites of the two distinct transgenic lines of the same promoter region may prompt downstream changes in the enrichment of epigenetic markers, cofactors, histones, or even distort higher order chromatin contacts. Due to the association of ZNF143 and permissive histone modifications [17, 84] and decreased amounts of H3K9me3 [85], chromatin immunoprecipitation analyses of epigenetic markers were pursued in this analysis.

Zebrafish provide for an excellent animal model to study vertebrate development. As previously established, ZNF143 is critical for the proper development of the zebrafish embryo [27, 28]. Here we employ the use of *tol2* transposition[119] to integrate a candidate promoter flanked by *egfp* and regulated by ZNF143 into a stably modified germline of zebrafish. The *medaka tol2* encodes for an autonomously active transposase that can catalyze the insertion of a non-autonomous *tol2* construct[119]. *Tol2* can then integrate a construct flanked by *tol2* sites into the genome as a single copy through a cut-and-paste mechanism [119]. Both the schematic of the transgene to study activation of ZNF143 and the method to establish transgenic fish lines are shown in Fig. 13.

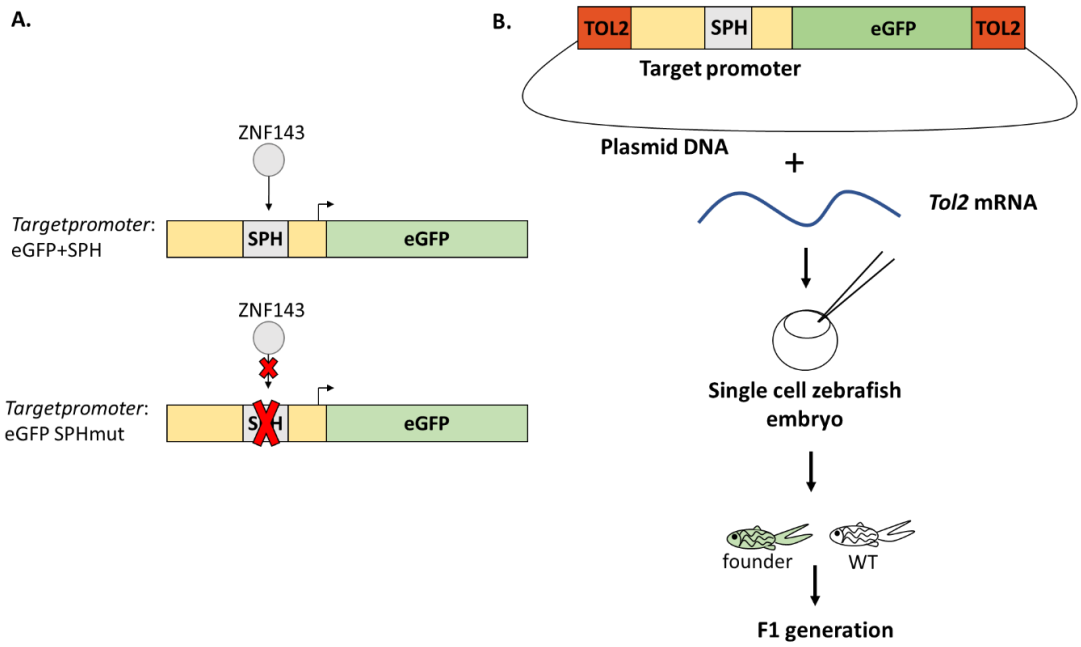


Figure 13. Construction of transgenic zebrafish lines.
A. Transgenic constructs used to construct transgenic fish lines including: 1. A fish line with a functional SPH site and 2. a fish line with a mutagenized SPH site within a target promoter that is regulated by ZNF143. SPHmut lines contain key conserved residues mutagenized rendering them unable to bind ZNF143. B. Method of transgenesis construction. The transgene to be incorporated into the genome is contained on the plasmid above. Both the plasmid and synthesized *tol2* mRNA are co-microinjected into a one cell staged embryo. Embryos are screened for transient fluorescence and raised to adulthood. Founder fish containing the transgene are validated through a PCR analysis. Resulting positive founder fish are outcrossed with a wild-type fish to establish the F_1 generation to be used in ChIP assays.

There are some limitations underlying this approach described in Fig.13 to study regulation by ZNF143. Firstly, the transposition of transgenes via *tol2* transposition is not a targeted insertion into the genome. *Tol2* is able to integrate anywhere in the genome with no preferences to a position on a chromosome [119]. The transgenic lines containing and lacking a functional SPH binding site may have integration events within different genetic contexts, such as within silent or enhancer regions. Therefore, multiple germline founder lines are necessary to establish consistency among significant results.

Secondly, this scope is limited to regulation at target promoters. A candidate promoter approach is not a representative approach for drawing global conclusions of how ZNF143 activates eukaryotic transcription. However, the mechanism of ZNF143 activation at target promoter regions is still a critical steppingstone towards learning the molecular mechanism of ZNF143 and could still shed light on how site-specific activator proteins regulate target genes.

Results

In order to study the mechanism of activation by ZNF143 at target promoters, first candidate genes regulated by ZNF143 had to be selected for study. Due to bipartite activation by ZNF143 at mRNA genes or small RNA genes transcribed by either RNAPII or RNAPIII, a candidate for both gene types were pursued for future study.

Additionally, there has been one direct protein interaction discovered with ZNF143, as previously described, with the CHD8 chromatin remodeler protein[80]. These proteins co-localize together and interact in co-immunoprecipitation experiments[80]. Past research has suggested that CHD8, especially the short isoform of CHD8, is a robust co-activator along with ZNF143[81]. Due to the direct protein interaction and potent co-activation of these two factors in transfection assays, a mRNA gene candidate and a small RNA gene candidate that is co-regulated by both ZNF143 and CHD8 were chosen for analysis. This includes a total of three gene candidates: 1. a mRNA encoding gene activated by ZNF143, 2. Small RNA encoding gene co-activated by ZNF143 and CHD8, 3. A mRNA gene co-activated by ZNF143 and CHD8.

Gene candidates for transgenesis were initially identified from preliminary ChIP-seq experiments performed in 24hpf zebrafish embryos (Fig. 14a). These are considered to be preliminary because they were not repeated, and the number of sequence reads for the ZNF143 ChIP-seq were not sufficient to identify all ZNF143 binding sites. ZNF143 binds to its consensus that is usually located within a couple hundred base pairs of the transcription start site[20]. Upon inspection, good candidate genes show enrichment of ZNF143 surrounding the TSS in the preliminary ChIP-seq. If enriched, candidate gene sequences are then inspected for a putative SPH site. 831 peaks were identified by ZNF143 ChIP-seq, while 1791 peaks were identified by CHD8 ChIP-seq. By visual inspection of gene scans, the ZNF143 results appeared to be of much higher quality with improved signal-to-noise comparison. Binding of ZNF143 at several candidate promoter regions is verified through a ChIP-PCR analysis (Fig. 14B). Moreover, several genes displayed binding enrichment for ZNF143(Fig. 14B). A good candidate gene that binds both ZNF143 and CHD8 in the preliminary ChIP-seq analysis is shown in Fig. 15. Three candidate genes with robust binding of ZNF143, or ZNF143 and CHD8, were selected for an analysis of activation by ZNF143 (Fig. 16). Both the *atic* (Fig. 15) and *u4* (data not shown) genes bind both ZNF143 and CHD8.

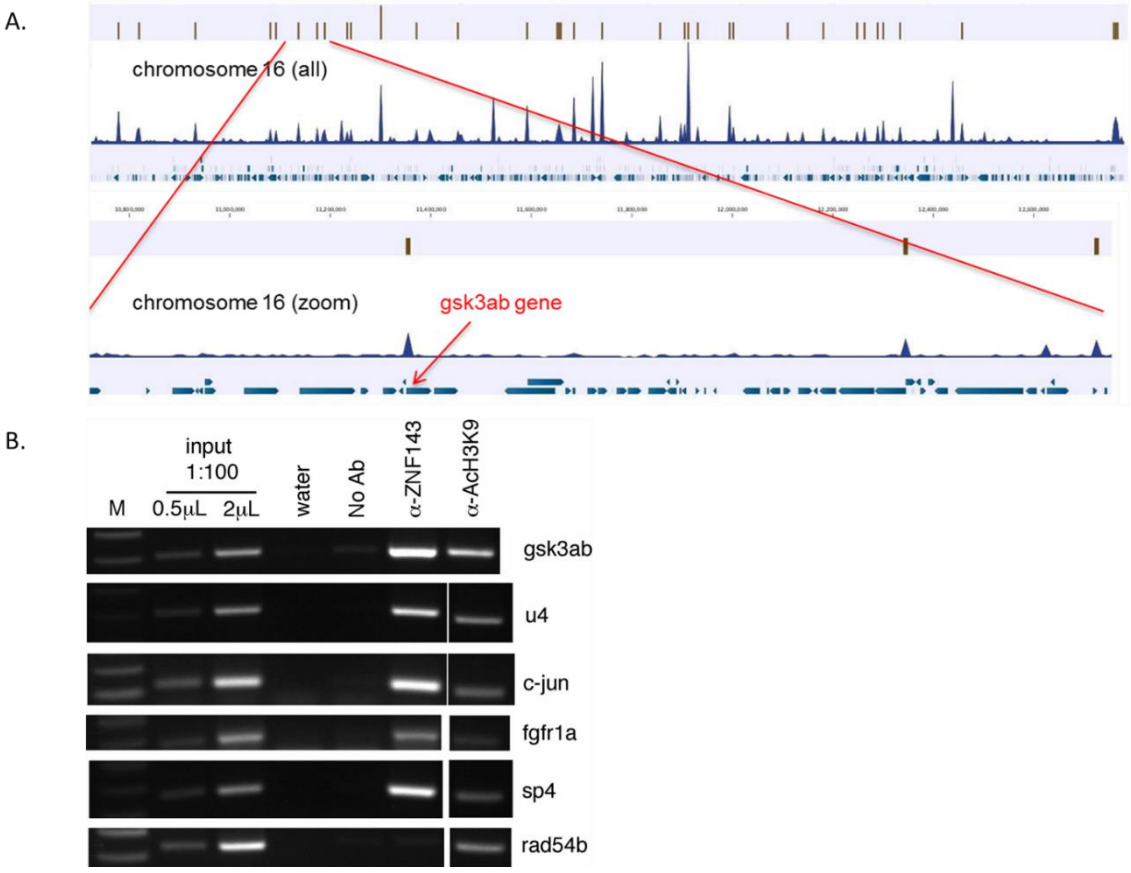


Figure 14. Preliminary determination for the binding of ZNF143 to candidate genes.
A.The binding profile of ZNF143 in 24hpf zebrafish embryos as measured by ChIP-seq. Preliminary genes were first selected through a ChIP-seq analysis. Significant reads are indicated by the vertical dashes and occur within 500 nucleotides from transcriptional start sites.
B. Validation of ChIP-seq ZNF143 gene targets in ZF4 cells. Selected example target genes were tested for binding by ZNF143. 24hpf embryo extracts were immunoprecipitated by α-ZNF143 or α-AcH3k9 as a positive control, detected through PCR, and electrophoresed on 1.5% agarose gels.

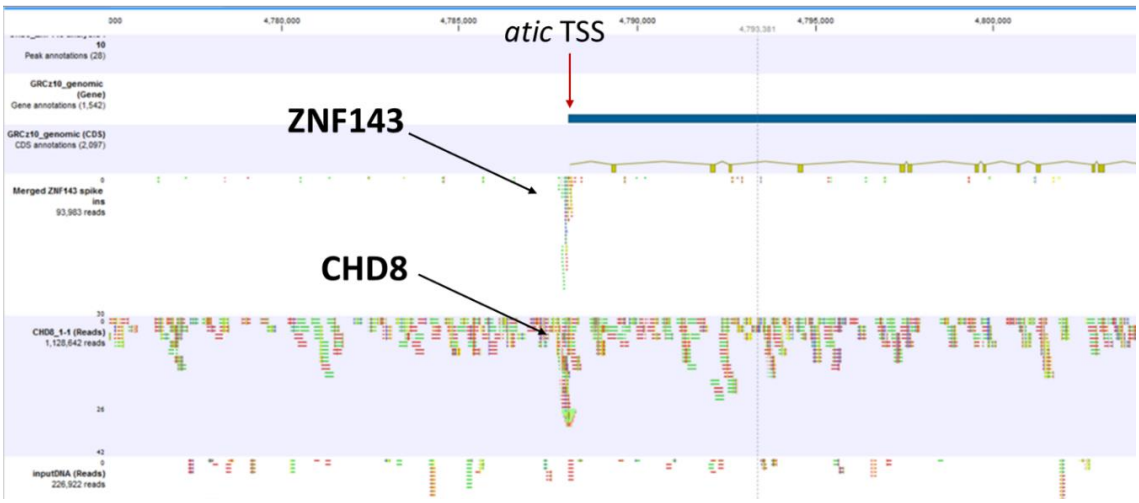


Figure 15. Example of a candidate gene co-bound by ZNF143 and CHD8.
Gene tracks for the *atic* promoter are displayed a preliminary ChIP-seq analysis. The solid blue line represents the *atic* gene. Colored dashes located underneath the gene demonstrate ZNF143 and CHD8 enrichment.

Once candidate genes were selected for robust binding in ChIP analyses (Fig. 16), they were tested for responsive activation by ZNF143 in a transient transfection luciferase assay in zebrafish ZF4 cells. In this assay, the candidate promoter that binds ZNF143 is cloned upstream of a luciferase reporter gene within a plasmid. A second effector plasmid with cDNA encoding for ZNF143 is co-transfected into cells to assess activation of ZNF143 on the target promoter driving luciferase. The amount of luciferase activity is directly controlled by transcriptional efficiency of the regulatory elements located within the promoter. Experiments focus on the SPH element that binds ZNF143. A model transgenic candidate gene displays both 1. Activation when a ZNF143 effector plasmid is transfected into cells containing the luciferase reporter plasmid with a native SPH site and 2. A reduction of luciferase activity when cells are transfected with the

reporter plasmid containing the mutagenized SPH element, whether or not ZNF143 levels are increased by co-transfection of the *znf143* effector plasmid. An acceptable gene candidate for co-activation by ZNF143 and CHD8 would have an increased amount of luciferase activity not only when the ZNF143 effector plasmid is added, but a further increase when an effector plasmid encoding CHD8s is added.

