Identification of *in vivo* Hox13 binding sites reveals an essential locus controlling zebrafish *brachyury* expression

Zhi Ye, Christopher R Braden, Andrea Wills and David Kimelman*
Department of Biochemistry
University of Washington
Seattle, WA 98195-7350

*Author for correspondence (kimelman@uw.edu)

Keywords: Hox genes, *brachyury*, Wnt signaling, Neuromesodermal progenitors

Abstract

During early embryogenesis the vertebrate embryo extends from anterior to posterior due to the progressive addition of cells from a posteriorly localized neuromesodermal progenitor (NMp) population. An autoregulatory loop between Wnt and Brachyury/Tbxt is required for the NMps to retain mesodermal potential, and hence normal axis development. We recently showed that the Hox13 genes help to support body axis formation and to maintain the autoregulatory loop, although the direct Hox13 target genes were unknown. Here, using a new method for identifying *in vivo* transcription factor binding sites, we identified over 500 potential Hox13 targets. Importantly, we found two highly conserved Hox13 binding elements far from the *tbxta* transcription start site, which also contain a conserved Tcf7/Lef1 (Wnt response) site. We show that the proximal of the two elements is sufficient to confer somitogenesis stage expression to a *tbxta* promoter that alone only drives NMp expression during gastrulation. Importantly, elimination of this proximal element produces shortened embryos due to aberrant formation of the most posterior somites. Our study provides a potential direct connection between Hox13 and regulation of the Wnt/Brachyury loop.

Summary statement

A novel screen for *in vivo* Hox13 binding sites in the zebrafish tailbud reveals a highly conserved enhancer element necessary for NMp expression of *brachyury* and body axis elongation.

INTRODUCTION

All vertebrate embryos form their anterior-posterior axis by progressively adding cells to the posterior end of the presomitic mesoderm and spinal cord from a neuromesodermal progenitor population (NMp) located in the tailbud at the most posterior end of the body (Henrique et al., 2015; Kimelman, 2016b; Mallo, 2020; Steventon and Martinez Arias, 2017; Wymeersch et al., 2021). A positive autoregulatory loop consisting of Wnt signals and the T-box transcription factor Brachyury/Tbxt, plays an essential role in the NMp population by sustaining the mesodermal potential of the NMps, such that loss of Brachyury causes a severe loss of somitic mesoderm and an increase in neural tissue (Gouti et al., 2014; Martin and Kimelman, 2012; Yamaguchi et al., 1999). In contrast, too much Wnt and Brachyury is also detrimental because it overpopulates the mesoderm at the expense of neural tissue (Bouldin et al., 2015; Garriock et al., 2015; Gouti et al., 2014; Jurberg et al., 2014; Martin and Kimelman, 2012; Tsakiridis et al., 2014; Turner et al., 2014). Hence the Wnt-Brachyury autoregulatory loop needs to be carefully controlled in order for the embryo to produce the correct amount of mesoderm and neural tissue as the body extends, but little is known about the precise mechanisms controlling the levels of these factors.

We recently showed that the Hox13 genes, specifically *hoxa13b* and *hoxd13a*, enhance the ability of the zebrafish *brachyury* ortholog, *tbxta/no tail*, to promote formation of the mesoderm during the somitogenesis stages through helping to maintain expression of *tbxta*, *wnt3a* and *wnt8* (Ye and Kimelman, 2020). However, the direct targets of these Hox13 factors were unknown. While the importance of the Hox genes in embryonic development has been known for decades, identifying *in vivo* targets has been a major challenge particularly in vertebrates, and only a limited number of verified vertebrate Hox targets are known (reviewed in Mann et al., 2009). Large scale *in vitro* screens have been very useful for identifying DNA binding motifs for Hox proteins, both in *Drosophila* and vertebrates (Berger et al., 2008; Jolma et al., 2013; Noyes et al., 2008), but the sequences are relatively non-specific making *in silico* predictions of Hox binding sites very challenging, and essentially impossible in cases where transcription is regulated by distal enhancer elements. However, an important finding from one of these studies that used protein produced in cell culture is that the posterior Hox factors,

Hox9-13, not only bind the canonical Hox motif shared by all Hox factors, but also bind a slightly different alternative binding motif (Jolma et al., 2013).

