Analysis of the Zebrafish *sox9b* Promoter:

Identification of Elements that Recapitulate Organ-Specific Expression of *sox9b*

Felipe R. Burns[a], Kevin A. Lanham[a,b], Kong M. Xiong[b,c,d], Alex J. Gooding[b],

Richard E. Peterson[a,b], and Warren Heideman[a,b]

Affiliations:

[a]Molecular and Environmental Toxicology Center, School of Medicine and Public Health, University of Wisconsin, 1300 University Avenue, Madison, WI 53706, USA

[b]Pharmaceutical Sciences Division, School of Pharmacy, University of Wisconsin, 777 Highland Avenue, Madison, WI 53705, USA

[c]Department of Biomolecular Chemistry, School of Medicine and Public Health, University of Wisconsin, 420 Henry Mall, Madison, WI 53705, USA

[d]Immunoassay Research and Development, Beckman Coulter Inc., 1000 Lake Hazeltine Dr., Chaska, MN 55318, USA

Corresponding author:

Dr. Felipe R. Burns
Telephone, Permanent: (608) 535-2155
Email: flpburns@gmail.com
Biopharmaceutical Services,
PPD Inc., 8551 Research Way #90
Middleton, WI 53562, USA

Author contributions

KMX, FRB, and AJG cloned different lengths of the promoter and made the expression constructs. FRB and KMX generated the various AB transgenic lines. FRB sequenced and conducted MatInspector and rVISTA analysis of the promoter. FRB took fluorescent images. FRB and KL designed and made the minimal promoter constructs. FRB injected and imaged transient expression analyses. FRB, KL, REP and WH wrote the manuscript with input from all authors.

Abbreviations

CCAAT, CAAT box; CNE, conserved non-coding elements; CPE, core promoter element; EGFP, enhanced green fluorescent protein; EGR2, early growth response protein 2; HIF1α, hypoxia inducible factor 1α; *SOX9,* SRY-related high-mobility box 9; TSS, transcriptional start site.

ABSTRACT

The SRY-related high-mobility box 9 (*SOX9*) gene is expressed in many different tissues. To better understand the DNA elements that control tissue-specific expression, we cloned and sequenced a 2.5 kb fragment lying 5' to the zebrafish *sox9b* gene transcriptional start site. Three regions of this clone contained stable secondary structures that hindered cloning, sequencing, and amplification. This segment, and smaller fragments, was inserted 5' of an EGFP reporter and transgenic fish were raised with the different reporters. Reporter expression was also observed in embryos directly injected with the constructs to transiently express the reporter. Heart expression required only a very short 5' sequence, as a 0.6 kb *sox9b* fragment produced reporter expression in heart in transgenic zebrafish, and transient experiments showed heart expression from a minimal *sox9b* promoter region containing a conserved TATA box and an EGR2 element (-74/+29 bp). Reporter expression in transgenic skeletal muscle was consistently lower than in other tissues. Jaw, brain, and notochord expression was strong with the full-length clone, but was dramatically reduced as the size of the fragment driving the reporter decreased from approximately 1.8 to 0.9 kb. The 2.5 kb region 5' of the *sox9b* contained 7 conserved non-coding elements (CNEs) that included putative hypoxia inducible factor 1α (HIF1α), CAAT box (CCAAT), early growth response protein 2 (EGR2), and core promoter elements. While a synthetic fragment containing all 7 CNEs produced some degree of reporter expression in muscle, jaw, heart and brain, the degree of reporter expression was considerably lower than that produced by the full length clone. These results can account for the tissue-specific expression of *sox9b* in the developing zebrafish.

INTRODUCTION

The SRY-related high-mobility box 9 (SOX9) transcription factor controls the development of a broad diversity of cell types in vertebrates. SOX9 is required for differentiation of glia cells (Stolt et al., 2003), heart valves (Akiyama et al., 2004), the notochord (Barrionuevo et al., 2006), and chondrocytes (Mori-Akiyama et al., 2003). In many cases, SOX9 is vital for multiple sequential steps during the course of tissue development and differentiation. As an example, *SOX9* is expressed early in development of primordial chondrogenic mesenchymal condensations, shortly after cranial neural crest cell migration, and is also required later in cartilage and bone growth (Bi et al., 2001; Akiyama et al., 2002).

The proper amount of *SOX9* is so vital for vertebrate gene expression that in mice, *Sox9* homozygous knockouts are lethal early in embryogenesis, and *Sox9* heterozygote knockouts die perinatally (Bi et al., 2001). Therefore, it is not possible to study *SOX9* function during later development with a simple mouse knockout model in certain vertebrates. Instead, conditional mice knockout models, preventing *Sox9* expression in specific cells and/or at specific times, are needed to study SOX9 protein function as development progresses (Bi et al., 2001; Akiyama et al., 2002).

As with mice, humans with only one functional copy of *SOX9* are haploinsufficient (Kwok et al., 1995). Heterozygous loss of *SOX9* function produces a syndrome known as campomelic dysplasia, characterized by cranial, musculoskeletal, heart, urogenital, and cognitive phenotypes. Many of these phenotypes, such as bowed limbs and decreased chondrogenesis, are thought to be caused by a disruption of extracellular matrix protein production.

SOX9 is a transcription factor that regulates transcription by binding to regulatory sequences near genes encoding extracellular matrix proteins and proteoglycans. A pulse in transcription of the related transcription factor SRY, encoded on the Y-chromosome, activates *SOX9* transcription in tissues destined to become testes. Thus, in humans *SOX9* is a critical target in male sexual differentiation.

The SOX9 protein has been found as a monomer, homodimerized with itself, or as a heterodimer with other SOX proteins (Bernard et al., 2003). An important role for these complexes is to increase production of extracellular matrix and cell membrane proteins (Lefebvre and de Crombrugghe, 1998) including aggrecan, a chondroitin sulfate proteoglycan, in cartilage derived cell lines (Sekiya et al., 2000) and *collagen 2a1* and *collagen 11a2* in chondrocytes (Bridgewater et al., 1998; Lefebvre and de Crombrugghe, 1998; Huang et al., 2000). Using RNA-seq, investigators discovered that *SOX9* regulates the expression of extracellular matrix proteins, remodeling enzymes, receptors, and transporters (Oh et al., 2014). While SOX9 can regulate target gene expression in a linear fashion by binding to and enhancing target gene transcription, its control is very intricate. Our comprehension of *SOX9* genetic and epigenetic regulation is ever increasing.

The control of *SOX9* by surrounding regulatory elements is complex. Chromosomal rearrangements and mutations in sequences over 1 Mb upstream can affect *SOX9* expression in humans, with minor to catastrophic effects on the developing fetus (Pfeifer et al., 1999). *SOX9* is often found in chromosomal regions referred to as gene deserts, with great non-coding spaces between the ORF and adjacent genes. In addition, numerous conserved non-coding elements (CNEs) have been found in

locations conserved across species (Bagheri-Fam et al., 2006). These CNEs often contain binding sites for transcription factors acting downstream of different of signaling pathways. Sonic and Indian hedgehog (GLI1), hypoxia factor 1α (HIF1), and bone morphogenic protein-2 (NF-Y [CCAAT-EBP]), signaling all interact with *SOX9* enhancers (Amarilio et al., 2007; Bien-Willner et al., 2007; Pan et al., 2008). Further, long-non-coding RNAs that surround the *SOX9* locus interact with the *SOX9* promoter and enhancers (Smyk et al., 2013).

Zebrafish are a useful model to study expression and regulation of *SOX9* due to their quick generation time, transparent external development, and the innovative molecular biological tools available. *SOX9* was duplicated during the teleost genome duplication event; therefore, zebrafish have two co-orthologs of *SOX9*: *sox9a* and *sox9b* (Yan et al., 2005). It is believed that with gene duplication the pair of genes originally shared completely overlapping functions and regulatory sequences, but have since diverged. This allows more specialization and the addition of new functions to one or the other of the pair. This is thought to be true of the *sox9a* and *sox9b* pair, with subfunction partitions, shared functions of *SOX9*, and divergent functions (Cresko et al., 2003).

As with other vertebrate *SOX9* genes, *sox9b* is located relatively far away from other genes in zebrafish. With the exception of non-protein-coding RNAs, the nearest neighboring protein coding gene, *sstr2*, is located more than 100,000 bp away from *sox9b* on chromosome 3 (Howe et al., 2013). Nonetheless, key regulatory elements are often not far from the 5' transcriptional start site of a gene. Interestingly, a zebrafish transgenic line with only 2450 base pairs (bp) of the *sox9b* regulatory region is able to

drive expression of a GFP reporter in a pattern that closely matches that of the endogenous *sox9b* mRNA (Plavicki et al., 2014).

This construct produces GFP expression in many cell types, including several known to require *Sox9* for development, including the midbrain and hindbrain, cardiomyocytes, and epicardial cells of the heart, vacuolated cells of the notochord, perichondrial cells and chondrocytes of Meckel's cartilage in the jaw, skeletal muscle and other tissues during development (Plavicki et al., 2014). Therefore, that short region of approximately 2.5 kb from *sox9b* carries elements that can produce tissue-specific transcription.