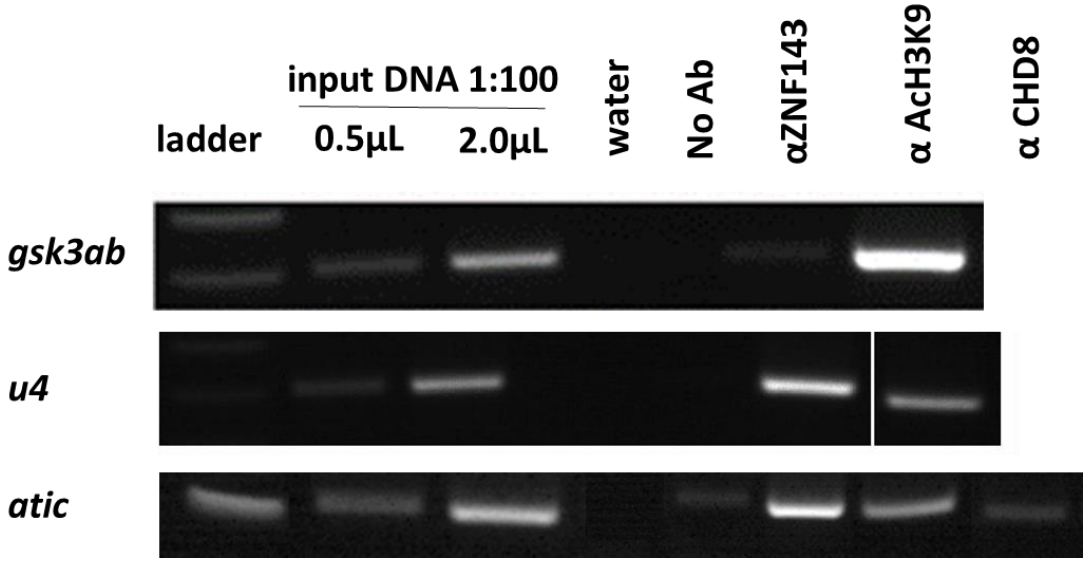


Figure 16. Transgenic gene candidates bind ZNF143.
ChIP-PCR analyses displaying binding of ZNF143 at the *gsk3ab*, *u4*, and *atic* promoter candidates in 24hpf embryos were performed to verify preliminary ChIP-seq binding. The *atic* gene also displays binding of CHD8. AcH3K9 antibody is shown as a positive control.

A small RNA encoding gene chosen for a transient transfection luciferase activation analysis by ZNF143 was the *u4* gene (Fig. 17). The *u4* gene encodes for the non-coding *u4* small nuclear RNA that is essential for pre-mRNA splicing in metazoans [120]. The distal region (DSE) of the *u4* gene contains an SPH element located a couple hundred base pairs upstream of the TSS[121]. Upon transfection of ZF4 cells with the

u4:luciferaseSPHmut construct, fold luciferase expression is drastically decreased when compared to *u4:luciferase* containing the SPH site (Fig. 17). In order to investigate the responsiveness of the *u4* promoter to ZNF143, a synthetic promoter containing 5 SPH sites was engineered. This included a deletion of DNA upstream of the PSE with the addition of multiple SPH sites closer to the TSS. Previous experiments in the lab with the human *u6* gene demonstrated that OCT and SPH motifs needed to be positioned close to the PSE in order to cause strong activating effects in such transfection experiments. When ZNF143 was added to the synthetic 5xSPH *u4* promoter, a marked increase in activation was able to be detected when compared to the synthetic *u4:luciferase* as a reference (Fig. 17). Co-activation by CHD8 was assessed by transfecting in plasmids encoding for either the long or the short isoform. Similarly to previously published experiments[81], CHD8s lacking the helicase domain of the full length CHD8 was a modest co-activator of ZNF143 at the *u4* promoter (Fig. 17). Interestingly, the CHD8 full length isoform did not significantly increase fold activation (Fig. 17). Nevertheless, the *u4* promoter displayed binding by ZNF143 (Fig. 16) and displayed co-activation by ZNF143 and CHD8 (Fig. 17). Because of these reasons, the *u4* promoter was used as a gene candidate for transgenesis.

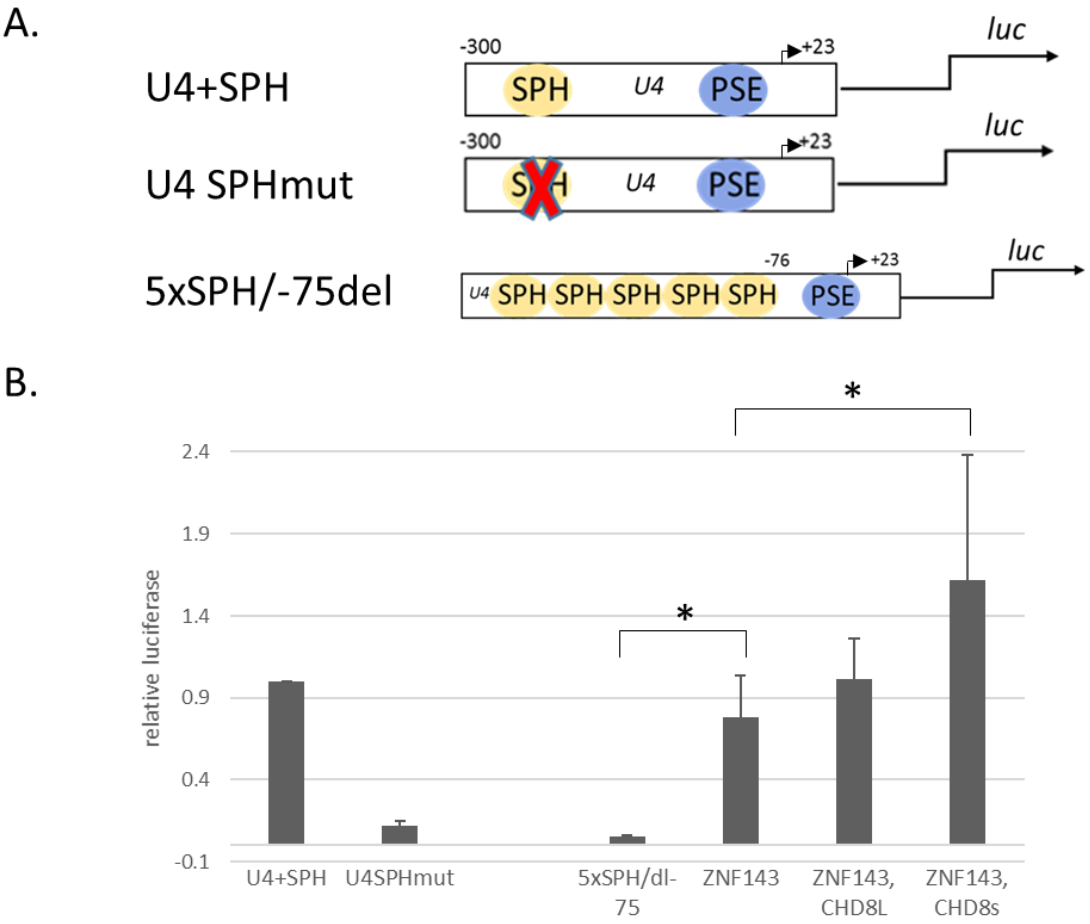


Figure 17. Co-activation by ZNF143 and CHD8s of the *u4* promoter.
A. Genetic constructs of the luciferase assay. The U4+SPH construct contains between -300 to +23nt of the native u4 promoter. The SPHmut construct contains between -300 to +23 of the native u4 with key residues of the SPH element mutated. The 5xSPH/-75del retains -76 to +23nt of the native U4 promoter with the addition of 5xSPH elements added on the 5' end of the gene.
B. The first two columns depict ZF4 cells that were transfected with the U4+SPH or U4 SPHmut constructs driving a firefly luciferase reporter as depicted in A. Fold luciferase activity of the U4SPHmut is relative to the U4+SPH construct. The four bars to the right depict ZF4 cells transfected with the synthetic 5xSPH/-75del construct driving synthetic firefly luciferase. ZNF143, CHD8L, or CHD8s represent the co-transfection of an effector plasmid encoding cDNA for ZNF143, CHD8L, and/or CHD8s respectively. Asterisks indicate $p<0.05$.

For an mRNA encoding gene candidate, the *5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase* (*atic*) gene was chosen for

transcriptional analysis to assess co-activation by CHD8 and ZNF143. The *atic* gene encodes for a protein involved in the catalysis of the last two steps of the *de novo* purine biosynthetic pathway [122]. When ZNF143 was co-transfected, the amount of luciferase expression increased 2-fold when compared to the native *atic:luciferase* as a reference (Fig. 18). Additionally, there was a dramatic loss of expression upon transfection of the mutagenized SPH construct (Fig. 18). This demonstrates the importance of site-specific regulation by ZNF143 for expression above a basal level at the *atic* promoter. Regarding co-activation by the CHD8s isoform, there was only a non-significant modest increase in activation when compared to activation only by ZNF143 (Fig. 18). This further increase in fold expression could represent a co-activation by CHD8s and was pursued as a transgenic gene candidate.

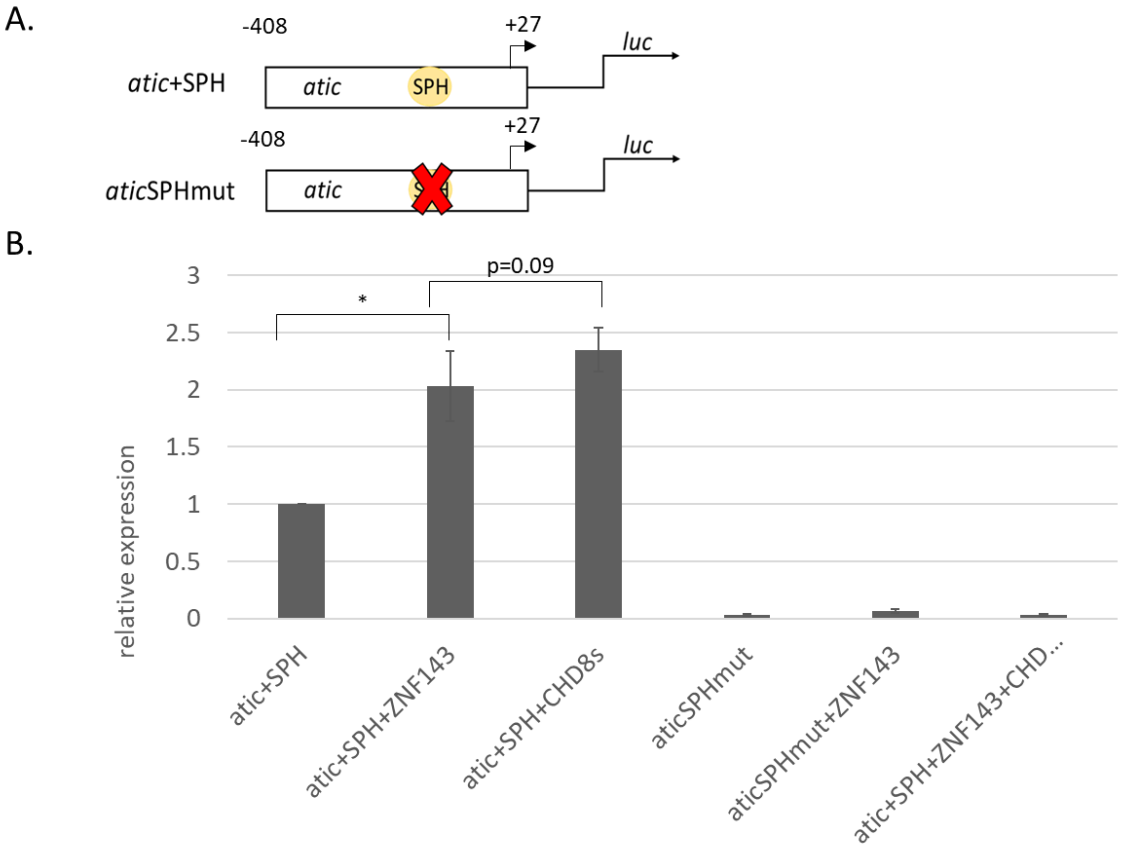


Figure 18. Co-activation by ZNF143 and CHD8s of the *atic* promoter.
A. Genetic constructs of the luciferase assay. The atic+SPH construct contains between -408nt +27nt of the native *atic* promoter driving a firefly luciferase reporter. The aticSPHmut construct contains the same fragment of the *atic* promoter with key residues of the SPH element mutagenized.
B. The first three columns depict ZF4 cells that were transfected with the atic+SPH constructs driving a firefly luciferase reporter outlined in A. The next three columns depict ZF4 cells transfected with the aticSPHmut construct driving synthetic firefly luciferase in A. ZNF143, and CHD8s represent the co-transfection of an effector plasmid encoding cDNA for ZNF143 and/or CHD8s respectively. Asterisks indicate $p<0.05$.

Finally, the third candidate promoter of an mRNA encoding gene that is regulated by ZNF143 is the *glycogen synthase kinase-3 alpha b* gene (*gsk3ab*). The *gsk3ab* gene encodes for the alpha isoform of a serine threonine kinase implicated in

metabolism and energy storage[123]. The highly homologous beta isoform, *gsk3b*, is an important regulator of the Wnt and PI3 kinase signal transduction pathways[124]. Upon the addition of ZNF143, fold luciferase expression increased 1.7 fold relative to a *gsk3ab*:*luciferase* reference (Fig. 19). Distinct from *u4* and *atic* gene candidates, the addition of CHD8s to *gsk3ab:luciferase* did not increase luciferase expression (Fig. 19), indicating a lack of co-activation by CHD8s. Additionally, no binding peak was observed for CHD8 at the *gsk3ab* gene promoter in the preliminary ChIP-seq (data not shown). As expected, transfection with the *gsk3ab:luciferaseSPHmut* construct significantly decreased in luciferase expression when compared to the *gsk3ab:luciferase* reference (Fig. 19). These findings implicate the *gsk3ab* promoter as a suitable gene candidate for a mRNA encoding gene regulated by ZNF143.