Here we show that elements within this region are conserved across different fish *sox9b* genes, and that fragments containing these different conserved non-coding elements can recapitulate tissue-specific expression in zebrafish. Additionally, we highlight secondary structure near the *sox9b* locus that can hinder cloning, sequencing, and genotyping reactions. This elucidates likely elements modulating *sox9b* expression and will be useful in making mutant variations to study tissue-specific *sox9b* function and expression, as well as in developing tissue-specific reporters.

MATERIALS and METHODS

Zebrafish

We used the AB zebrafish strain expressing the various transgenes, as indicated. Zebrafish were maintained as described previously (Westerfield, 2000). Briefly, adults were housed in 38L glass aquaria with recirculating reverse osmosis water supplemented with Instant Ocean Sea Salts (fish water, 60 mg/L; Aquarium Systems,

Mentor, OH) at about 27°C. This water was filtered through biofilter media, activated charcoal and an ultraviolet light sterilizer. Water used to rear embryos was supplemented with Instant Ocean Sea Salts (60 µg/ml) and methylene blue (50 µM). Powdered brine shrimp flakes (Aquatic Eco-systems Inc., Apopka, FL) and TetraMin Tropical flakes (Tetra Holding, Blacksburg, VA) were fed to adult zebrafish twice daily. Before imaging embryos or larvae, they were euthanized by tricaine overdose (Westerfield, 2000). Adhering to the National Institutes of Health's "Guide for the Care and Use of Laboratory Animals," all procedures involving zebrafish were approved by the Animal Care and Use Committee of the University of Wisconsin-Madison.

Cloning and amplifying *sox9b* genomic fragments

Templates and primers. We used the zebrafish genome build Zv7 to design PCR primers. A full list of primers used for cloning, sequencing, and secondary structure analysis is provided in Supplemental Table 1. Using the build information, we amplified fragments extending in the 5' direction from near the transcriptional start site of *sox9b* mRNA (Supplemental Figure 1C) at +29 bp from the transcriptional start site of *sox9b* on Chromosome 3 in Zv9 (Cunningham et al., 2015). We have recently published the complete sequence obtained along with a complete description of overlaps with builds Zv7 and Zv9 (Plavicki et al., 2014). This, along with Supplemental Figure 1C allows one to determine the precise boundaries of the fragments produced relative to the longest cloned sequence and relative to the different build overlaps.

PCR Conditions. Genomic DNA was extracted from zebrafish embryos or larvae (Westerfield, 2000) and diluted 1:10 (9.2 ng/µl) or 1:100 (0.92 ng/µl) for use as a PCR template. Final concentrations of the following reaction mixture were used: 2.5 mM of

$MgCl_2$ (LA PCR™ Buffer II, TaKaRa La Taq™), 0.4 mM of each dNTP (TaKaRa La Taq™, Japan), 0.2 µM each of the upstream and downstream primers (see Supplemental Table 1, IDT®), 1.25 units of LaTaq™ (TaKaRa La Taq™, Clontech, RR002A, Japan) and 1 µl of PCR template 1:10 (9.2 ng/µl) or 1:100 (0.92 ng/µl). PCR amplification conditions with annealing times of 20-30 s and 1 min/kb extension times were considered traditional methods as per manufacturers' protocols. PCR amplifications with annealing times of 2.5 m and extension times of 1.5 min/kb were considered the modified method.

Sub-Cloning. The various PCR products were inserted into the pCR™ II-TOPO® TA cloning ® kit vector for propagation (ThermoFisher Scientific, 4600-40, Grand Island, NY) following the manufacturer's protocol. The various *sox9b* promoter and recombinant clones were then inserted into the pDestTol2pA2 Gateway® expression vector driving EGFP (Kwan et al., 2007) following standard and manufacturer's protocols (Thermo Fisher Scientific, Grand Island, NY).

Sequencing *sox9b* genomic fragments

Clones were sequenced at the UW Biotechnology Center with BigDye® Terminator v3.1 Cycle Sequencing kit (ThermoFisher Scientific, 4337455, Grand Island, NY) following the manufacturer's protocol (Sanger et al., 1977; Smith et al., 1986). In some cases, a modified method was developed in consultation with the UW Biotechnology Center for sequencing through GT-rich repeats. This incorporated dideoxy GTP (dGTP, BigDye® Terminator v3.0 Ready Reaction Cycle Sequencing Kit, ThermoFisher Scientific, 4390229, Grand Island, NY) instead of the typical dITP, spiked

to the BigDye® reaction mixture at a ratio of 3 to 1, along with a denaturation step prior to setting up the sequencing reactions, and modified thermocycling times and temperatures. The sequences were aligned using CodonCode Aligner (CodonCode Corporation, www.codoncode.com).

Our initial efforts produced a 2,450 bp fragment of the *sox9b* promoter, and by using techniques that allow PCR through GT repeats, we were able to clone out sequentially smaller fragments from AB zebrafish chromosomal DNA. All of the clones that we produced were anchored with their 3' ends extending 29 bp into the transcribed portion of *sox9b* (sox9b009 mRNA, Ensembl Zv9 version) (Cunningham et al., 2015), indicating that these independently obtained and overlapping sequences truly represent the DNA immediately upstream of *sox9b*. In addition, primer pair "l" produced amplicons of the predicted length from wild type AB zebrafish chromosomal DNA that extended from the -26 position far into the first exon of *sox9b* (+491), confirming that our clones are indeed located immediately 5' of the transcriptional start of *sox9b*.

Using traditional PCR conditions we found that some PCR amplicons, later found to contain tracts of GT repeats, yielded faint or no bands (Supplemental Figure 1C, c, f, k). A close look at the chromatographs show that Sanger sequencing reactions failed to produce products through bps -1329 to -1341 in the reverse direction. The addition of dGTP, DMSO, and TEsl to the sequencing reactions resulted in PCR products proceeding through regions of secondary structure (Supplemental Figure 1F). The sequence shows an inverted repeat GT sequence located at (-1407/-1395 and -1329/-1341) and two other GT rich regions at (-1985/-1874) and (-164/-51). These three regions are in red colored font in Figure 1.

sox9b:*EGFP* transgenic and transient embryos

Reporter constructs were injected into 1-2 cell stage AB embryos and injected embryos were raised to adulthood. The injection conditions and concentrations of plasmids and mRNA were the same as detailed previously (Lanham et al., 2014). Adult founder fish were identified by outcrossing potential founders to wild type AB fish and screening offspring for EGFP expression. EGFP positive embryos were raised to produce the F1 line. EGFP-positive F1 adults were outcrossed to wild type AB fish and screened to produce the F2 line. Images shown are from representative F1 and F2 embryos and larvae. Transient expression analysis was conducted by injecting the various reporter constructs into the 1-2 cell stage of AB embryos. The *sox9b* clones were inserted upstream of EGFP in the Tol2 transgenesis system as previously described to make reporter constructs (Kwan et al., 2007). The plasmids underwent restriction digest analysis and were also sequenced to verify that proper lengths of the *sox9b* promoter were constructed. The lines were genotyped using a primer specific to the attB4 site (F' genotype) and a unique sequence in the *sox9b* promoter (Supplemental Table 1, R').

Organ-specific expression incidence

Incidence of occurrence for transgenic expression was calculated as the expression of EGFP in a particular tissue in a particular transgenic line divided by the total number of transgenic lines, for a particular length of an injected *sox9b* promoter. Incidence of transient expression was calculated for notochord, heart, skeletal muscle, jaw, and brain, respectively, as number of injected fish expressing EGFP in that organ, divided by the total number of fish injected with the *sox9b* promoter construct, both

survived the injection and expressed EGFP in at least one organ. Scoring for jaw expression was based entirely on expression in Meckel's cartilage.

Conserved noncoding element analysis

All putative transcription factor binding sites provided were analyzed by MatInspector (v8.0) (Cartharius et al., 2005). Inter- and intraspecies CNE analysis was conducted using rVista v2.0 analysis (Loots and Ovcharenko, 2004) and visualized with VISTA (Mayor et al., 2000). The criteria for conservation were 70% match over 100 base pairs. MatInspector was used with default settings. Minimal promoter and enhancers were constructed using IDT's gBlocks ® (www.idtdna.com/site, Coralville, IA). The minimal promoter and enhancers were constructed from the conserved non-coding element sequences found using rVISTA and MatInspector analysis of the -2421/+29*sox9b* region and contained appropriate attB sites for recombination and cloned into p5E 5' entry vector of the Tol2 transgenesis system. The sequences are provided in Supplemental Figure 1. Supplemental Figure 2 show's that the core promoter elements in blue are comprised of the putative TATA, NKX3-2, SMARCA, and CPEBP elements, which are also coded blue in the representation in Figure 1b for convenience.

Image acquisition and analysis

Transgenic and transient expression images of larvae were taken with an Olympus DP72 digital camera on an Olympus S2X16 microscope with a GFP filter (excite 470nm; emit 525nm). All images are displayed with the following conditions: 4x total magnification, 288 dpi, 1000 ms, ISO 400 (Olympus), brightness 150, contrast 100 (Adobe Photoshop).

RESULTS

A set of *sox9b* promoter fragments containing conserved non-coding elements.