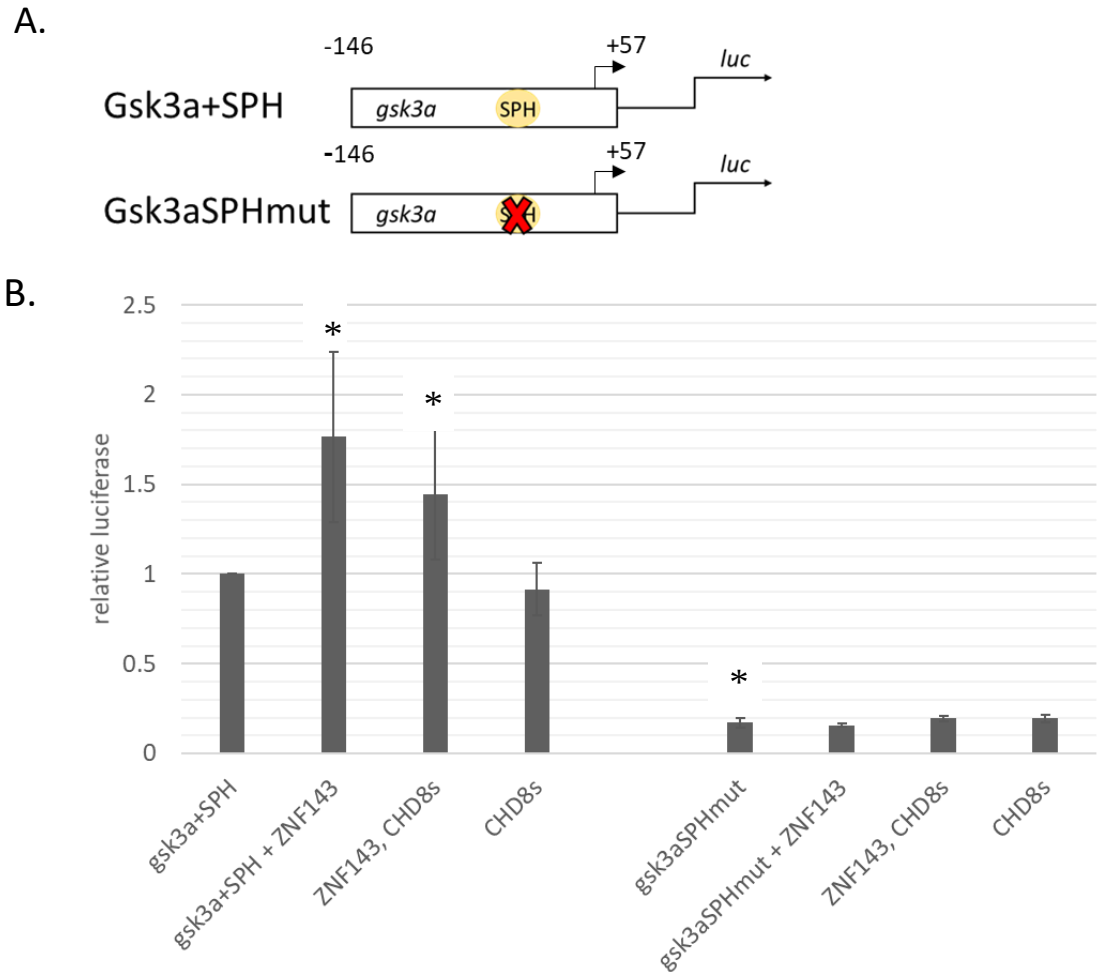


Figure 19. Activation of the *gsk3ab* gene by ZNF143.
A.Genetic constructs of the luciferase assay. The gsk3a+SPH construct contains between -146 to +57 nt of the native *gsk3ab* promoter driving a firefly luciferase reporter. The gsk3aSPHmut construct contains the same fragment of the *gsk3ab* promoter with key residues of the SPH element mutagenized.
B. The first four columns depict ZF4 cells that were transfected with the gsk3a+SPH construct driving a firefly luciferase reporter as depicted in A. The next four columns depict ZF4 cells transfected with the gsk3aSPHmut construct driving synthetic firefly luciferase in A. ZNF143, and CHD8s represent the co-transfection of an effector plasmid encoding cDNA for ZNF143 and/or CHD8s respectively. Asterisks indicate $p<0.05$.

Once all three gene candidates were chosen, transgenic fish were constructed via *tol2* transgenesis as outlined in Figure 12. Founders were screened for fluorescence and the presence of a transgene was detected in a PCR analysis from genomic DNA isolated

from F_1 embryos. Transgenic founders for the *atic* and *gsk3ab* promoters were verified via a PCR analysis in Fig. 20.

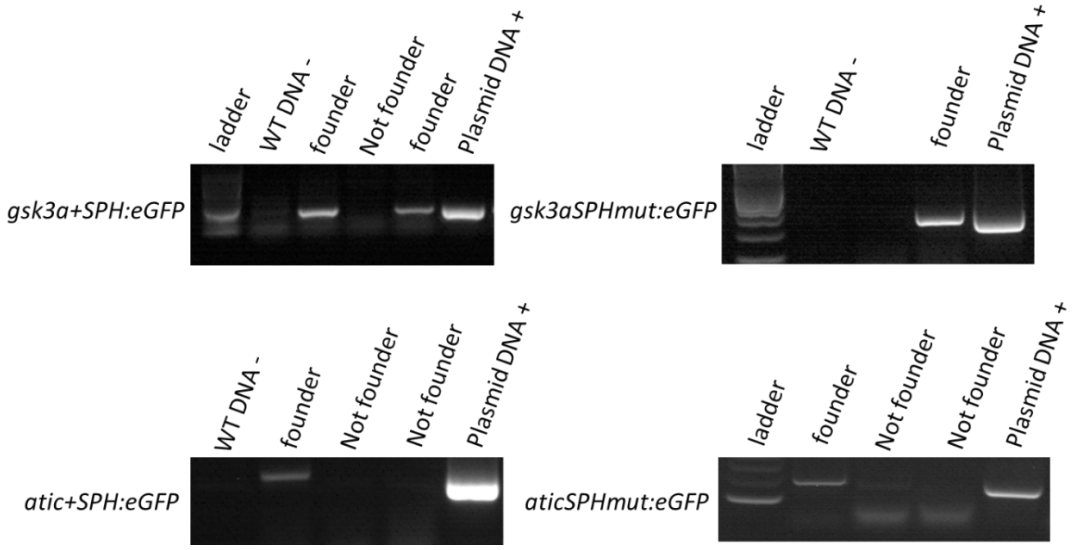


Figure 20. Verification of founders via PCR.
PCR products were amplified using a primer set that included a forward primer in the target promoter region with a reverse primer in the flanking *eGFP* sequence. Genomic DNA obtained from 24hpf embryos was used as a negative control. A plasmid containing the transgene was used as a positive control.

Founders for the *gsk3ab* promoter containing the SPH motif displayed strong fluorescence in in the brain (Fig. 21). This included the forebrain, midbrain, and hindbrain (Fig. 21). Additionally, *gsk3a+SPH:eGFP* founders contained fluorescence in the heart (Fig. 22). The expressed *gsk3ab+SPH:eGFP* protein or RNA is maternally deposited into the embryo, as ubiquitous fluorescence is expressed detected at 1.5 hours past fertilization (Fig. 23). As development of the embryo continues to 24hpf and 39hpf stages, fluorescence in brain structures become more apparent over a background axial expression (Fig. 23). *Gsk3abSPHmut:eGFP* founders had a decreased fluorescence when

compared to *gsk3ab* containing the SPH motif (Fig. 20). This points to likely regulation by ZNF143 in driving different expression levels of fluorescence between the two lines at the *gsk3ab* promoter.

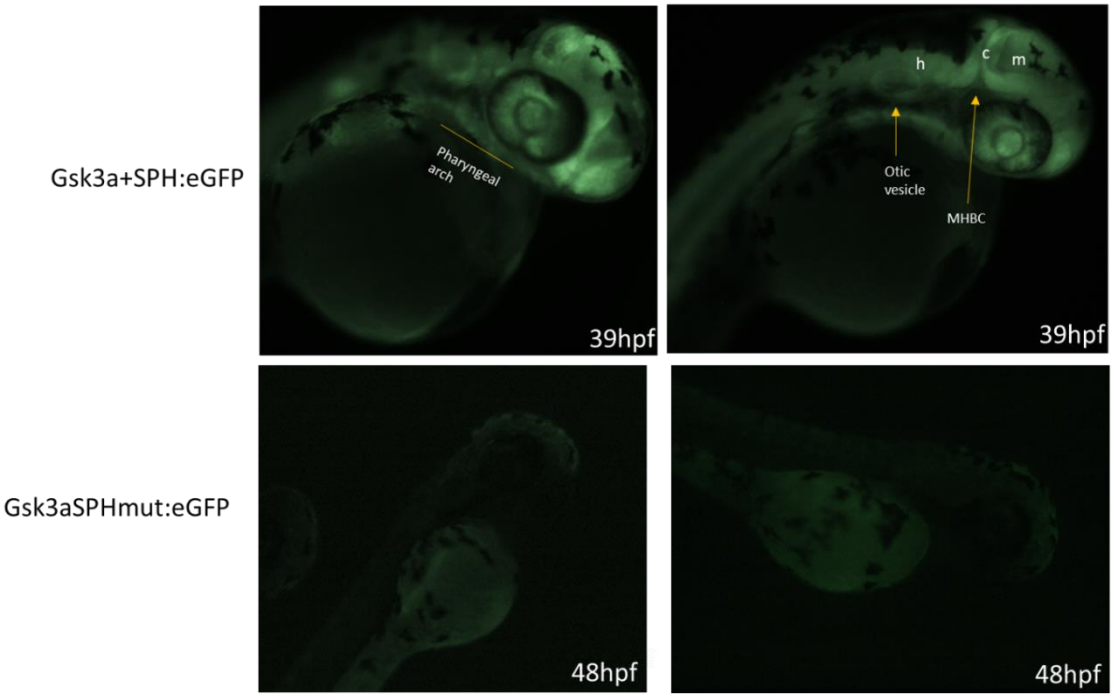


Figure 21. Fluorescence patterns of *gsk3ab+SPH:eGFP* and *gsk3abSPHmut:eGFP*. Fluorescence images for *gsk3a+SPH:eGFP* and *gsk3aSPHmut:eGFP* in zebrafish embryos at 39hpf and 48hpf developmental stages respectively. m=midbrain, c=cerebellum, h=hindbrain, MHBC= midbrain-hindbrain boundary constriction.

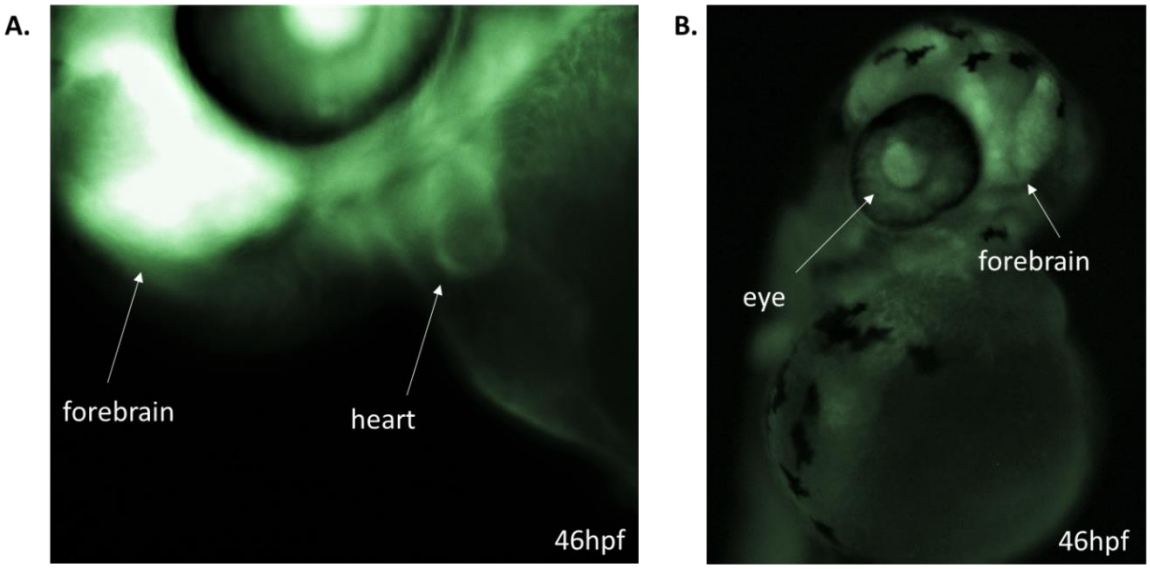


Figure 22. Gsk3a+SPH:eGFP zebrafish contain strong eGFP expression in the brain and heart.
Fluorescence images of *gsk3a+SPH:eGFP* raised to 46hpf. A. Zoomed lateral image of the forebrain, eye, heart, and yolk. B. Ventral view of the head, eye, and forebrain.

Founders of the *atic+SPH:eGFP* transgene displayed a non-specific pervasive fluorescence throughout the developing embryo (Fig. 23). The accompanying *aticSPHmut:eGFP* line lacked fluorescence similarly to the *gsk3abSPHmut:eGFP* line in Fig. 20 (results not shown). Founders for the *u4* were identified via PCR analysis but did not display fluorescence (results not shown). For this reason, *u4* was discontinued from further analysis.

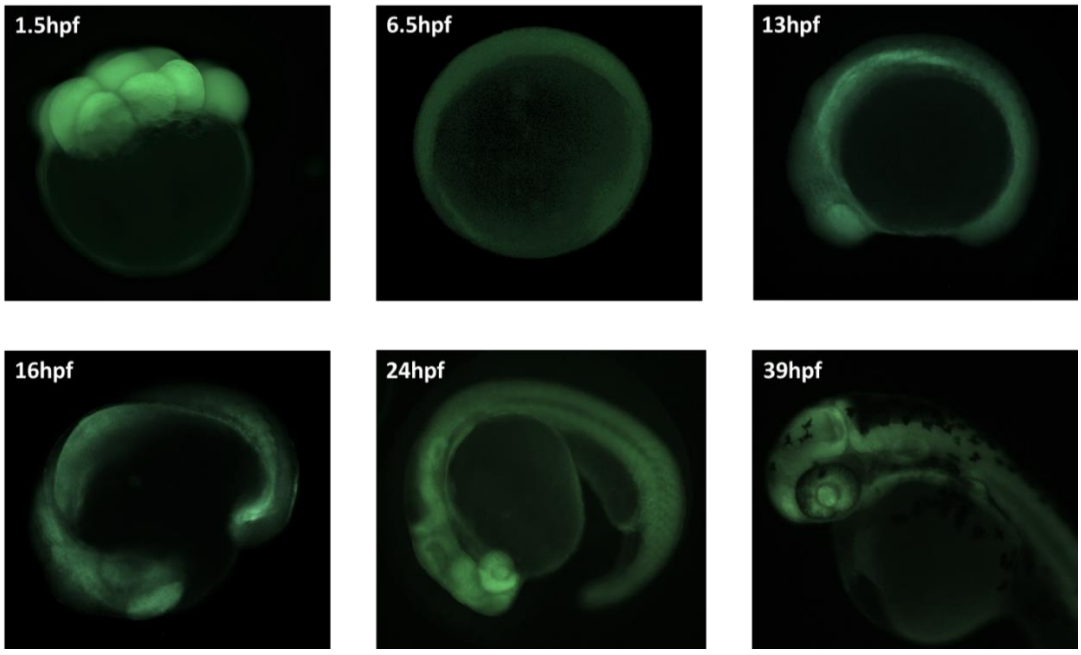


Figure 23. Developmental time course *gsk3ab+SPH:eGFP*.
Fluorescence images of *gsk3ab+SPH:eGFP* in zebrafish embryos at varying stages between 1.5hpf and 39hpf.

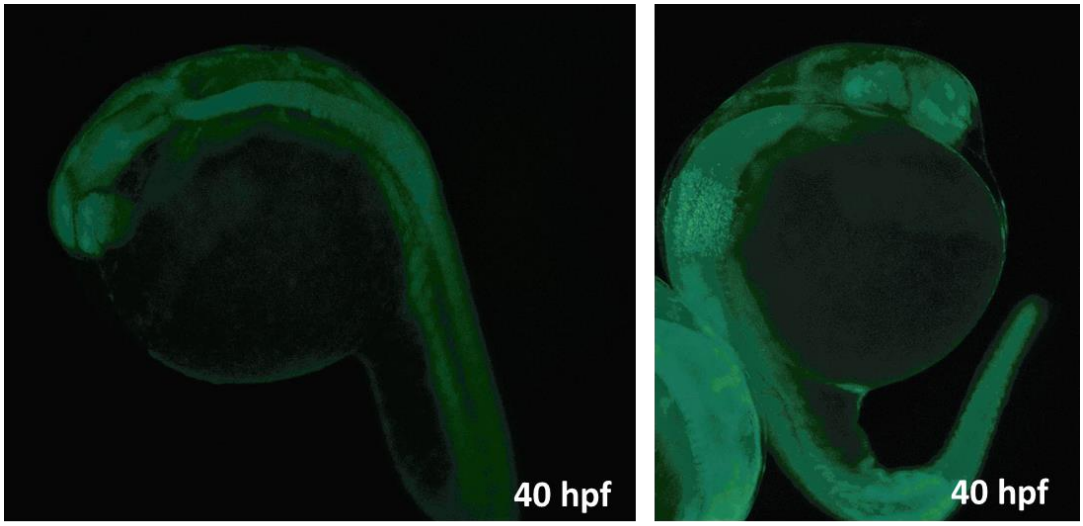


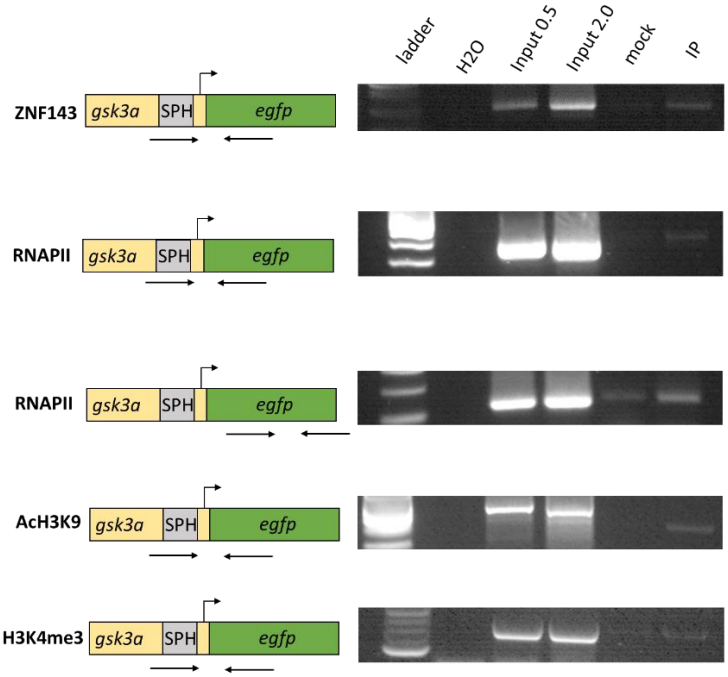
Figure 24. Fluorescence images of *atic+SPH:eGFP*.
Fluorescence images displaying a ubiquitous expression of *atic+SPH:eGFP* in zebrafish embryos at 40hpf.

After the establishment of a transgenic line, chromatin immunoprecipitation (ChIP) experiments were performed to probe for distinguishing features of promoter regions regulated by ZNF143. Before exploring the possibility of any enriched epigenetic modifications within transgenic promoter regions, the presence of ZNF143 was first determined. It was postulated that transgenic constructs containing SPH sites would likely bind ZNF143, whereas transgenic line constructs with a mutated SPH site would not. This verification of ZNF143 binding is critical for drawing any later conclusions about the enrichment of RNA polymerase and epigenetic markings.

While it is expected that a higher enrichment of RNAPII would be present at transgenes containing SPH sites when compared to transgenes with mutagenized SPH sites, the amount of RNAPII enrichment along different regions of the transgene can provide mechanistic information regarding where polymerase is being recruited or retained on the transgene. The ratio of RNAPII enrichments with primers encompassing the promoter region compared to the amount of RNAPII enriched downstream within the gene body can be compared between transgenic lines with the SPH binding site and transgenes containing a mutagenized SPH site. A transgene with an SPH site containing a higher ratio of RNAPII enriched within TSS to RNAPII enriched within the gene body when compared to a transgene with a mutagenized site could be indicative of ZNF143 aiding transcription initiation. Alternatively, a higher RNAPII gene body enrichment to RNAPII at TSS enrichment ratio at the transgene containing an SPH site could suggest that ZNF143 is aiding in transcription elongation. Delineating between these possibilities could help narrow the step of transcription that ZNF143 assists.

There has been evidence of enrichment of various active epigenetic modifications at genes bound by ZNF143. ZNF143 is enriched in regions of chromatin that harbor active histone epigenetic modifications [17], and recently has been associated with a reduction in H3K9me3[85]. Due to these associations, chromatin immunoprecipitations were performed to identify the presence of epigenetic modifications at promoters regulated by ZNF143, to determine whether these enrichments were reduced upon mutagenesis of the SPH site. As expected, the *gsk3ab+SPH:eGFP* bound ZNF143 in the presence of an SPH site (Fig. 25). RNAPII was enriched both in the region encompassing the TSS, although the high background level did not support statistical significance, and along the body of the *eGFP* reporter (Fig. 25). Furthermore, *gsk3ab+SPH:eGFP* had enrichment of two active epigenetic modifications, H3K9ac and H3K4me3 (Fig. 25). Enrichment of a typical repressive modification, H3K9me3, was not able to be detected at *gsk3ab+SPH:eGFP* (results not shown).

A.



B.

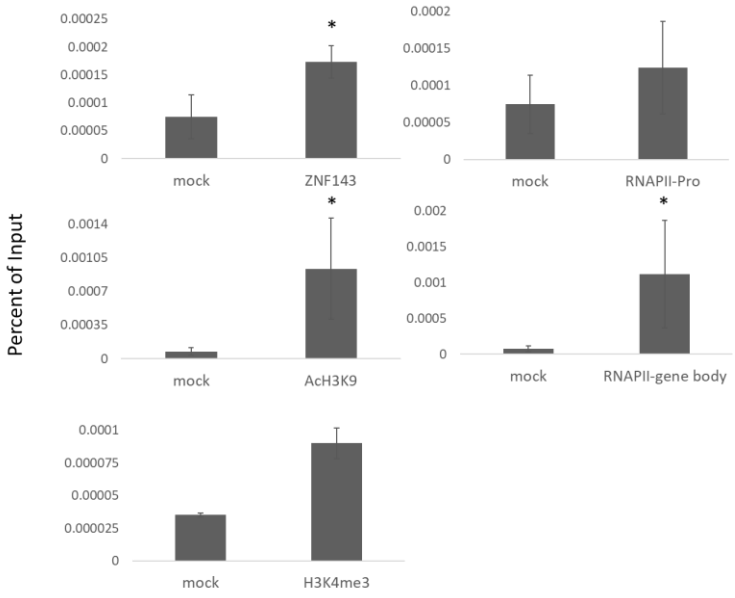


Figure 25. ChIP analysis of 24hpf *gsk3ab+SPH:eGFP embryos*.

A. ChIP was performed in 24hpf embryos with the antibody denoted. Primer sets that were used are characterized underneath the transgene. Lane 2 represents a water control for the PCR analysis. The mock lane represents a no antibody control added to chromatin. B. ChIP-qPCR analysis was performed with the results normalized to the % input method. Starred enrichments have $p<0.05$.

Regions of euchromatin and heterochromatin have associated repertoires of epigenetic modifications that mark promoter regions [125]. One developmentally regulated class of genes in zebrafish contain enrichments of classically active marks, such as H3K9ac, H4ac, and H3K4me3, and lacks transcriptionally repressing marks such as H3K9me3 or H3K27me3[125]. The *gsk3ab* gene has been shown to function in an array of cellular functions including cellular differentiation and growth[126]. The promoter of *gsk3ab+SPH:eGFP* might be representative of this class of promoter as it is enrichedwith H3K9ac and H3K4me3, and lacking H3K9me3.

During the process of raising embryos of F_1 fish line to one month of age, *gsk3abSPHmut:eGFP* have had few fish that survived to adulthood. Only low numbers of *gsk3abSPHmut:eGFP* F_1 embryos were produced. While some chromatin has been obtained for this transgenic line, there is around a 50% mortality rate in embryos surviving to 24hpf (results not shown). One ChIP replicate for the *gsk3aSPHmut:eGFP* transgenic line using ZNF143 and RNAPII antibodies is shown in Fig. 26.

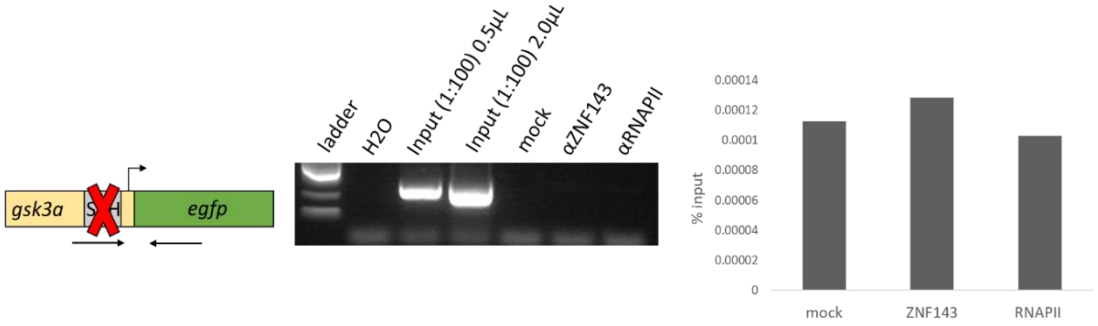


Figure 26. ChIP analysis of ZNF143 and RNAPII on gsk3abSPHmut:eGFP. ChIP analysis performed in 24hpf embryos with ZNF143 and RNAPII antibodies. Primer sets that were used are characterized underneath the transgene. Lane 2 represents a water control for the PCR analysis. The mock lane represents a no antibody control added to chromatin. A ChIP-qPCR analysis was performed with results normalized to the % input method.

ChIP experiments on *gsk3aSPHmut:eGFP* demonstrate a loss of ZNF143 and RNAPII binding within the region encompassing the TSS (Fig. 26). This loss is shown by the relative signal to mock ratio. The relative signal to mock ratio for ZNF143 and RNAPII are 2.33 and 1.66 respectively for *gsk3a+SPH:eGFP*. This ratio decreases to 1.13 for ZNF143 and 0.91 for RNAPII in single replicate experiments for *gsk3aSPHmut:eGFP*. Future ChIP experiments would need to be done to verify these results. Additionally, future ChIP experiments with antibodies for histone modifications such as H3K4me3, H3K9ac, and H3K9me3 would be performed.

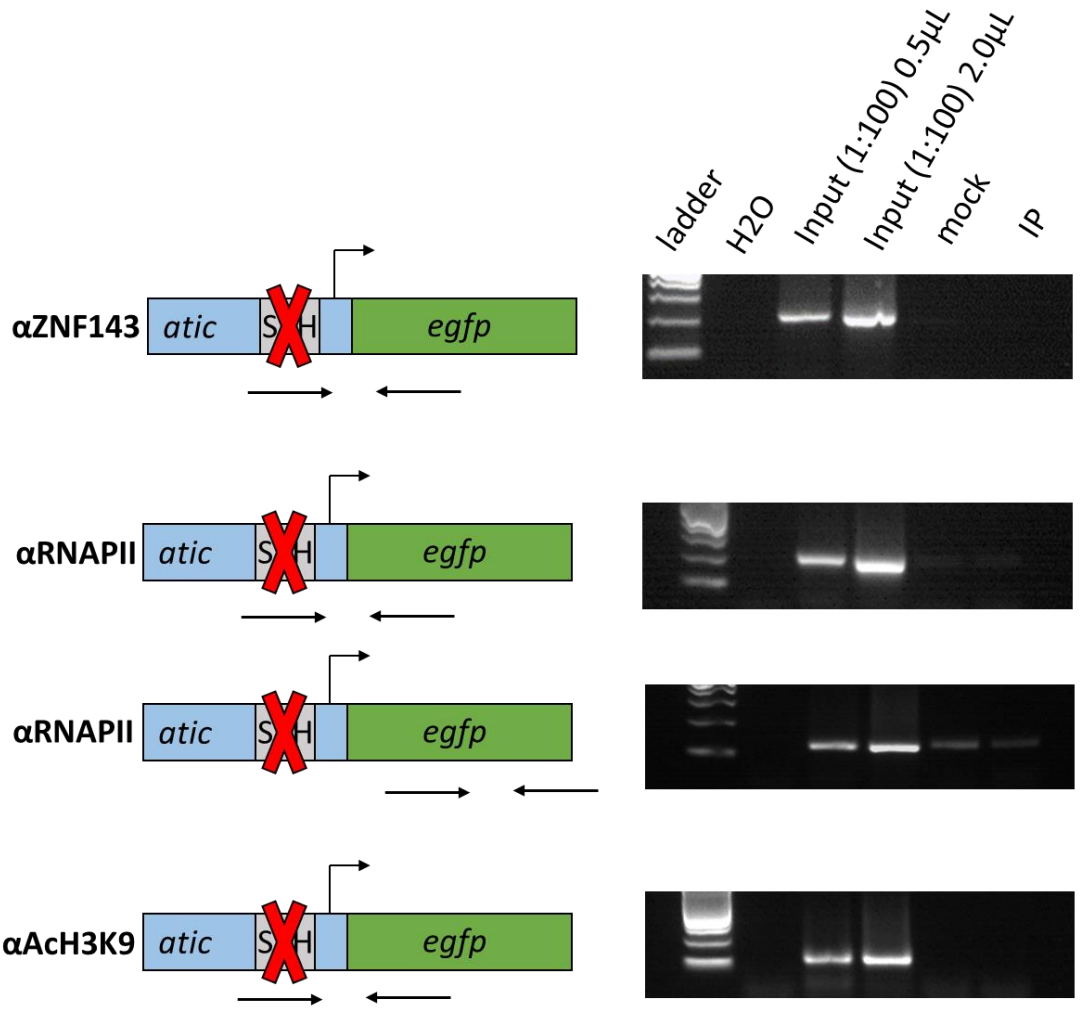


Figure 27. ChIP analysis of the *aticSPHmut:eGFP* transgenic line.
ChIP was performed in 24hpf embryos with the antibody denoted. Primer sets that were used are characterized underneath the transgene. Lane 2 represents a water control for the PCR analysis. The mock lane represents a no antibody control added to chromatin.