We have previously reported transgenic lines carrying a *sox9b*:EGFP reporter construct made from the longest upstream fragment that we cloned, -2421/+29 (Burns et al., 2015; Plavicki et al., 2014). This construct clearly showed expression in brain, heart, skeletal muscle myocytes, notochord and the jaw cartilages. The pattern of EGFP reporter mRNA expression closely matched that of the normal *sox9b* mRNA, indicating that we had cloned enough of the regulatory sequences of *sox9b* to produce a reporter that matches the normal expression pattern. With this in hand, we set out to determine whether sub-sections of our approximately 2.5 kb clone could produce more restricted expression patterns, providing clues about whether or not different sequences controlled expression in the different tissues. We approached this by making deletions of the *sox9b* sequence from the 5' end.

To determine the regions of the *sox9b* promoter that can produce tissue-specific expression, varying lengths of the -2421/+29 *sox9b* promoter and enhancer region were amplified as shown in Figure 1A. These segments were then coupled to EGFP using the pDestTol2pA2 expression vector for production of transgenic lines.

In addition, we used rVISTA (Figure 1B) to identify conserved non-coding elements in common between the zebrafish -2421/+29 region and the corresponding 5' regions of *sox9* in cavefish, medaka, stickleback, gar, and coelacanth. As lobe-finned fish, gar and coelacanth have diverged prior to the teleost genome duplication, and each carry only one copy of *sox9*. The medaka *sox9a* gene is similar to the

developmental role of zebrafish *sox9b* in the skeleton (Dong et al., 2012). The position of these elements across the different species, and how they are systematically removed in the shorter clones is shown in Figure 1A and B. The positions of these elements relative to the entire -2421/+29 sequence and GT repeats are shown in figure 1C. As can be seen, not all of the elements are conserved between all of the different species pairs. We selected a group of 7 CNEs to investigate further.

Conserved non-coding sequences are believed to encode important information for the regulation of transcription. MatInspector revealed putative transcription factor sites in the CNEs discovered by rVISTA. The core promoter elements outlined in blue (Figure 1B and C), spanning positions -32 to -4, contained binding motifs for core promoter element binding protein (CPEBP) binding sequence, TATA binding protein (TATA), NK3 homeobox 2 (NKX3-2/BAPX), and SWI/SNF related actin dependent regulator A3 (SMARCA3).

A more distal highly conserved region, from -74 to -35, outlined in green, contained putative binding elements for early growth response protein 2 (EGR2), progesterone receptor (PRE1), and KRAB zinc finger 1 (ZF01). EGR has been shown to bind at many core promoters and may be necessary for other enhancer element activity (Zaiou et al., 1998).

The conserved sequence, from -174 to -160 highlighted in brown, was found to contain a putative NFY/CCAAT box element (NFY/CCAAT) and a growth factor independence 1 element (GFI1) in brown. Within the pink highlighted CNE, from -383 to -332, we identified a putative CAAT/CCAAT box (CCAAT), linked closely to putative PREB, p53, OCT1, TCF11, IRX5, MEIS/HOX9A, OCT 3/4, FAST1, and HFH8 elements.

At the other end of our largest clone, positions -1648 to -1628 highlighted in orange, we found a small CNE containing putative ARID5a and SP1 elements. We also found a CNE, at -2135 to -2103, highlighted in grey that was most closely conserved between zebrafish and cavefish. This contained putative binding sites for ATF, HIF1α, AHRR, and XBP1.

Expression analysis of increasingly shorter lengths of the *sox9b* promoter

To determine the regions of the *sox9b* promoter that can produce tissue-specific expression, varying lengths of the -2421/+29*sox9b* promoter and enhancer region were cloned into the pDestTol2pA2 expression vector driving EGFP expression (Figure 2A). Different founders were established for each length and founders were denoted with an alphabetical identifier (A,B,C,D, or E). The shorter the length of the promoter, the more difficult it was to identify founders due to decreased expression of EGFP. Although for some constructs we obtained limited numbers of founders, the expression pattern for each construct was consistent. In other words, the differences in pattern that we did observe were between fish carrying different constructs rather than between individual founders injected with the same construct: the expression patterns appeared to be sequence-dependent.

Transgenic 96 hpf larvae were imaged under identical conditions, and the expression patterns were compared (Figure 2B). The -2421/+29, -1884/+29 and -1816/+29 lines all produced expression in brain, skeletal muscle, notochord, heart, and jaw. The -910/+29 transgenic line showed reduced expression that was confined primarily to the heart and to a lesser extent the brain. In some instances, the -910/+29 transgenic line showed expression in skeletal muscle, and failed consistently to produce

GFP in the notochord expression. The -562/+29 transgenic line produced GFP expression that was confined to the heart.

Compilations of the results from the transgenic reporter lines are shown in Figure 3 (see also Figure 6). Deletions from the 5' end showed that different regions drive expression of the *sox9b* reporter in different tissues. For example, the originally bright notochord expression incidence decreased dramatically from 100% to 50% as the 5' end was shortened from -1884 to -1816 and was completely absent in shorter constructs. In contrast, the step from -1884 to -1816 did not reduce GFP expression in brain and heart.

In some cases we observed reduced expression of GFP in the jaw and heart for the -1884/+29 transgenic lines compared with either the longer or shorter lines (Fig 3). We do not know whether this represents positional effects for the individual founders, or a real drop in expression. One possibility is that this construct reveals a negative element that is balanced by upstream positive elements in the longer fragments and then is gone in the shorter ones. While this could explain the result, the fact that we see expression in jaw and heart in some instances with this line suggests the former explanation of founder effects.

Interestingly, the incidence of expression in the jaw was highest in the -2421/+29 and gradually decreased as the lengths got shorter. Shortening to -910/+29 left only heart and very limited brain and jaw expression. Heart expression persisted even with the -562/+29 fragment, indicating that some sequence even closer to the transcriptional start produces heart expression in these constructs. This shows the presence of

specific regions directly upstream of *sox9b* that recapitulate expression in different tissues.

Transient *sox9b*:EGFP injection.

To streamline analysis, and to provide an alternative method of analysis, we turned to transient expression experiments in which we examined embryos directly injected with the test construct.

In general, the transient injection results were consistent with the experiments with stable transgenic lines (Figures 3 and 6). As with the stable transgenics, jaw, brain, and notochord expression was largely absent with the -910/+29 and shorter constructs; however, could be observed with a new 1386/+29 construct. Similarly, heart and muscle expression persisted from the -910/+29 construct in both transgenic and transient experiments, and heart expression was observed from the -562/+29 fragment in both transient and stable experiments. However, with muscle there was some discrepancy in that the stable lines showed no muscle expression from the shortest piece, -562/+29, while this construct produced some muscle expression in the transient experiments.

In addition the transient experiments showed muscle expression in what were clearly cardiomyocytes. This is consistent with our previous report using the longest clone in a stable transgenic. When comparing the stable and transient experiments, the stable transgenics produced a signal that tended to be stronger at the somite boundaries. This signal was present across the entire somite, rather than restricted to individual myocytes as shown with the transients. Some of the difference can be attributed to the imaging in different experiments, as well as the relatively greater

brightness in the stable lines. However, it appears that the stable lines express GFP in cell types other than myocytes in the muscle; cell types that we have not identified. These may be neurons or blood vessels. Without this background fluorescence from other cells, the myocytes stand out more distinctly in the transient experiments. Further, gain and image acquisition settings in this method were different then described previously.

Notochord and brain reporter expression dropped when comparing the full-length construct with the -1884/+29 construct in transient experiments (Fig 5). This may indicate that the transient experiments are more sensitive to losses of elements in the constructs.

Interspecies conserved non-coding elements (CNEs) near *sox9b* is sufficient for eGFP expression in skeletal muscle and heart

The conservation of non-coding elements between species suggests a functional role for these elements. To test this, we first assembled a GFP reporter containing the very 3' portion of the *sox9b* clone containing the CPEBP, TATA, and EGR2 elements as a minimal promoter to which we could add different elements. These elements are defined in Figure 1 relative to the entire clone, and in Supplementary Figure 2. A second construct was made containing the minimal promoter along with the other 5 CNEs identified in Figure 1 (Figure 5B). These two constructs were used in transient expression experiments in which fertilized embryos were transformed with these constructs at the 1 and 2 cell stage and visualized at 96 hpf to score them for expression.

In these experiments, with mosaic expression and the plasmid not evenly distributed throughout the developing fish (Supplemental Figure 3), we used incidence of GFP signal to score expression. The full-length -2421/+29 construct as a positive control yielded 68% of the injected fish expressing EGFP in heart, 77% expressed EGFP in skeletal muscle, 73% in notochord, 36% in the jaw, and 68% in the brain (Figure 3C; n = 22).

In contrast, *sox9b* CPE constructs with just the minimal *sox9b* promoter produced expression in the heart in 83% of the injected individuals (n = 12) and 67% incidence in skeletal muscle, but no signs of expression in notochord (0%) and brain (0%). There was only a low incidence (17%) of expression in the jaw (Figure 3C).

When the additional 5 CNEs were coupled to the CPE construct, expression in muscle brain and jaw was partially restored: of 12 injected fish scored, reporter expression was found in skeletal muscle (67%), heart (50%), and jaw (33%). This construct produced little or no incidence of expression in notochord (17%) and brain (0%). Figure 6 summarizes the results produced with transient expression experiments.