Preliminary ChIP experiments performed on *aticSPHmut:eGFP* did not show significant enrichments of ZNF143, RNAPII, or AcH3K9 (Fig. 27). This suggests that a mutagenized SPH site has disrupted binding of ZNF143 at this promoter. While hard to draw conclusions without comparison to *atic+SPH:eGFP*, these preliminary ChIP

experiments performed on *aticSPHmut:eGFP* corroborate results from the *gsk3aSPHmut:eGFP* line. These both suggest that a mutagenized SPH site leads to a downstream loss of RNAPII. Furthermore, the *aticSPHmut:eGFP* ChIP experiments suggest that this loss may be correlated with a loss of AcH3K9.

Discussion

Promoter regions of eukaryotic genes receive combinatorial signals from trans-acting factors that direct complex programs of gene expression. Activator proteins partake in a key role of this process through increasing gene expression above a basal level. While there are general mechanisms established regarding transcriptional activation, the mechanisms of many activators are still unknown. One such activator protein is ZNF143 that regulates at least 2000 promoter regions in mammals[19] and is critical for many biological processes.

ZNF143 binds an array of different genes including both small RNA encoding and mRNA encoding genes transcribed by RNAPIII or RNAPII [105]. Here a candidate promoter approach was implemented with the goal of discovering features of promoter regions targeted for activation by ZNF143.

A search for candidate promoters was carried out with the goal of finding a mRNA gene regulated by ZNF143, a mRNA gene co-regulated by ZNF143 and CHD8, and lastly a small RNA gene regulated by ZNF143 and CHD8. Each of these genes were subjected to binding and activation analyses by ZNF143, or both ZNF143 and CHD8. The *u4* gene, essential for pre-mRNA splicing[120], was demonstrated to be a justifiable

small RNA candidate promoter region that binds both ZNF143 and CHD8. In addition, a synthetic *u4* promoter containing 5 SPH elements driving firefly luciferase displayed strong activation by ZNF143. The *atic* and *gsk3ab* genes were chosen as mRNA encoding genes that bind ZNF143. Unlike *gsk3ab,* that demonstrated activation solely by ZNF143, the *atic* promoter was shown to be co-activated by both ZNF143 and CHD8 in luciferase assays. These three candidate promoter regions were chosen for further analysis.

One aspect of activation by ZNF143 neglected in this analysis is a small RNA gene transcribed by RNAPIII. ZNF143 transcribes several gene types transcribed by RNAPIII including the *U1*, *U4C*, *U6*, *Y4*, *7SK*, and *H1 RNA* genes[11, 12, 31-36]. While the *u4* gene was pursued as a small RNA gene candidate, it would be an example of a promoter type still transcribed by RNAPII. The mechanism by which ZNF143 activates genes transcribed by RNAPIII remains unknown.

Both mRNA gene encoding promoter regions regulated by ZNF143, *gsk3ab* and *atic*, directed strong expression of *eGFP* within developing embryos. These transgenic mRNA or protein products are maternally deposited into the embryo of zebrafish and direct a strong expression of *eGFP* throughout development. The *atic* gene, involved in *de novo* purine biosynthesis [122], is reported to be expressed spatially within neural structures, digestive system, and yolk in developing zebrafish through in-situ hybridization analyses [100]. While an accentuation of fluorescence within neural structures was observed in *atic+SPH:eGFP* founders, there is ubiquitous background expression present throughout development until 48hpf in transgenic embryos. It is

likely that the *atic* promoter used in transgenesis in this study may be lacking enhancer(s) necessary for tissue-specific expression, as it only contained over 400bp of 5' flanking sequence. Conversely, the *gsk3ab* promoter used in this study contained over 1kb of 5' flanking sequence and may contain necessary enhancer regions. The spatial and tissue-specific expression pattern of *gsk3ab* has not been characterized in zebrafish [100]. Here the expression of *gsk3ab+SPH:eGFP* is pervasive all throughout early development of the embryo, and becomes more distinct in brain structures nearing 24hpf.

Founders containing a mutagenized SPH element showed a marked decrease in fluorescence during early development in zebrafish. Furthermore, preliminary ChIP experiments performed on the *gsk3abSPHmut:eGFP* and *aticSPHmut:eGFP* lines demonstrate a reduction of both ZNF143 and RNAPII binding encompassing the TSS region of *eGFP*. ChIP analyses were not performed on the *atic+SPH:eGFP* line due to the loss of reproduction among the founder line. Due to this, the F_1 generation was not able to be obtained. Additionally, *aticSPHmut:eGFP* demonstrated a loss of AcH3K9. Future ChIP-qPCR experiments need to be performed to support reductions of ZNF143 and RNAPII, and ChIP experiments need to be performed on *gks3aSPHmut:eGFP* to investigate if AcH3K9 enrichment is also reduced at this promoter. This preliminary result suggests that a mutated SPH element reduced ZNF143 recruitment and downstream RNAPII binding. A reduction of enrichment of ZNF143 and RNAPII relative to transgenic constructs lacking SPH elements indicates that ZNF143 is increasing expression above a basal level of *gsk3ab* and/or *atic* promoter regions. It is

possible that ZNF143 recruits histone acetyltransferases to promoter regions to make them more permissible to transcription. However, it remains unknown if histone modifications such as AcH3K9 are directly correlated to activation by ZNF143.

One mechanism by which activator proteins can stimulate transcription is by recruiting cofactors that are able to modify chromatin structure[127]. Due to a direct interaction with CHD8[128], and evidence of epigenetic modifications being associated with ZNF143 enriched promoters [17], it is possible that ZNF143 interacts through a co-factor to direct changes within chromatin structure at gene targets. This is still a valid possibility. Through utilizing primers encompassing the region around the TSS of *gsk3a+SPH:eGFP* transgenic lines, permissive modifications including H3K4me3 and AcH3K9 were detected. Although there is a slight chance that multiple founder transgenes inserted into actively transcribed regions of chromatin, this suggests that the endogenous *gsk3ab* likely contains permissive histone modifications within the promoter. Future experiments on *gsk3abSPHmut:eGFP* need to be performed to uncover any reduction of these histone modifications. Conversely, it could be possible that SPHmut transgenes gain silencing epigenetic modifications, such as H3K9me3. This is unlikely because H3K9me3 was not detected at *aticSPHmut:eGFP* (data not shown).

Methods

Plasmid constructions

Transgenesis plasmids for the *gsk3ab*, *atic*, and *u4* genes were constructed using the pT2AW3 vector, a promoter-less version of pT2AL200R150G available from the Kawakami lab [129]. pT2AW3 allows for *tol2* mediated transgenesis in zebrafish.

pT2AW3/atic

Primers amplifying the *atic* promoter insert (-407 to +27) from zebrafish gDNA included an engineered *BamHI* site and *HindIII* site added into the 5'-end and 3'end, respectively. These primer sequences were *atic* Forward: 5' CGGAGGATCCCTGTGGTATCT GTTTGAACA 3', and *atic* Reverse: 5'CGGAAAGCTTACACGTGGATGCAATAG TTT3.' The PCR product was cloned into an intermediary plasmid, pGEM-T (Promega A3600) according to manufacturer's instructions and digested using *BamH1* and *HindIII*. The cut insert and double digested pT2AW3 with *BamH1* and *HindIII* were ligated using T4 ligase (NEB M0202S). Sequences were verified using dideoxy sequencing.

pT2AW3/aticSPHmut was generated from pT2AW3/atic using site-directed mutagenesis with the following primer sequences:

5'TCAAAAACTACAGTTagatctATGCATCTACGCGTA3' and 5' TACGCGTAGATGCATagatctAACTGTAGTTTTTGA 3'. The lower case letters signify nucleotides used to disrupt the SPH motif. This sequence was verified using dideoxy sequencing.

pT2AW3/*u4*

The *u4+SPH* and *u4SPHmut* sequences were amplified out of a previously constructed plasmids (pGL3/zU4#1 and pGL3/zU4SPHmut) using a forward primer with an engineered *EcoRI* on the 5' end, and a reverse primer specific to that vector (GL primer 2). These sequences were *u4* Forward: 5' CGGAGAATTCGTTTTCGCTTTAAAGC 3' and Reverse (GL primer 2): 5'CTTTATGTTTTTGGCGTCTTCCA3'.
This fragment contained an engineered *EcoRI* on the 5' end and a *HindIII* site on the 3' end. These PCR products were cloned into an intermediary pGEM-T plasmid and digested out using *EcoR1* and *HindIII*. These inserts and a double digested pT2AW3 were ligated using a T4 ligase. These plasmids were verified using dideoxy sequencing.

pT2AW3/*gsk3ab*

This plasmid was constructed previously by Courtney Pippit in the Kunkel lab. It contained the zebrafish *gsk3ab* promoter from approximately -1.1kb to +57. The SPH motif in the promoter was disrupted in pT2AW3/gsk3abSPHmut using site-directed mutagenesis. The sequence modified in the SPHmut promoter was TTagatctAAGACAATGC, with lower case letters signifying the exact changes.

Zebrafish husbandry

AB/TL WT zebrafish were maintained using standard methods with protocols approved by Texas A&M University Animal Care and Use Committee (AUP #2016-0102, AUP #2013-0082)

Transgenesis Construction

In vitro transcription of *tol2* mRNA:

Tol2 cDNA included in pCS-TP, supplied by the Lekven laboratory in the Biology Department, TAMU, was linearized with *Not1*. *In vitro* transcription was performed using a MAXIscript SP6 transcription kit (Invitrogen AM1310) following manufacturer's instructions. *Tol2* mRNA was resuspended with 20uL of water following an ethanol precipitation. The size of *tol2* mRNA (2500nt) was verified using agarose gel electrophoresis and the concentration was determined using a Nanodrop spectrophotometer.

Microinjections:

Approximately 1nL of injection mix was injected into one cell stage embryos using the following injection mix:

5.75uL water
2.5uL of appropriate pT2AW3 plasmid DNA (50ng/uL)
0.75uL of *tol2* mRNA (300ng/uL)
1 uL of phenol red stock (2%)

Zebrafish were screened for transient fluorescence and raised to adulthood. Founders were detected through fluorescence and verified via PCR analysis of genomic DNA. Positive founders were outcrossed and raised to adulthood to establish the F_1 generation. Embryos of the F_1 generation were used in subsequent ChIP analyses.

Genomic zebrafish embryo DNA extractions

10-20 embryos were dechorionated manually and transferred to a 1.5mL Eppendorf tube with a transfer pipet. As much liquid was removed as possible. 100uL of extraction

buffer (10mM Tris (pH=8.1), 10mM EDTA, 0.2M NaCl, 0.5% SDS, 0.2mg/ml Proteinase K) was added and incubated at 50C for 3 hours with occasional mixing. The resulting genomic DNA mix was centrifuged for 1 min at top speed at room temperature. The supernatant was transferred to a clean tube and 200uL of 95% ethanol was added and mixed. This was either allowed to sit on ice or transferred to a -20C freezer for 30min. The resulting mixture was centrifuged for 10min at top speed at room temperature. The supernatant was then removed and discarded. 200uL of 70% ethanol was added to wash the DNA pellet and re-centrifuged for 2min at room temperature at max speed. The pellet was then allowed to dry and dissolved in 20uL of TE buffer. This was incubated for at least 30 minutes on ice to allow time for the DNA pellet to resuspend and then stored at -20C.

PCR Genotyping

Genotyping of transgenic founders was performed through a PCR analysis using a primer set that included a forward primer in the promoter of interest and an *eGFP* +60 reverse primer.

Primer sequences:

gsk3ab Forward: 5' CGC ACA ATG CTA TTG CAG AAT ATT G 3'
atic Forward: 5' TAG GAC AAA ACT GAT TTT AAA C 3'
u4 Forward: 5' CGC GTT TCG TCT GAT AGG CAG CC 3'
egfp Reverse: 5' GTC CAG CTC GAC CAG GAT G 3'

PCR genotyping reaction

1.0uL primers (10uM each)
1.0uL genomic DNA
8uL water
10uL GoTaq Green Master Mix 2x (Promega REF M7128)

Genotyping cycling conditions

95C 5min
30x [95C 1min
56C 1min
72C 2min]
72C 5min

Luciferase transfection experiments

Transient transfection experiments within ZF4 zebrafish cells lines were performed following a previous protocol [27].

ZNF143 and CHD8 polyclonal antibodies; ChIP-seq

Recombinant zebrafish ZNF143 protein was prepared by Jessica Tracy, recombinant CHD8 protein by Gary Kunkel, and antisera were tested and ChIP-seq samples prepared by Gary Kunkel. Briefly, recombinant full-length zebrafish ZNF143 protein containing 6HIS and myc N-terminal tags was expressed in E. coli from pET28/mycznf143 plasmid. Likewise, a fragment of the zebrafish CHD8 protein containing amino acids 1 – 591 was subcloned into pET28b (pET28/chd8A*) and expressed in E. coli. The insoluble fractions from cell lysates were suspended in 6M urea-containing buffer, and proteins were purified partially using HIS-BIND (Novagen) chromatography in 6M urea-containing buffers. A final purification step used SDS-PAGE preparative gels.

After brief staining, the intense and well-separated ZNF143 or CHD8 fragment bands were excised and used as antigens for preparation of rabbit polyclonal antisera by Cocalico Biologicals (Reamstown, PA). The antisera from the TAMU163 rabbit was used for ZNF143 ChIP experiments, while that from the TAMU165 rabbit was optimal for CHD8 ChIP experiments. Antisera were affinity-selected by binding/elution to nitrocellulose pieces of an immunoblot derived from the appropriate protein. Chromatin from a total of approximately 2500 formaldehyde-crosslinked 24hpf zebrafish embryos was used for each ChIP following a scaled-up protocol of that described in this Methods section. Selected DNAs were submitted to Texas AgriLife Genomics and Bioinformatics Services for sequencing with the Illumina HiSeq2500. The ZNF143 sample consisted of a spike-in to one lane. Sequencing reads were aligned to the zebrafish genome using the CLC Genomics Workbench version 8 program (Qiagen) with the following output: ZNF143: 50 nt read lengths; 1,161,758 reads mapped; 673 ZNF143 peaks identified. CHD8: 250 nt read lengths; 25,723,706 reads mapped; 1791 CHD8 peaks identified.

Additional antibodies used in ChIP

Antibodies used in chromatin immunoprecipitation experiments that were purchased from a manufacturer are listed below.

Anti-RNAPII RPB1 (Biolegend- Cat 664912 Lot B281694)

Acetyl- Histone H3 Lys9 (AcH3K9) (Pierce #MA5-11195 Lot OJ1766371)

H3K9me3 (Diagenode Cat SN-146-100 Lot 001)

H3K4me3 (Epicypher Cat 13-0041 Lot 20083002-42)

Chromatin Immunoprecipitation of zebrafish embryos

Formaldehyde cross-linking

24hpf embryos were first dechorionated and then crosslinked using 1.85% formaldehyde and allowed to incubate for 15 minutes at room temperature. 0.125M glycine was used to quench the cross-linking reaction at room temperature for 5 minutes. Embryos were washed 3 times with cold PBS. Embryos were then either stored at -80C upon removal of PBS or continued to sonification.

Sonification/chromatin prep

200-300 embryos were resuspended in lysis buffer (10mM Tris (pH=8.0), 10mM NaCl, 0.5% IGEPAL). 5uL of Sigma protease inhibitor cocktail (P8340) plus 5uL of 100mM PMSF solution were added to embryos and homogenized manually with a dounce. After incubation on ice for 15 minutes nuclei were centrifuged at 3500 RPM for 5min at 4C. The resulting pellet was resuspended in 0.2mL of sonification buffer (1% SDS, 10mM EDTA, 50mM tris (pH=8.0)) plus 2uL of PMSF and mixed. 0.8mL of sonification buffer without SDS and 8uL of 100mM PMSF was added. Chromatin was sheared using a microtip sonicator (Branson Sonifier 250: 20 second intervals, duty cycle: constant level, output control: 1.5). Chromatin was incubated on ice for 5 minutes between bursts. This was repeated for a total of 12 times). Sonicated DNA was centrifuged at 14,000 RPM, for 10minutes at 4C. Aliquots of 50 embryo equivalents were used for ChIP samples. Sonicated chromatin was 200-300bp in length, after analysis on 1.5% agarose gels.

A/G magnetic bead prep

150uL of A/G beads (Pierce Protein A/G Magnetic Beads 1896188) were pulled on a magnetic stand and resuspended with 1.0mL of ChIP wash buffer I (50mM NaCl, 2mM EDTA, 0.2% Sarkosyl). This was incubated at room temperature on a Labquake rotisserie shaker. This was repeated three times. Beads underwent a final wash with 0.5mL of wash buffer I. 50uL of 20mg/mL BSA and 100ug of E. coli tRNA was added and allowed to incubate on the Labquake for 3 hours at 4C. Beads were pulled on magnetic stand and a final suspension of were resuspended with 150uL of dilution buffer. Prepped beads were stored at 4C.

Immunoprecipitation using A/G beads

An aliquot of chromatin was diluted with an equal volume of dilution buffer (200mM NaCl, 2% Triton X-100). A 1:200 volume of Sigma protease inhibitor and 1:200 volume of 100mM PMSF was added along with 10uL of prepped protein A/G beads. The beads were incubated on the Labquake for 4hr at 4C. Beads were pulled on magnetic stand and supernatant transferred to a new tube. 5uL of prepped A/G beads and 3ug of purchased antibody (or 5uL of ZNF143 whole antiserum) were added to the supernatant. This was incubated overnight on the Labquake at 4C.

Wash Steps/Crosslink reversal

Beads were pulled using a magnetic stand and the supernatant was discarded. Beads were resuspended with 1mL wash buffer I (50mM NaCl, 2mM EDTA, 0.2% Sarkosyl)

and incubated at room temperature on the Labquake for 5minutes. Beads were pulled on the magnetic stand, the supernatant was discarded, and resuspended with wash buffer II (11mM Tris-HCl (pH=8.1), 250mM NaCl, 0.5% IGEPAL, 1% deoxycholic acid for αRNAPII, αH3K4me3, αH3K9me3 ChIPs) (αZNF143, αacH3K9 ChIPs used the more stringent wash buffer II containing 11mM Tris-HCl (pH=8.1), 500mM LiCl, 1% IGEPAL, 1% deoxycholic acid) and incubated on the Labquake for 5minutes at room temperature. The supernatant was discarded. This process was repeated once more with wash buffer I, twice with wash buffer II, three times with wash buffer I, and then resuspended with 200uL of elution buffer (50mM Tris (pH=8.0, 10mM EDTA, 1% SDS). Beads were eluted at 55C for 1 hour. Beads were vortexed and pulled on magnetic stand. The supernatant was transferred to a new tube and 0.67uL of 20mg/mL Proteinase K (NEB P8102S) was added. This was incubated at 55C for 2 hours. The temperature was increased to 65C and incubated overnight. Sample DNA was then cleaned up using a Qiagen PCR cleanup kit and eluted into 50uL of EB (Qiaquick PCR Purification Kit #28104).

Input DNA preparation

An aliquot of starting chromatin was prepped in parallel with the reversal of cross links for immunoprecipitated samples. An equal volume of dilution buffer was added to starting chromatin. 0.67uL of 20mg/mL Proteinase K and 1.5uL of 10mg/mL RNase A (Thermo Scientific EN0531) was added and incubated at 55C for 2 hours, then increased

to 65C to incubate overnight. A phenol chloroform extraction was performed on input DNA. This was resuspended with 50uL of EB (Qiaquick PCR Purification Kit #28104).

Detection of IP DNA via PCR

A negative water control was performed as a PCR negative control. A no antibody control was performed along immunoprecipitated samples. Two positive control titrated input DNA controls were performed as described below. Primer sets included a forward primer in the promoter of interest and an eGFP +60Rev primer as reported in the transgenic genotyping PCR analysis. ChIP-qPCR C_t values were normalized using the %input method as described in [130].

ChIP-PCR

Immunoprecipitated samples

3.0uL of immunoprecipitated DNA (mock or IP)
1.0uL (10uM primers)
10uL of GoTaq Green Master Mix 2x (Promega REF M7128)
Water to fill to 20uL

Input samples

0.5uL or 2.0uL of 1:100 dilution of input DNA
1.0uL (10uM primers)
10uL of GoTaq Green
Water to fill to 20uL

Cycling conditions

94C 3 minutes

30x [95C 1min
60C 2min
72C 2min]

72C 7min

ChIP-qPCR

Immunoprecipitated Samples

1.3uL of water
0.7uL (10uM primers)
3uL of IP sample
5uL of BioRad iTaq SYBR Green Supermix (#1725120)

Input

3.8uL of water
0.7uL (10uM primers)
0.5uL input DNA
5uL SYBR Supermix

Cycling conditions

95C 10:00m
39x
95C 0:15s
59C 0:30s
Plate read
72C 0:30s
Plate read
65C 0:31s
60x
65C 0:05s (+0.5C/cycle-ramp 0.5C/sec)
Plate read

Fluorescence microscopy

Fluorescence images were taken in live embryos on a Zeiss Axio Imager.M2 fluorescence microscope at 5x objective. Images were processed using Zen Pro.

CHAPTER IV : CONCLUSIONS

Appropriate regulation of gene expression in eukaryotes must be responsive to a wide range of cellular cues. One step by which gene expression is controlled is at the level of transcription initiation. A diverse set of mechanisms underly this regulatory step. *Cis*-regulatory elements contained in DNA sequences that bind *trans*-acting regulatory factors, such as activator proteins, facilitate the process of initiation. Activator proteins are bipartite, consisting of a DNA binding domain and an activation domain. Molecular targets of many known activation domains remain ambiguous, and therefore mechanisms of transcriptional activation are poorly understood.

ZNF143 is a common transcriptional activator protein, binding an estimated 2000 promoter regions within vertebrates[19]. Versatile in its regulation, ZNF143 is able to activate transcription at small RNA and mRNA encoding promoter regions transcribed by RNAPII and RNAPIII[15]. However, there are many unknowns regarding how ZNF143 fulfills its activator role within the cell. As outlined in Chapter I, there is evidence of a diverse set of potential ZNF143 activation mechanisms. These include activation through facilitation of a pre-initiation complex assembly[41], association with a chromatin remodeler cofactor [128], stimulating chromatin interactions with distal regulatory elements[17], or associations with stimulatory transcription histone modifications [17]. The aim of this dissertation was

to further explore the mechanism of ZNF143 through a characterization of the genetic context of target promoters bound and activated by ZNF143.

This research employs zebrafish as a useful model to study the mechanism of activation by ZNF143. However, a previously uncharacterized paralog of *znf143*, *znf143a*, has been identified in zebrafish. This paralog likely arose due to an additional genome duplication event in teleost fish. To understand activation by *znf143*, the contributions of *znf143a* had to be characterized (Ch. II). These ZNF143 proteins contain a 65% amino acid identity between the two proteins[28]. Due to their redundancy, it is possible that paralogous genes either acquire new functions or are thought to be lost over time [131]. Retention of *znf143a* could be for several non-mutually exclusive reasons: including a gene dosage effect [132], providing a buffer against deleterious mutations [104], subfunctionalization[133], or neofunctionalization[134].

Transient transfection assays in zebrafish ZF4 cells revealed that the activation activity of *znf143b* is maintained for *znf143a* [28]. Despite similar tissue-specific expression patterns, quantitative expression analyses revealed that a higher amount of *znf143a* is expressed in early embryogenesis of zebrafish when compared to *znf143b*[28]. This provides evidence for the subfunctionalization of expression levels for *znf143a*, as there is a partitioning of quantitative expression levels of the two *znf143* genes. Additionally, knockdown experiments caused morphological defects in zebrafish embryos when targeting either *znf143* gene. In these experiments, the complementary *znf143* paralog was unable to fully compensate the activity of the

paralog targeted for knockdown. This may be a gene dosage effect whereas *znf143a* contributes to the fitness of zebrafish by providing an increase in overall expression of *znf143*. Transcription factors are commonly dosage-sensitive genes, as they are a class of proteins highly retained after whole genome duplication events [132]. ZNF143 may be representative of a transcription factor gene requiring a high gene dosage for the fitness of zebrafish.

There are areas of potential divergence between the *znf143* paralogs that we do not yet understand. There is considerable divergence in amino acid sequence within the C-terminus of the ZNF143 proteins, and a slight divergence within the small RNA gene activation domain[28]. The C-terminus of ZNF143 remains uncharacterized, and therefore cannot be discounted as a potential domain that might confer different functions between paralogs. Overall, *znf143a* and *znf143b* are important for activation and fitness of a developing zebrafish. Loss of either paralog leads to brain and axial developmental defects.

In order to investigate the context of a promoter regulated by ZNF143, a candidate promoter approach was employed for the construction of zebrafish transgenic lines. This involved an initial identification of promoter regions containing SPH motifs that both bind and are activated by ZNF143. ZNF143 shares a protein interaction with CHD8 [128] and displays increased co-activation at target promoters by CHD8 in transfection analyses (Chapter III). For these reasons, binding and co-activation by CHD8 in concert with ZNF143 was included in searching for promoter candidates. This analysis included three promoters representing three

different classes: 1. A small RNA encoding promoter co-regulated by ZNF143 and CHD8, 2. An mRNA encoding promoter regulated by ZNF143, and 3. An mRNA encoding promoter co-regulated by ZNF143 and CHD8. The *u4*, *gsk3ab*, and *atic* promoters were identified as fitting these categories of promoters respectively and were incorporated as model promoters for ZNF143 regulation in transgenesis experiments.

The resulting *gsk3ab+SPH:eGFP* transgenic line demonstrated binding of ZNF143 and RNAPII to the transgene. This was suspected as both *atic* and *gsk3ab* transgenic lines drive strong fluorescence during the development of zebrafish embryos. Furthermore, *gsk3ab+SPH:eGFP* contained enrichment of histone modifications that are associated with typically active promoter or enhancer regions. The active modifications in this study included both H3K4me3 and AcH3K9. This corroborates previous research suggesting that cell type specific ZNF143 binding sites often harbor active epigenetic modifications such as H3K4me1 or H3K4me3 [17].

Furthermore, preliminary ChIP-PCR experiments revealed a loss of ZNF143, RNAPII, and AcH3K9 enrichment at the *aticSPHmut:eGFP* transgene. There was also a loss of ZNF143 and RNAPII enrichment at *gsk3abSPHmut:eGFP*. While future experiments need to be performed for statistical significance, these preliminary results suggest that *atic* and *gsk3ab* are activated by ZNF143. Although still unknown, there might be an indirect association between ZNF143 and AcH3K9 modifications at the *atic* promoter. *Gsk3a+SPH:eGFP* also contained enrichment of

H3K4me3. Future ChIP experiments should be performed to detect a possible reduction of H3K4me3 enrichment at *gsk3abSPHmut:eGFP*. A simple potential mechanism of ZNF143 activation through an association with active histone modifications at target promoter regions is outlined in Fig. 28.