DISCUSSION AND CONCLUSION

The results reported are the summation of many experiments with different lengths of 5' *sox9b* sequence driving a GFP reporter. We used both stable transgenic lines as well as transient expression. The stable lines can suffer from positional insertion effects. Here we report patterns that we saw in common between the offspring of at least 2 founders for each construct. Despite the possibility of positional effects, we found that the expression patterns from a given construct varied relatively little: the

variation between different constructs far outweighed variations we saw within a group of founders with the same construct. For both the transient and stable experiments we report patterns we observed in common for a given construct.

We used transient expression as a means to gather more data using a different approach less plagued by position of insertion. On the other hand, we found that we had to compensate for mosaic expression with the transient experiments. An example of this is shown in Supplemental Figure 3. Nonetheless, we were gratified to find that in general the results we obtained with the transient experiments were consistent with those observed with the stable transgenic experiments. Unless otherwise indicated the results presented are representative of what we observed consistently among animals carrying the same construct. Because of the variation observed, and the discrepancies we have described, some of the results have been presented with qualifiers such as +/-, or comments about where experiments diverge from each other in conclusion.

***Sox9b* Regions Required for Tissue-Specific Expression.**

Notochord expression dropped as the 5' end was shortened from -1884 to -1816 and absent in -910/+29 transgenics. Analysis of the transient data shows some expression in the notochord with injection of the -1386/+29 fragment, and confirms absence of notochord expression from the -910/+29 construct. These results are consistent with more than one notochord-specific element lying within the -1884 to -910 interval: notochord expression was reduced but not absent after removing the segment from -1884 to -1816. This indicates that one region needed for peak notochord expression in these conditions lies within that interval. As the deletions progressed further to -910, the remaining notochord expression disappeared, indicating the loss of

an additional element with a role in notochord expression. Such a model explains the incremental drops in reporter expression as the clone was shortened.

Expression in brain followed a similar pattern suggesting two regulatory sites lost as we progressed from the strong signal that was a hallmark of the longer -2421/+29 construct to the reduced signal produced by the -1816 construct. Brain expression was largely decreased when the length was reduced from -1816 to -910.

Expression in jaw was retained even as the 5' end was reduced to -1816, but then disappeared with the constructs were shortened from -1816 to -910 in the transgenic lines (Figure 2 and 3). Some limited brain and jaw expression was observed in transient experiments with the -1386/+29 construct. Altogether, this suggests expression in the jaw requires a sequence located between -1816 and -1386.

Expression of EGFP in the heart was detectable in all transgenic lines, including -562/+29 (Figure 2 and Figure 3). This indicates that expression in skeletal muscle and heart is enhanced either by a single shared element or by distinct, tissue-specific elements that both lie close to each other near the transcriptional start site.

In transgenics carrying even the full-length clone, the EGFP signal in skeletal muscle was lower in intensity than in other tissues. Although *SOX9* is expressed in skeletal muscle (Nie, 2006; Plavicki et al., 2014), the lower expression in this tissue compared to other tissues is consistent with the finding that ectopic over expression of *SOX9* drastically inhibits myocyte growth (Schmidt et al., 2003). Thus, muscle expression may normally be lower than in other tissues.

Multinucleated skeletal muscle cells have the potential to amplify non-specific expression. We cannot rule out some level of non-specific signal. However, the pattern

of skeletal muscle expression that we observed, in which individual muscle cells light up, while adjacent myocytes express low levels of reporter was consistently observed. This pattern was observed with the full-length -2421/+29 reporter in multiple stable transgenic lines, as well as in the transient experiments. A non-specific mechanism would be expected to produce expression in the entire region. In addition, we saw a similar pattern of individual myocyte expression with the stable full-length transgenic founders. Although harder to see because of higher overall expression in muscle, the pattern is the same. This is easier to see in previously published images from Plavicki *et al*, showing individual myocytes in the caudal fin (Plavicki et al., 2014). Furthermore in previous work with transient and stable transgenic reporters using the same plasmid backbone and other promoters such as *cmlc2* we have seen no myocyte expression even at high exposure (Lanham et al., 2014).

Specific Elements and Tissue-Specific Expression.

We constructed an EGFP reporter composed of the sequence from -74 to 0 to represent a minimal promoter. Although we have no evidence for binding, this region contains putative EGR2, TATA, and CPEBP binding sites (Figure 1). Transient experiments with this short segment produced heart expression in 83% of injected fish (Figure 5 and 6). As with the experiments using longer fragments, we saw both heart and skeletal muscle expression with this short basal construct. However, with the shorter fragments observed no reporter expression in the brain, jaw, or notochord.

Despite the fact that it contains a putative NKX3-2 element (Zeng et al., 2002), the minimal *sox9b* promoter produced a low incidence of EGFP expression in the Meckel's cartilage of the jaw. This is consistent with the low incidence of transgenic and

transient expression in the jaw and the brain using the -562/+29*sox9b*:EGFP construct, which also contains this putative NKX3-2 element.

Inclusion of the additional CNEs including the hypoxia and ARID5A elements increased expression in the jaw and brain (Figure 5C). However, even when all 7 CNEs were present, the reporter derived from individual elements did not produce the expression seen with the full-length clone. This indicates that other elements besides our 7 CNEs or physical distances between CNEs and perhaps other factors are important to recapitulate the expression seen with the longer length reporters. As indicated, we have no evidence showing the binding of any specific protein to the cloned sequences.

***sox9b* 5' Sequence.**

We have several reasons for believing that our sequence accurately reflects the endogenous zebrafish *sox9b* promoter region in our AB strain. We produced long amplicons anchored into the first exon of *sox9b* (Supplemental Figure 1C). These same regions are able to recapitulate *sox9b* expression in numerous tissues consistent with expression patterns seen with endogenous *sox9b* (23). We also successfully cloned increasingly shorter lengths that perfectly overlapped the original longer clones (Supplemental Figure 1C).

As indicated, the *sox9b* promoter region that we have cloned varies significantly from what is reported in the current zebrafish genome build, GRCz10/danRer10. We note the possibility of strain differences in sequence: our sequence was from an AB strain cultured in our lab for more than a decade, while the Ensembl sequence is based on a sophisticated Sanger-AB-Tubingen (SAT) strain. In addition, *sox9b* is telomeric.

The discrepancies are not too surprising as the sequence 5' of *sox9b* changed steadily across different zebrafish genome builds: Zv6, Zv7, Zv8 (Howe et al., 2013; Plavicki et al., 2014). Accuracy of genome builds can be compromised by secondary structure. When sequence reactions terminate at long GT repeat regions, it can be difficult for algorithms to construct accurate builds containing these areas. Long repeats match any region carrying the same repeat, potentially placing them in locations other than the true genomic match.

Our discovery of stable secondary structures very near the transcriptional start site of *sox9b* and development of methods to work through these should be helpful to other investigators working with *sox9b*. Sequencing reactions that can proceed through these secondary structures are helpful in determining what sequence lies on the other side of the GT repeat domain and can assist in build fidelity, but are not always practical for large scale genomic sequencing projects to accomplish. Without our efforts beginning with primers designed from an earlier build, the anomaly would never have been detected. Our results further highlight the need for individual labs to contribute to the genome assembly for their model organism and region of interest.

ACKNOWLEDGEMENTS

We thank Dorothy Nesbit for animal husbandry and Glen Threadgold from the Wellcome Trust Sanger Institute for discussions involving our sequence. We thank Jeremy Niece of the University of Wisconsin Biotech Sequencing Center for developing methods to sequence through secondary structure. We also thank Andrew Sentkowski and Chelsea O'Driscoll for assistance with visual data collection and plasmid isolation. We thank Drs. Henry G. Tomasiewicz, Michael Taylor, and Chunyue Yin for helpful suggestions and discussions involving the sequence and expression patterns. This research was supported by NIH grant R01 ES012716 and University of Wisconsin Sea Grant Institute, National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce grant number NA 16RG2257, Sea Grant Project Numbers R/BT-17, R/BT-20 and R/BT-22 (KL, KMX, WH & REP). FRB was supported by a supplement to NIH grant R01 ES012716 for underrepresented minorities and by an Advanced Opportunity Fellowship from the University of Wisconsin, Science and Medical School, Graduate Research Scholars Program and by Brooks and Alison. The Hilldale Undergraduate Fellowship was awarded to and provided support for AJG. The supportive funding took no part in study design, data collection, analysis, interpretation, writing the manuscript, and in the decision to submit the manuscript for peer review.

References

Akiyama, H., Chaboissier, M.C., Behringer, R.R., Rowitch, D.H., Schedl, A., Epstein, J.A. and de Crombrugghe, B., 2004. Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. Proc Natl Acad Sci U S A 101, 6502-7.

Akiyama, H., Chaboissier, M.C., Martin, J.F., Schedl, A. and de Crombrugghe, B., 2002. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Genes Dev 16, 2813-28.

Amarilio, R., Viukov, S.V., Sharir, A., Eshkar-Oren, I., Johnson, R.S. and Zelzer, E., 2007. HIF1alpha regulation of Sox9 is necessary to maintain differentiation of hypoxic prechondrogenic cells during early skeletogenesis. Development 134, 3917-28.

Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Gunther, T., Schule, R., Kemler, R., Mallo, M., Kanzler, B. and Scherer, G., 2006. Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. Dev Biol 291, 382-97.

Barrionuevo, F., Taketo, M.M., Scherer, G. and Kispert, A., 2006. Sox9 is required for notochord maintenance in mice. Dev Biol 295, 128-40.

Bernard, P., Tang, P., Liu, S., Dewing, P., Harley, V.R. and Vilain, E., 2003. Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. Hum Mol Genet 12, 1755-65.

Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R. and de Crombrugghe, B., 2001. Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci U S A 98, 6698-703.

Bien-Willner, G.A., Stankiewicz, P. and Lupski, J.R., 2007. SOX9cre1, a cis-acting regulatory element located 1.1 Mb upstream of SOX9, mediates its enhancement through the SHH pathway. Hum Mol Genet 16, 1143-56.

Bridgewater, L.C., Lefebvre, V. and de Crombrugghe, B., 1998. Chondrocyte-specific enhancer elements in the Col11a2 gene resemble the Col2a1 tissue-specific enhancer. J Biol Chem 273, 14998-5006.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. and Werner, T., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21, 2933-42.

Cresko, W.A., Yan, Y.L., Baltrus, D.A., Amores, A., Singer, A., Rodriguez-Mari, A. and Postlethwait, J.H., 2003. Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. Dev Dyn 228, 480-9.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Giron, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Kahari, A.K., Keenan, S., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Aken, B.L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S.M., Spudich, G., Trevanion, S.J., Yates, A., Zerbino, D.R. and Flicek, P., 2015. Ensembl 2015. Nucleic Acids Res 43, D662-9.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J.C., Koch, R., Rauch, G.J., White, S., Chow, W., Kilian,

B., Quintais, L.T., Guerra-Assuncao, J.A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S.F., Laird, G.K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Elliot, D., Threadgold, G., Harden, G., Ware, D., Begum, S., Mortimore, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Lloyd, C., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., et al., 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498-503.

Huang, W., Zhou, X., Lefebvre, V. and de Crombrugghe, B., 2000. Phosphorylation of SOX9 by cyclic AMP-dependent protein kinase A enhances SOX9's ability to transactivate a Col2a1 chondrocyte-specific enhancer. Mol Cell Biol 20, 4149-58.

Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P. and Chien, C.B., 2007. The Tol2kit: A multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn 236, 3088-3099.

Kwok, C., Weller, P.A., Guioli, S., Foster, J.W., Mansour, S., Zuffardi, O., Punnett, H.H., Dominguez-Steglich, M.A., Brook, J.D., Young, I.D. and et al., 1995. Mutations in SOX9, the gene responsible for Campomelic dysplasia and autosomal sex reversal. Am J Hum Genet 57, 1028-36.

Lanham, K.A., Peterson, R.E. and Heideman, W., 2014. Expressing a constitutively active AHR mutant in cardiomyocytes: replication of dioxin toxicity. Tox Sci, in revision.

Lefebvre, V. and de Crombrugghe, B., 1998. Toward understanding SOX9 function in chondrocyte differentiation. Matrix Biol 16, 529-40.

Loots, G.G. and Ovcharenko, I., 2004. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. Nucleic Acids Res 32, W217-21.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S. and Dubchak, I., 2000. VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16, 1046-7.

Mori-Akiyama, Y., Akiyama, H., Rowitch, D.H. and de Crombrugghe, B., 2003. Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. Proc Natl Acad Sci U S A 100, 9360-5.

Nie, X., 2006. Sox9 mRNA expression in the developing palate and craniofacial muscles and skeletons. Acta Odontol Scand 64, 97-103.

Oh, C.D., Lu, Y., Liang, S., Mori-Akiyama, Y., Chen, D., de Crombrugghe, B. and Yasuda, H., 2014. SOX9 regulates multiple genes in chondrocytes, including genes encoding ECM proteins, ECM modification enzymes, receptors, and transporters. PLoS One 9, e107577.

Pan, Q., Yu, Y., Chen, Q., Li, C., Wu, H., Wan, Y., Ma, J. and Sun, F., 2008. Sox9, a key transcription factor of bone morphogenetic protein-2-induced chondrogenesis, is activated through BMP pathway and a CCAAT box in the proximal promoter. J Cell Physiol 217, 228-41.

Pfeifer, D., Kist, R., Dewar, K., Devon, K., Lander, E.S., Birren, B., Korniszewski, L., Back, E. and Scherer, G., 1999. Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. Am J Hum Genet 65, 111-24.

Plavicki, J.S., Baker, T.R., Burns, F.R., Xiong, K.M., Gooding, A.J., Hofsteen, P., Peterson, R.E. and Heideman, W., 2014. Construction and characterization of a sox9b transgenic reporter line. Int J Dev Biol 58, 693-9.

Sanger, F., Nicklen, S. and Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74, 5463-7.

Schmidt, K., Glaser, G., Wernig, A., Wegner, M. and Rosorius, O., 2003. Sox8 is a specific marker for muscle satellite cells and inhibits myogenesis. J Biol Chem 278, 29769-75.

Sekiya, I., Tsuji, K., Koopman, P., Watanabe, H., Yamada, Y., Shinomiya, K., Nifuji, A. and Noda, M., 2000. SOX9 enhances aggrecan gene promoter/enhancer activity and is up-regulated by retinoic acid in a cartilage-derived cell line, TC6. J Biol Chem 275, 10738-44.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B. and Hood, L.E., 1986. Fluorescence detection in automated DNA sequence analysis. Nature 321, 674-9.

Smyk, M., Szafranski, P., Startek, M., Gambin, A. and Stankiewicz, P., 2013. Chromosome conformation capture-on-chip analysis of long-range cis-interactions of the SOX9 promoter. Chromosome Res 21, 781-8.

Stolt, C.C., Lommes, P., Sock, E., Chaboissier, M.C., Schedl, A. and Wegner, M., 2003. The Sox9 transcription factor determines glial fate choice in the developing spinal cord. Genes Dev 17, 1677-89.

Westerfield, M., 2000. The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 4th ed. Univ. of Oregon Press, Eugene, OR.

Yan, Y.L., Willoughby, J., Liu, D., Crump, J.G., Wilson, C., Miller, C.T., Singer, A., Kimmel, C., Westerfield, M. and Postlethwait, J.H., 2005. A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132, 1069-83.

Zaiou, M., Azrolan, N., Hayek, T., Wang, H., Wu, L., Haghpassand, M., Cizman, B., Madaio, M.P., Milbrandt, J., Marsh, J.B., Breslow, J.L. and Fisher, E.A., 1998. The full induction of human apoprotein A-I gene expression by the experimental nephrotic syndrome in transgenic mice depends on cis-acting elements in the proximal 256 base-pair promoter region and the trans-acting factor early growth response factor 1. J Clin Invest 101, 1699-707.

Zeng, L., Kempf, H., Murtaugh, L.C., Sato, M.E. and Lassar, A.B., 2002. Shh establishes an Nkx3.2/Sox9 autoregulatory loop that is maintained by BMP signals to induce somitic chondrogenesis. Genes Dev 16, 1990-2005.

Figure Legends

Figure 1. Conserved non-coding element analysis. (A) Representation of five cloned *sox9b* promoter lengths used in stable transgenic experiments. **(B)** CNEs identified with the putative enhancer elements as indicated. **(C)** Sequence of longest cloned *sox9b* fragment. Repetitive sequences and regions of secondary structure are shown as red font. The CNEs are indicated by boxes shaded to match the panels above.

Figure 2. Expression constructs and summary of expression. (A) Expression constructs with different lengths of the *sox9b* promoter. The founders obtained for each segment are designated as lines A,B,C,D, or E. **(B)** Lateral images of 96 hpf zebrafish from each stable transgenic line, using confocal microscopy showing EGFP reporter expression as green fluorescence. Abbreviations: brain, b; heart, h; jaw, j; skeletal muscle, m; and notochord, n. The + indicates consistently observed; the - indicates consistently not observed; +/- indicates sporadic or weak occurrence in offspring.

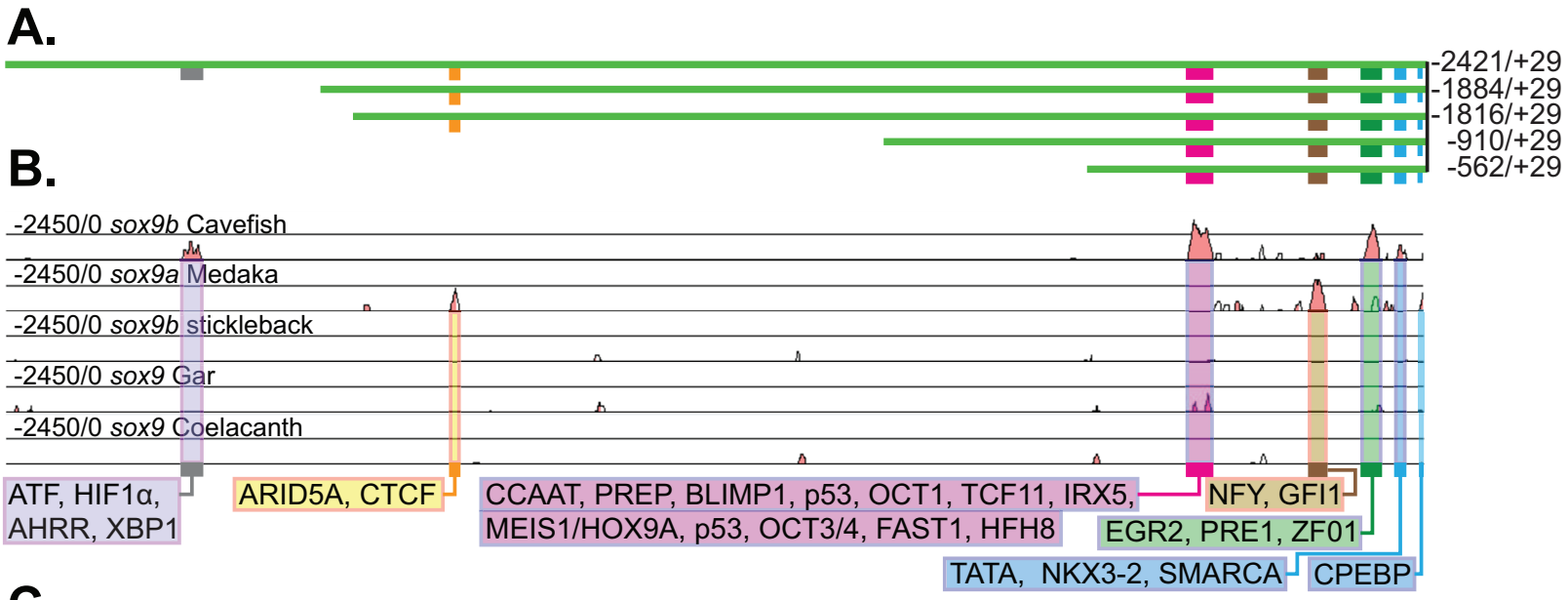
Figure 3. Schematic of stable transgenic results. Graphic representations of *sox9b:*EGFP expression are shown for each heart, skeletal muscle, notochord, brain, and jaw as indicated. The X-axis indicates the length of the different fragments tested. The Y axis shows a qualitative assessment of all data collected. For reference, the locations of the CNEs are shown at the bottom of the figure.

Figure 4. Transient expression analysis of different lengths 5' of *sox9b*. Lateral images of 96 hpf zebrafish from transient experiments using the indicated promoter sequences. Newly fertilized eggs were injected as described in the Methods, and images were recorded using confocal microscopy showing EGFP reporter expression as green fluorescence. Abbreviations: brain, b; heart, h; jaw, j; skeletal muscle, m; and notochord, n. The + indicates consistently observed; the - indicates consistently not observed; +/- indicates sporadic or weak occurrence in offspring.

Figure 5. Expression of interspecies conserved non-coding elements. (A) Minimal *sox9b* promoter construct (*sox9b* CPE) driving EGFP expression. **(B)** *sox9b* core promoter with the additional fish CNEs (*sox9b* CPE + fish CNE's) driving EGFP expression. **(C)** Lateral views in the green channel of 96 hpf zebrafish transiently expressing EGFP under control of the various *sox9b:EGFP* constructs. Abbreviations are: brain, b; heart, h; jaw, j; skeletal muscle, m; and notochord, n.

Figure 6. Schematic of transient and stable transgenic results. Graphic representations of *sox9b:*EGFP expression are shown for each heart, skeletal muscle, notochord, brain, and jaw as indicated by the initials. The height of each letter represents the incidence of expression of the data collected. **(A)** Results for all lengths tested are shown. For reference, the locations of the CNEs are shown in each segment, using the color-coding from Figure 2. **(B)** Results for constructs carrying individual CNEs.

Figure(s)



C.

>-2421/+29*sox9b 5'*

```
CGTTTGCTACGTCCAGCTAATGTGCTTGTAAATAGAGAGTGGGCAGCTACAGGTGCGAGTGTCACATTTAGAGCATGGGCAATATGCATTTGTGTTTGATTCTAA 105
ATGGTGTTTGGGTCATTTTTGACCCGTCCGCCAGCAGAAACGTTGCCGGCAGCGCTTCAACTGATCACTGCGTGTGCTGGAGTGTGCTGTACACGTCTCTGTGA
TGCCCGTCCTCGAAAGCCGAGCAGGCCGGGAGTCTCCAGCAGAGTTTATGACTGATCAGAGTGTGACTTCAGCAGATCAGGTGTGGCCGTCAGAGACGTGTG
AGGGATGATTCCAGCTGAAGCTGTTGGGTCAATGATGAGTCCTGTTCAAACACTCGGGCCGTGACGCTGCGCTCTGATCCGACGCTCGTACAGCCGTTCCTGAC
TCTGTGCTGTGAGAGAGATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 520
GTGTGTGTGTGTGTGTGTGTGTGTGTGTTCCGCTGGCAGGCCTGTGGAAGCTGGCACATCTGAGGGGAAAATCACAGAGCGGTGCAAACTCATTCCTCACAGG
TCCAGCTGCAGATCTGTCCGCTCCGCCGTCATCTGTGTGTTCTGGTTCTGCTGGATCTCCACCCGCTGGAGTTCCTCACATTCACCACAGAGACTCTACTAGACCA
CAGCAAACAACTCTGCTGACGGGGACTGCCATCAGCTCAGCCGCTCGGAATATTGAGCAGGGGGCGGGGCTTTCTTTAGAAAGTAAGAGGGGTGTGGTCAAG
AATATAGATGTGCAGTTGCTATGGAATCCCTCAGATTGACATCAACAGAAGGAGCCGCTCCAAACACTGAAGCACACGCACTGACTTTGACTGAAGAGTGCCAG
AACACACTGATTACATTAATTATTGCTGAAAAAAACATCCCTGCTGAATATATACATATATATATATACTGTATATATATGTGTGTGTGTGTTGTGGTTCAATGCTGGTC 1045
CCAATGCTGGTCAGGGACAGTAAAACCAGCATTGACACACACACACACACACCCTCTCTCTCGGCTGTCCGCTGCAGTGTCCGCGAGCGTGAACAGGACATGA
TTTAGCAGTTGACTTTCACTGTTTGCCTCTCAGTTCTTAACACTGGAGCTCGAAGTGAAGCTGAGAAGTTTCCCAGCGCTGCAGTGTGTGTGACTGAGGAGAG
AGCTGCGGTATCACTGATCACACACACACACACACACACACACACACACGCTGCTCCACCTGTGAACACATGAGACTCACGCTGGCCTTTAATGCTCTCACAGATGGA
GATGGAGAGGAGCGTCAGAAAGTAAAACCTTCTCCTGCTGAACACTAGAGAAGATATTTTGAAGAATGCTGGACGCCTGTTGTGACCACTCACTTCCATTTGTT 1459
ATTCCTACTATGTGAGCCAATGGTGACCGGTTTCCAACATTCCTCAAAATATCTTCTTTTGTCTTCAACAGACTGAAGAAACTCATAAAGGTGTAAATGCACATGA
GGGGGGTAATGAGTGAGTGAATCTGTGTTTTGGGCTGAGCGGACCCTTTAAACACTGAGGCTGGCCCTTTCTGCTGGCTCCAACACAGTTTATACAGGATTATT
CAGATCAGAGATGAGCTTATCCTGCTGCGCTGTGTGCAGAACAGCATTTCAGATTGTGTGGCCATCTCTGCGCTGCCATTGAGCAGTGTTTTAGTGGGATCATCT
AAAATGCAGACGCAATAGATGAGTGTGGAAGCCAGCAGAGGACTGCAGTAACCATGCTGGAATCTCTCAGCATTAGAAGACTCTCCACTACAACACTGACTCTA
CCCTCAACACAATTTAGAGAAGCTGATAAAGCGGTCCCGGAGCCAACATGAGCTGCTCACGGGTGTGTTCTGACGCTCAGAGACGCTGCTCAGACACACATCT 1981
GTCATCTATTAAACATGACTCCTCCGGCAGCTCATGAGGAACACACGCCCGACCCACGGCCACTGAAAATGACATGTCCACAGCATCACCAGCACAAATAAACA
CAGAGAAACTCAGAGCTGCACTGTAGATGTGTGGATACCGCAGTACTCACAGCACAGGACGGCCATTTCCAGACACTCAACACACGAGATGTTACACAGTTCCC
CACTACACCACACATTCACTAGAGCCGGCCACAGATCCATCATGCATAATCCCGTTCAGCAAATCACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTGCTCCTCCCCTCTGCACTATATAAGTG
TCGTGCTCTGGAGAGGGAGCAGAATCTCCTCCTGGATTTCTGGACTCTC-3'                                                 2500
```

Transcription Start

Fig. 1, Burns et al.