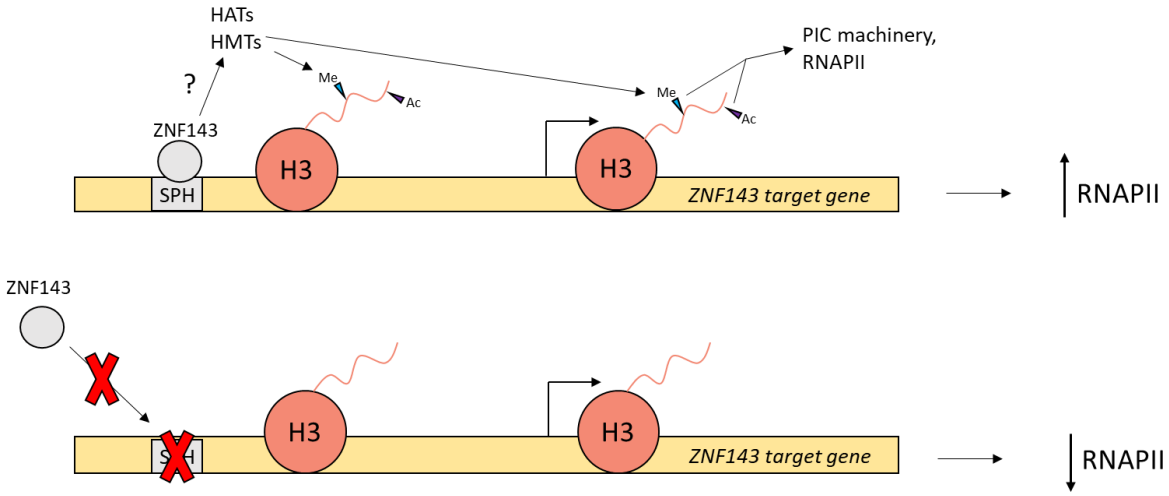


Figure 28. Potential mechanism of activation of target promoters by ZNF143 through active histone modifications.
ZNF143 binds to SPH motifs within target promoter regions. Histone H3 of the -1 and +1 nucleosomes is displayed. Promoter regions regulated by ZNF143 are associated with active histone modifications. ZNF143 may aid recruitment of histone acetyltransferases or histone methyltransferases, although many co-factors of activation by ZNF143 are still unidentified. Target promoter regions lacking a functional SPH motif contain decreased enrichment of active histone modifications.

Future ChIP experiments should address the binding of CHD8 and ZNF143 at the *atic* transgenes. Due to a direct interaction between CHD8 and ZNF143[128], a lack of ZNF143 binding could impact the recruitment of CHD8 at the *aticSPHmut* promoter. This has yet to be explored. While active histone modifications were not able to be detected at the *aticSPHmut:eGFP* transgene, future experimentation would also address the presence of these same modifications at *atic+SPH:eGFP*.

There are several mechanisms that have addressed the potential binding specificity of CHD8, including both the association of CHD8 with active histone modifications and recruitment of CHD8 via protein interactions [135]. It may be possible that active histone modifications are involved with recruitment of CHD8 in addition to a potential ZNF143-CHD8 recruitment model.

Co-activation by CHD8short at ZNF143 regulated promoter regions remains to be explored. One possible model of activation by ZNF143 could be that ZNF143 binds to the SPH motif within a promoter region, recruits CHD8, and thereby remodels nucleosomes in a manner that renders a promoter more accessible to transcription. However this model does not work for co-activation by CHD8short due to the lack of a helicase domain critical for the remodeling of nucleosomes in this isoform [81]. A possible feature of co-activation by ZNF143 and CHD8short is that ZNF143 binds to a SPH element, recruits CHD8short, thus preventing recruitment of the full length CHD8 isoform[81]. The long isoform of CHD8, CHD8l, is also associated with repression and thus a de-repression mechanism could be responsible for observed activation by the short CHD8 isoform[81].

Overall, ZNF143 is a critical regulator of vertebrate promoter regions. Both its versatility in regulating a variety of promoter types and ubiquitous expression among most tissues renders ZNF143 likely to be involved in multiple mechanisms of activation. Furthermore, the lack of structural information regarding activation domains leaves our understanding of transcriptional activators, including ZNF143, obscure. Within zebrafish evolution, an additional paralog of *znf143*, required for the fitness of developing

zebrafish, is retained and contributes to the overall activation by ZNF143. Although more evidence is needed, ZNF143 may be associated with positive histone modifications that facilitate recruitment of RNAPII to promoter regions containing SPH motifs.

REFERENCES

1. Turowski, T.W. and D. Tollervey, *Transcription by RNA polymerase III: insights into mechanism and regulation.* Biochemical Society transactions, 2016. **44**(5): p. 1367-1375.

2. Haberle, V. and A. Stark, *Eukaryotic core promoters and the functional basis of transcription initiation.* Nature Reviews Molecular Cell Biology, 2018. **19**(10): p. 621-637.

3. Nikolov, D.B. and S.K. Burley, *RNA polymerase II transcription initiation: A structural view.* Proceedings of the National Academy of Sciences, 1997. **94**(1): p. 15-22.

4. Gupta, K., et al., *Zooming in on Transcription Preinitiation.* Journal of Molecular Biology, 2016. **428**(12): p. 2581-2591.

5. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-192.

6. Cantin, G.T., J.L. Stevens, and A.J. Berk, *Activation domain–mediator interactions promote transcription preinitiation complex assembly on promoter DNA.* Proceedings of the National Academy of Sciences, 2003. **100**(21): p. 12003.

7. Erkina, T.Y. and A.M. Erkine, *Nucleosome distortion as a possible mechanism of transcription activation domain function.* Epigenetics & Chromatin, 2016. **9**(1): p. 40.

8. Spiegelman, B.M. and R. Heinrich, *Biological Control through Regulated Transcriptional Coactivators.* Cell, 2004. **119**(2): p. 157-167.

9. Spitsbergen, J.M. and M.L. Kent, *The state of the art of the zebrafish model for toxicology and toxicologic pathology research--advantages and current limitations.* Toxicologic pathology, 2003. **31 Suppl**(Suppl): p. 62-87.

10. Roebuck, K.A., R.J. Walker, and W.E. Stumph, *Multiple functional motifs in the chicken U1 RNA gene enhancer.* Mol Cell Biol, 1987. **7**(12): p. 4185-93.

11. Myslinski, E., A. Krol, and P. Carbon, *Optimal tRNA((Ser)Sec) gene activity requires an upstream SPH motif.* Nucleic Acids Res, 1992. **20**(2): p. 203-9.

12. Schaub, M., et al., *Staf, a promiscuous activator for enhanced transcription by RNA polymerases II and III.* EMBO J, 1997. **16**(1): p. 173-81.

13. Kubota, H., et al., *Transcriptional regulation of the mouse cytosolic chaperonin subunit gene Ccta/t-complex polypeptide 1 by selenocysteine tRNA gene transcription activating factor family zinc finger proteins.* J Biol Chem, 2000. **275**(37): p. 28641-8.

14. Ngondo-Mbongo, R.P., et al., *Modulation of gene expression via overlapping binding sites exerted by ZNF143, Notch1 and THAP11.* Nucleic Acids Res, 2013. **41**(7): p. 4000-14.

15. Schuster, C., A. Krol, and P. Carbon, *Two distinct domains in Staf to selectively activate small nuclear RNA-type and mRNA promoters.* Mol Cell Biol, 1998. **18**(5): p. 2650-8.

16. Xie, D., et al., *Dynamic trans-acting factor colocalization in human cells.* Cell, 2013. **155**(3): p. 713-724.

17. Bailey, S.D., et al., *ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters.* Nat Commun, 2015. **2**: p. 6186.

18. Wen, Z., et al., *ZNF143 is a regulator of chromatin loop.* Cell Biol Toxicol, 2018. **34**(6): p. 471-478.

19. Myslinski, E., et al., *A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human Staf/ZNF143 in mammalian promoters.* J Biol Chem, 2006. **281**(52): p. 39953-62.

20. Roebuck, K.A., et al., *Octamer and SPH motifs in the U1 enhancer cooperate to activate U1 RNA gene expression.* Mol Cell Biol, 1990. **10**(1): p. 341-52.

21. Gonzalez, D., et al., *ZNF143 protein is an important regulator of the myeloid transcription factor C/EBPα.* Journal of Biological Chemistry, 2017. **292**(46): p. 18924-18936.

22. Schaub, M., A. Krol, and P. Carbon, *Flexible zinc finger requirement for binding of the transcriptional activator staf to U6 small nuclear RNA and tRNA(Sec) promoters.* J Biol Chem, 1999. **274**(34): p. 24241-9.

23. Schaub, M., A. Krol, and P. Carbon, *Structural organization of Staf-DNA complexes.* Nucleic Acids Res, 2000. **28**(10): p. 2114-21.

24. Minezaki, Y., et al., *Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation.* J Mol Biol, 2006. **359**(4): p. 1137-49.

25. Boija, A., et al., *Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.* Cell, 2018. **175**(7): p. 1842-1855.e16.

26. Myslinski, E., A. Krol, and P. Carbon, *ZNF76 and ZNF143 are two human homologs of the transcriptional activator Staf.* J Biol Chem, 1998. **273**(34): p. 21998-2006.

27. Halbig, K.M., A.C. Lekven, and G.R. Kunkel, *The transcriptional activator ZNF143 is essential for normal development in zebrafish.* BMC Mol Biol, 2012. **13**: p. 3.

28. Huning, L. and G.R. Kunkel, *Two paralogous znf143 genes in zebrafish encode transcriptional activator proteins with similar functions but expressed at different levels during early development.* BMC Mol Cell Biol, 2020. **21**(1): p. 3.

29. Ishiguchi, H., et al., *ZNF143 activates gene expression in response to DNA damage and binds to cisplatin-modified DNA.* Int J Cancer, 2004. **111**(6): p. 900-9.

30. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**.

31. Myslinski, E., et al., *An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene.* Nucleic Acids Res, 2001. **29**(12): p. 2502-9.

32. Murphy, S., C. Di Liegro, and M. Melli, *The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter.* Cell, 1987. **51**(1): p. 81-87.

33. Bark, C., et al., *Genes for human U4 small nuclear RNA.* Gene, 1986. **50**(1-3): p. 333-44.

34. Lund, E. and J.E. Dahlberg, *True genes for human U1 small nuclear RNA. Copy number, polymorphism, and methylation.* J Biol Chem, 1984. **259**(3): p. 2013-21.

35. Kunkel, G.R., et al., *U6 small nuclear RNA is transcribed by RNA polymerase III.* Proc Natl Acad Sci U S A, 1986. **83**(22): p. 8575-9.

36. Maraia, R.J., et al., *The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes.* Nucleic acids research, 1994. **22**(15): p. 3045-3052.

37. Mattaj, I.W., et al., *An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes.* Nature, 1985. **316**(6024): p. 163-7.

38. Hernandez, N., *Transcription of vertebrate snRNA genes and related genes*, in *Transcriptional regulation*, Y.K.R. MacKnight S.L., Editor. 1992, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 1103 1135.

39. Halbig, K.M., A.C. Lekven, and G.R. Kunkel, *Zebrafish U6 small nuclear RNA gene promoters contain a SPH element in an unusual location.* Gene, 2008. **421**(1–2): p. 89-94.

40. Kunkel, G.R., et al., *Identification of a SPH element in the distal region of a human U6 small nuclear RNA gene promoter and characterization of the SPH binding factor in HeLa cell extracts.* Gene Expr, 1996. **6**(2): p. 59-72.

41. Kunkel, G.R. and J.D. Hixson, *The distal elements, OCT and SPH, stimulate the formation of preinitiation complexes on a human U6 snRNA gene promoter in vitro.* Nucleic Acids Res, 1998. **26**(6): p. 1536-43.

42. Cheung, C.H., Q.N. Fan, and W.E. Stumph, *Structural requirements for the functional activity of a U1 snRNA gene enhancer.* Nucleic Acids Res, 1993. **21**(2): p. 281-7.

43. Danzeiser, D.A., O. Urso, and G.R. Kunkel, *Functional characterization of elements in a human U6 small nuclear RNA gene distal control region.* Mol Cell Biol, 1993. **13**(8): p. 4670-8.

44. Mach, C.M., B.W. Hargrove, and G.R. Kunkel, *The Small RNA gene activator protein, SphI postoctamer homology-binding factor/selenocysteine tRNA gene transcription activating factor, stimulates transcription of the human interferon regulatory factor-3 gene.* J Biol Chem, 2002. **277**(7): p. 4853-8.

45. Saur, D., et al., *Complex regulation of human neuronal nitric-oxide synthase exon 1c gene transcription. Essential role of Sp and ZNF family members of transcription factors.* J Biol Chem, 2002. **277**(28): p. 25798-814.

46. Barski, O.A., et al., *Regulation of aldehyde reductase expression by STAF and CHOP.* Genomics, 2004. **83**(1): p. 119-29.

47. Di Leva, F., et al., *Human synaptobrevin-like 1 gene basal transcription is regulated through the interaction of selenocysteine tRNA gene transcription activating factor-zinc finger 143 factors with evolutionary conserved cis-elements.* J Biol Chem, 2004. **279**(9): p. 7734-9.

48. Grossman, C.E., et al., *ZNF143 mediates basal and tissue-specific expression of human transaldolase.* J Biol Chem, 2004. **279**(13): p. 12190-205.

49. Gerard, M.A., A. Krol, and P. Carbon, *Transcription factor hStaf/ZNF143 is required for expression of the human TFAM gene.* Gene, 2007. **401**(1-2): p. 145-53.

50. Anno, Y.N., et al., *Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription.* Nucleic Acids Res, 2011. **39**(8): p. 3116-27.

51. Paek, A.R., et al., *IGF-1 induces expression of zinc-finger protein 143 in colon cancer cells through phosphatidylinositide 3-kinase and reactive oxygen species.* Exp Mol Med, 2010. **42**(10): p. 696-702.

52. Paek, A.R., C.H. Lee, and H.J. You, *A role of zinc-finger protein 143 for cancer cell migration and invasion through ZEB1 and E-cadherin in colon cancer cells.* Mol Carcinog, 2014. **53 Suppl 1**: p. E161-8.

53. Verma, V., et al., *Loss of zinc-finger protein 143 contributes to tumour progression by interleukin-8-CXCR axis in colon cancer.* J Cell Mol Med, 2019. **23**(6): p. 4043-4053.

54. Chen, Z.H., et al., *Eukaryotic initiation factor 4A2 promotes experimental metastasis and oxaliplatin resistance in colorectal cancer.* J Exp Clin Cancer Res, 2019. **38**(1): p. 196.

55. Sadłecki, P., et al., *Expression of zinc finger transcription factors (ZNF143 and ZNF281) in serous borderline ovarian tumors and low-grade ovarian cancers.* Journal of ovarian research, 2019. **12**(1): p. 23-23.

56. Paek, A.R. and H.J. You, *GAIP-interacting protein, C-terminus is involved in the induction of zinc-finger protein 143 in response to insulin-like growth factor-1 in colon cancer cells.* Mol Cells, 2011. **32**(5): p. 415-9.