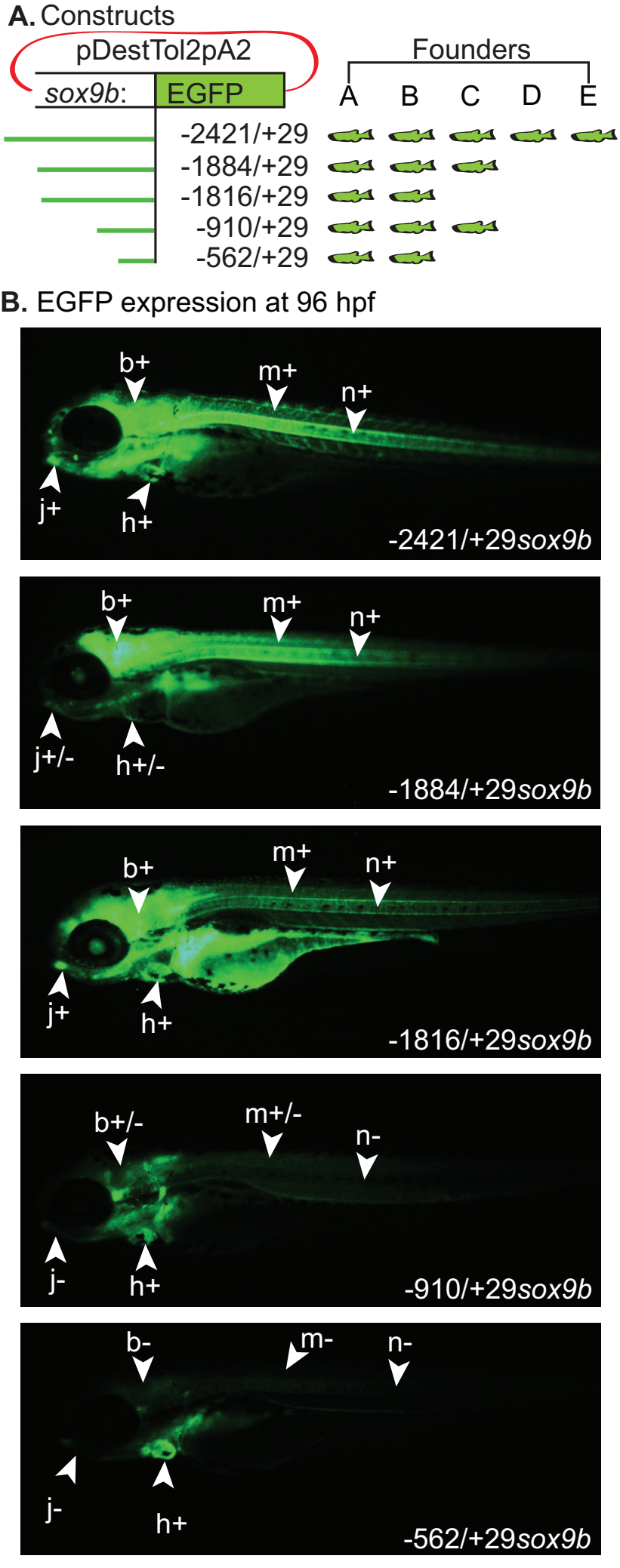


Fig. 2, Burns et al.

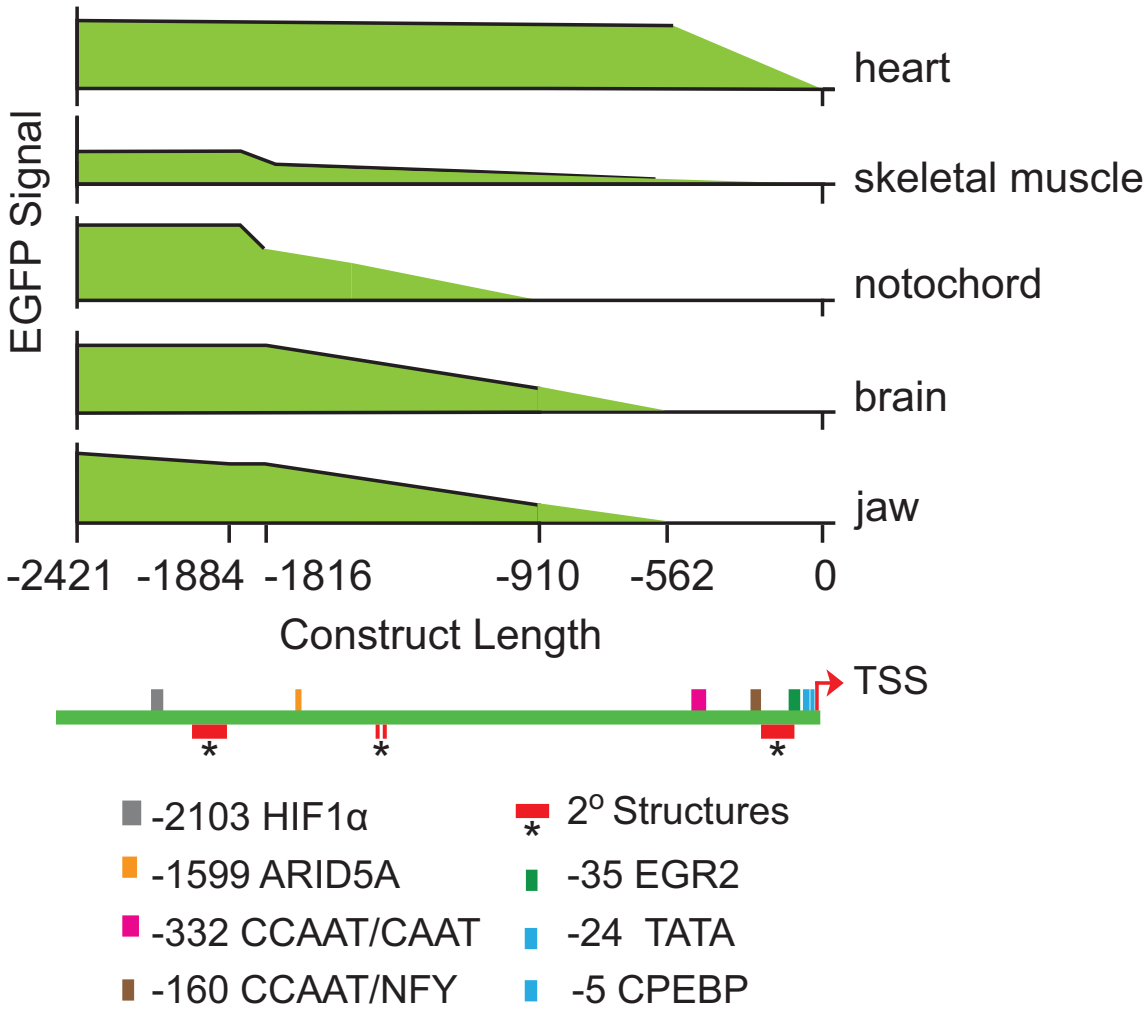
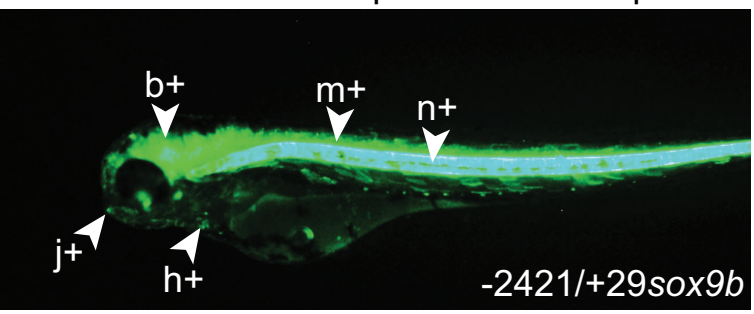
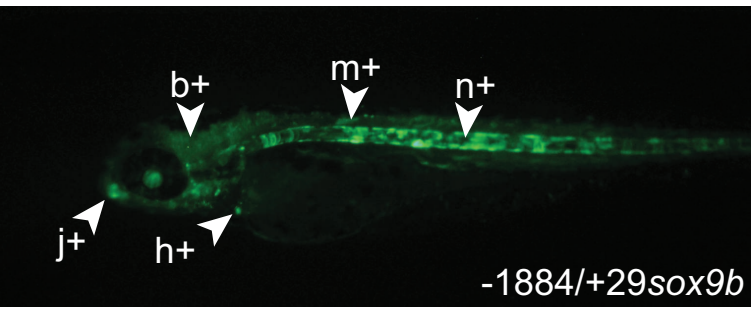


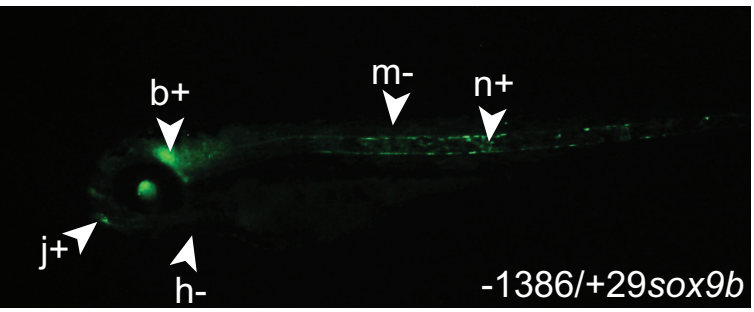
Fig. 3, Burns et al.

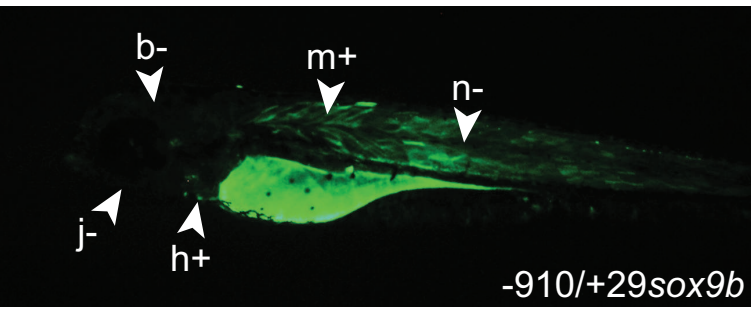
A. Constructs

B. Transient EGFP expression at 96 hpf









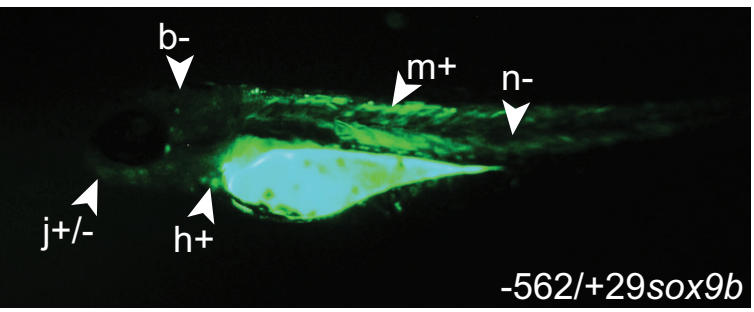


Fig. 4, Burns et al.

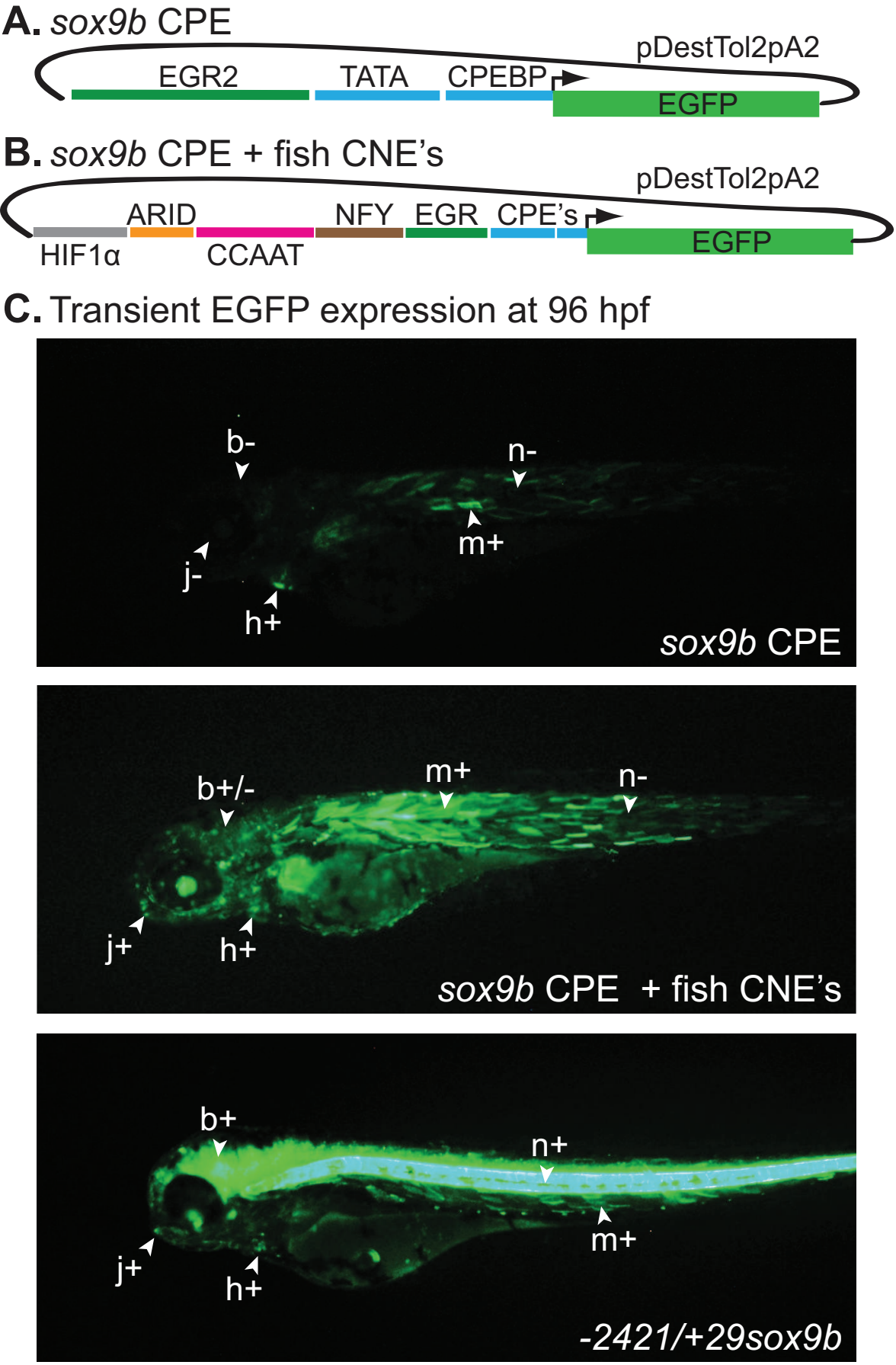


Fig. 5, Burns et al.

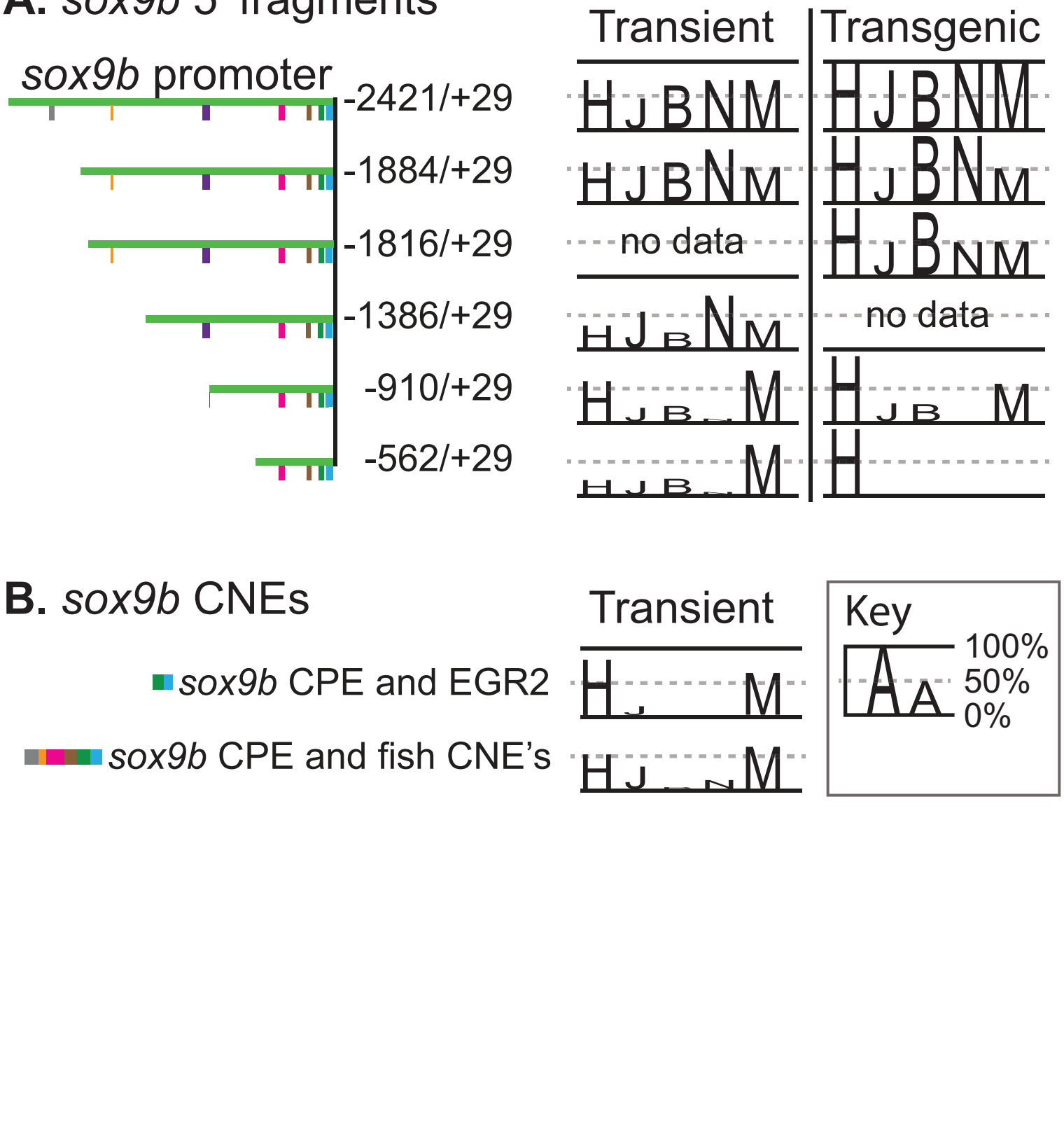


Fig. 6, Burns et al.