57. Paek, A.R., et al., *Zinc finger protein 143 expression is closely related to tumor malignancy via regulating cell motility in breast cancer.* BMB Rep, 2017. **50**(12): p. 621-627.

58. Wei, S., et al., *ZNF143 enhances metastasis of gastric cancer by promoting the process of EMT through PI3K/AKT signaling pathway.* Tumour Biol, 2016. **37**(9): p. 12813-12821.

59. Myslinski, E., et al., *Transcription of the human cell cycle regulated BUB1B gene requires hStaf/ZNF143.* Nucleic Acids Res, 2007. **35**(10): p. 3453-64.

60. Bharadwaj, R. and H. Yu, *The spindle checkpoint, aneuploidy, and cancer.* Oncogene, 2004. **23**(11): p. 2016-2027.

61. Izumi, H., et al., *Role of ZNF143 in tumor growth through transcriptional regulation of DNA replication and cell-cycle-associated genes.* Cancer Sci, 2010. **101**(12): p. 2538-45.

62. Izumi, H., et al., *Forced Expression of ZNF143 Restrains Cancer Cell Growth.* Cancers (Basel), 2011. **3**(4): p. 3909-20.

63. Ye, B., W. Shen, and Z. Zhao, *ZNF143 is dynamically bound to a subset of its interphase sites during mitosis.* Biochemical and Biophysical Research Communications, 2020. **523**(2): p. 293-298.

64. Lee, Tong I. and Richard A. Young, *Transcriptional Regulation and Its Misregulation in Disease.* Cell, 2013. **152**(6): p. 1237-1251.

65. Kawatsu, Y., et al., *The combination of strong expression of ZNF143 and high MIB-1 labelling index independently predicts shorter disease-specific survival in lung adenocarcinoma.* British Journal Of Cancer, 2014. **110**: p. 2583.

66. Paek, A.R., et al., *The Role of ZNF143 in Breast Cancer Cell Survival Through the NAD(P)H Quinone Dehydrogenase 1(-)p53(-)Beclin1 Axis Under Metabolic Stress.* Cells, 2019. **8**(4).

67. Zhao, Y., et al., *Construction of disease-specific transcriptional regulatory networks identifies co-activation of four gene in esophageal squamous cell carcinoma.* Oncol Rep, 2017. **38**(1): p. 411-417.

68. Torigoe, T., et al., *Cisplatin resistance and transcription factors.* Curr Med Chem Anticancer Agents, 2005. **5**(1): p. 15-27.

69. Uchiumi, T. and D. Knag, *[Cisplatin sensitivity transcriptional factor and mitochondrial DNA maintenance protein].* Rinsho Byori, 2009. **57**(10): p. 978-86.

70. Wakasugi, T., et al., *ZNF143 interacts with p73 and is involved in cisplatin resistance through the transcriptional regulation of DNA repair genes.* Oncogene, 2007. **26**(36): p. 5194-203.

71. Cabezas, O.R., et al., *Polycystic Kidney Disease with Hyperinsulinemic Hypoglycemia Caused by a Promoter Mutation in Phosphomannomutase 2.* J Am Soc Nephrol, 2017. **28**(8): p. 2529-2539.

72. Kim, Y., et al., *A Mutation in ZNF143 as a Novel Candidate Gene for Endothelial Corneal Dystrophy.* J Clin Med, 2019. **8**(8).

73. Radziwon, A., et al., *Single-base substitutions in the CHM promoter as a cause of choroideremia.* Hum Mutat, 2017. **38**(6): p. 704-715.

74. Chen, X., et al., *Zfp143 regulates Nanog through modulation of Oct4 binding.* Stem Cells, 2008. **26**(11): p. 2759-67.

75. Blomen, V.A., et al., *Gene essentiality and synthetic lethality in haploid human cells.* Science, 2015. **350**(6264): p. 1092-6.

76. Zhao, X., P.S. Pendergrast, and N. Hernandez, *A positioned nucleosome on the human U6 promoter allows recruitment of SNAPc by the Oct-1 POU domain.* Mol Cell, 2001. **7**(3): p. 539-49.

77. Murphy, S., *Differential in vivo activation of the class II and class III snRNA genes by the POU-specific domain of Oct-1.* Nucleic acids research, 1997. **25**(11): p. 2068-2076.

78. Carbon, P. and A. Krol, *Transcription of the Xenopus laevis selenocysteine tRNA(Ser)Sec gene: a system that combines an internal B box and upstream elements also found in U6 snRNA genes.* The EMBO journal, 1991. **10**(3): p. 599-606.

79. James Faresse, N., et al., *Genomic Study of RNA Polymerase II and III SNAP(c)-Bound Promoters Reveals a Gene Transcribed by Both Enzymes and a Broad Use of Common Activators.* PLoS Genet, 2012. **8**(11).

80. Yuan, C.C., et al., *CHD8 associates with human Staf and contributes to efficient U6 RNA polymerase III transcription.* Mol Cell Biol, 2007. **27**(24): p. 8729-38.

81. Kunkel, G.R., et al., *CHD8short, a naturally-occurring truncated form of a chromatin remodeler lacking the helicase domain, is a potent transcriptional coregulator.* Gene, 2018. **641**: p. 303-309.

82. Wang, H., et al., *Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells.* Proceedings of the National Academy of Sciences, 2011. **108**(36): p. 14908.

83. Sathyan, K.M., et al., *An improved auxin-inducible degron system preserves native protein levels and enables rapid and specific protein depletion.* Genes Dev, 2019. **33**(19-20): p. 1441-1455.

84. Vinckevicius, A., J.B. Parker, and D. Chakravarti, *Genomic Determinants of THAP11/ZNF143/HCFC1 Complex Recruitment to Chromatin*. Mol Cell Biol, 2015. **35**(24): p. 4135-46.

85. Zhang, L., et al., *ZNF143-mediated H3K9 trimethylation upregulates CDC6 by activating MDIG in hepatocellular carcinoma.* Cancer Research, 2020: p. canres.3226.2019.

86. Crews, S.T. and J.C. Pearson, *Transcriptional autoregulation in development.* Current biology : CB, 2009. **19**(6): p. R241-R246.

87. Ngondo, R.P. and P. Carbon, *Transcription factor abundance controlled by an auto-regulatory mechanism involving a transcription start site switch*. Nucleic Acids Research, 2014. **42**(4): p. 2171-2184.

88. Ngondo, R.P. and P. Carbon, *ZNF143 is regulated through alternative 3'UTR isoforms.* Biochimie, 2014. **104**: p. 137-46.

89. Phillips-Cremins, J.E., et al., *Architectural protein subclasses shape 3D organization of genomes during lineage commitment.* Cell, 2013. **153**(6): p. 1281-1295.

90. Holwerda, S.J.B. and W. de Laat, *CTCF: the protein, the binding partners, the binding sites and their chromatin loops.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2013. **368**(1620): p. 20120369-20120369.

91. Heidari, N., et al., *Genome-wide map of regulatory interactions in the human genome.* Genome Res, 2014. **24**(12): p. 1905-17.

92. Ye, B.Y., et al., *ZNF143 is involved in CTCF-mediated chromatin interactions by cooperation with cohesin and other partners.* Molecular Biology, 2016. **50**(3): p. 431-437.

93. Zhang, K., et al., *Systematic identification of protein combinations mediating chromatin looping.* Nat Commun, 2016. **7**: p. 12249.

94. Mourad, R. and O. Cuvier, *TAD-free analysis of architectural proteins and insulators.* Nucleic acids research, 2018. **46**(5): p. e27-e27.

95. Yang, J. and V.G. Corces, *Chromatin insulators: a role in nuclear organization and gene expression.* Advances in cancer research, 2011. **110**: p. 43-76.

96. Klenova, E., et al., *YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders*. J Neurosci, 2004. **24**(26): p. 5966-73.

97. Uusküla-Reimand, L., et al., *Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders.* Genome Biol, 2016. **17**(1): p. 182.

98. Clarkson, C.T., et al., *CTCF-dependent chromatin boundaries formed by asymmetric nucleosome arrays with decreased linker length.* Nucleic Acids Res, 2019. **47**(21): p. 11181-11196.

99. Ishihara, K., M. Oshimura, and M. Nakao, *CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling*. Molecular Cell, 2006. **23**(5): p. 733-742.

100. Ruzicka, L., et al., *The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources.* Nucleic Acids Res, 2019. **47**(D1): p. D867-d873.

101. Maconochie, M., et al., *PARALOGOUS HOX GENES: Function and Regulation.* Annual Review of Genetics, 1996. **30**(1): p. 529-556.

102. Gabaldón, T. and E.V. Koonin, *Functional and evolutionary implications of gene orthology*. Nature reviews. Genetics, 2013. **14**(5): p. 360-366.

103. Kryuchkova-Mostacci, N. and M. Robinson-Rechavi, *Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs.* PLoS computational biology, 2016. **12**(12): p. e1005274-e1005274.

104. Dandage, R. and C.R. Landry, *Paralog dependency indirectly affects the robustness of human cells.* Molecular systems biology, 2019. **15**(9): p. e8871-e8871.

105. Schuster, C., et al., *Staf, a novel zinc finger protein that activates the RNA polymerase III promoter of the selenocysteine tRNA gene*. EMBO J, 1995. **14**(15): p. 3777-87.

106. Rincon, J.C., et al., *Molecular cloning of a cDNA encoding human SPH-binding factor, a conserved protein that binds to the enhancer-like region of the U6 small nuclear RNA gene promoter*. Nucleic Acids Res, 1998. **26**(21): p. 4846-52.

107. Mourad, R. and O. Cuvier, *Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation*. PLoS Comput Biol, 2016. **12**(5): p. e1004908.

108. Michaud, J., et al., *HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy*. Genome Res, 2013. **23**(6): p. 907-16.

109. Parker, J.B., et al., *Host Cell Factor-1 Recruitment to E2F-Bound and Cell-Cycle-Control Genes Is Mediated by THAP11 and ZNF143*. Cell Reports. **9**(3): p. 967-982.

110. Guschanski, K., M. Warnefors, and H. Kaessmann, *The evolution of duplicate gene expression in mammalian organs*. Genome Res, 2017. **27**(9): p. 1461-1474.

111. Postlethwait, J.H., et al., *Zebrafish comparative genomics and the origins of vertebrate chromosomes*. Genome Res, 2000. **10**(12): p. 1890-902.

112. Huang, X. and W. Miller, *A time-efficient, linear-space local similarity algorithm*. Advances in Applied Mathematics, 1991. **12**(3): p. 337-357.

113. Qi, Lei S., et al., *Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression*. Cell. **152**(5): p. 1173-1183.

114. Mandegar, M.A., et al., *CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs*. Cell Stem Cell, 2016. **18**(4): p. 541-53.

115. Long, L., et al., *Regulation of transcriptionally active genes via the catalytically inactive Cas9 in C. elegans and D. rerio*. Cell research, 2015. **25**(5): p. 638-641.

116. El-Brolosy, M.A. and D.Y.R. Stainier, *Genetic compensation: A phenomenon in search of mechanisms*. PLoS genetics, 2017. **13**(7): p. e1006780-e1006780.

117. Conant, G.C. and K.H. Wolfe, *Turning a hobby into a job: How duplicated genes find new functions*. Nature Reviews Genetics, 2008. **9**: p. 938.

118. El-Brolosy, M.A., et al., *Genetic compensation triggered by mutant mRNA degradation*. Nature, 2019. **568**(7751): p. 193-197.

119. Kawakami, K., *Tol2: a versatile gene transfer vector in vertebrates*. Genome Biol, 2007. **8 Suppl 1**: p. S7.

120. Shukla, G.C. and R.A. Padgett, *U4 small nuclear RNA can function in both the major and minor spliceosomes*. Proceedings of the National Academy of Sciences, 2004. **101**(1): p. 93.

121. McNamara, K.J., et al., *Transcriptional signals of a U4 small nuclear RNA gene.* Nucleic Acids Res, 1987. **15**(22): p. 9239-54.

122. Cheong, C.G., et al., *Crystal structures of human bifunctional enzyme aminoimidazole-4-carboxamide ribonucleotide transformylase/IMP cyclohydrolase in complex with potent sulfonyl-containing antifolates.* J Biol Chem, 2004. **279**(17): p. 18034-45.

123. Ali, A., K.P. Hoeflich, and J.R. Woodgett, *Glycogen synthase kinase-3: properties, functions, and regulation.* Chem Rev, 2001. **101**.

124. Ali, A., K.P. Hoeflich, and J.R. Woodgett, *Glycogen Synthase Kinase-3: Properties, Functions, and Regulation.* Chemical Reviews, 2001. **101**(8): p. 2527-2540.

125. Lindeman, L.C., et al., *Chromatin states of developmentally-regulated genes revealed by DNA and histone methylation patterns in zebrafish embryos.* Int J Dev Biol, 2010. **54**(5): p. 803-13.

126. Forde, J.E. and T.C. Dale, *Glycogen synthase kinase 3: A key regulator of cellular fate.* Cellular and Molecular Life Sciences, 2007. **64**(15): p. 1930-1944.

127. Green, M.R., *Eukaryotic Transcription Activation: Right on Target.* Molecular Cell, 2005. **18**(4): p. 399-402.

128. Yuan, C.-C., et al., *CHD8 Associates with Human Staf and Contributes to Efficient U6 RNA Polymerase III Transcription.* Molecular and Cellular Biology, 2007. **27**(24): p. 8729.

129. Urasaki, A., G. Morvan, and K. Kawakami, *Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition.* Genetics, 2006. **174**(2): p. 639-49.

130. Lin, X., L. Tirichine, and C. Bowler, *Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species.* Plant methods, 2012. **8**(1): p. 48-48.

131. Inoue, J., et al., *Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling.* Proceedings of the National Academy of Sciences, 2015. **112**(48): p. 14918.

132. Edger, P.P. and J.C. Pires, *Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes.* Chromosome Research, 2009. **17**(5): p. 699.

133. MacCarthy, T. and A. Bergman, *The limits of subfunctionalization.* BMC Evolutionary Biology, 2007. **7**(1): p. 213.

134. Assis, R. and D. Bachtrog, *Neofunctionalization of young duplicate genes in <em>Drosophila</em&gt.* Proceedings of the National Academy of Sciences, 2013. **110**(43): p. 17409.

135. Wade, A.A., et al., *Common CHD8 Genomic Targets Contrast With Model-Specific Transcriptional Impacts of CHD8 Haploinsufficiency*. Frontiers in Molecular Neuroscience, 2019. **11**(481).