One major problem for identifying *in vivo* Hox binding sites in vertebrate embryos is the severe paucity of high quality and highly specific antibodies to the Hox proteins, which is an even bigger problem for zebrafish than for mammals due to less commercial interest. In addition, approaches such as ChIP-seq are not feasible for tissues such as the tailbud, which contains only a few hundred cells. We have solved the first of these problems by creating two transgenic zebrafish lines that allow us to produce controlled levels of Hoxa13b with a small epitope tag, circumventing the need for a Hox13 specific antibody. Secondly, we have adapted a relatively new technology called CUT&RUN for use on zebrafish embryos, which permits quantitative identification of *in vivo* binding sites for transcription factors using very limited numbers of cells (Meers et al., 2019; Skene and Henikoff, 2017).

Using these approaches we have identified at least 576 candidate Hoxa13b target genes in the zebrafish tailbud during the middle of the somite forming stages. Utilizing epigenetic markers in parallel, we have further characterized the targets associated with transcriptional activation to identify genes that are potentially activated by Hoxa13b. Importantly, we found two transcriptionally active Hoxa13b binding sites far upstream of the *tbxta* transcription start site, which have been conserved over hundreds of millions of years of evolution. These enhancer elements also each contain a conserved perfect consensus binding site for Tcf7/Lef1, which are the Wnt regulated transcription factors in cells, revealing how Hox factors and Wnt signaling could intersect to promote *tbxta* expression. We show that the proximal of the two elements, which in zebrafish embryos shows higher levels of Hoxa13b binding in the CUT&RUN analysis, is sufficient to drive posterior expression of *tbxta* in the tailbud when joined to a proximal *tbxta* promoter, resolving the previously puzzling inability in zebrafish to find a *tbxta* promoter that sustains expression in the NMps during axis extension. Importantly, we show that deletion of the proximal enhancer element using CRISPR produces shorter embryos due to defects in the most posterior somites, as well as reduced expression of *tbxta* in the NMps. Finally, we reveal an unexpected role for the transcription factor Rbpj, which binds near the Hox13 binding sites in the *tbxta* gene, in regulating *tbxta* expression and posterior body

formation. In addition to contributing a rich source of data for understanding gene expression within the tailbud, our results provide a potential molecular basis for understanding how Hox13 proteins act to promote axis formation, and reveal a critical Hox-binding enhancer element regulating the Brachyury/Wnt autoregulatory loop.

RESULTS

Adaptation of CUT&RUN for use in zebrafish explants

In order to solve the lack of quality antibodies to Hoxa13b, we produced two different transgenic lines that each add a FLAG epitope tag to Hoxa13b. One of the lines produced a heat-shock inducible Hoxa13b-FLAG, whereas the other placed the Hoxa13b-FLAG tag at the endogenous *hoxa13b* locus. The reason that we used both approaches is that expression level of endogenous *hoxa13b* in the tailbud is low, as is true for all the Hox13 genes (Ye and Kimelman, 2020), and we were concerned that the low expression from the endogenous *hoxa13b* locus might cause us to miss some of the authentic Hoxa13b target genes because of a weak signal. By using both approaches we hoped to maximize our chances of identifying the complete set of Hoxa13b binding sites. We also note that inducible epitope-tagged Hox factors have been very recently used in differentiated mouse embryonic stem cells to identify Hox targets in that system, confirming the utility of this approach (Bulajic et al., 2020).

To produce the inducible transgenic line, 27 amino acids containing three FLAG tags was added to the C-terminus of Hoxa13b (Figure 1A). In order to monitor expression, GFP was placed after the 2A peptide, which allows GFP to be co-expressed with Hoxa13b as a separate protein. This coding region was placed under control of a zebrafish heat shock promoter to make the line *HS:hoxa13b-FLAG-GFP*. We previously showed that a very similar transgenic line that produces a non-epitope tagged version of Hoxa13 produces a severe axis formation defect when heat shocked at 40°C (Ye and Kimelman, 2020). This new line produces the identical phenotype when heat shocked under the same conditions, demonstrating that the epitope tag does not impair Hoxa13b function (Figure S1).

The second transgenic line was produced by taking the genomic *hoxa13b* sequence, which includes one intron, and placing the FLAG-2A-GFP sequences at the C-terminus of the coding region. This was flanked by a unique DNA sequence called *Mbait* that is not present in

the zebrafish genome, allowing the plasmid to be cut *in vivo* by a specific *Mbait* gRNA without creating off-target effects (Kimura et al., 2014). This construct was injected into zebrafish embryos along with Cas9, the *Mbait* gRNA, and gRNAs targeting the 5' and 3' untranslated regions of the endogenous *hoxa13b* locus in order to insert the modified *hoxa13b* at the endogenous *hoxa13b* locus. We identified one line that showed the expected expression of the FLAG-tagged protein at the C-terminus of the embryo in the NMps and mesoderm (Figure 1C-G and Video S1). This knockin (KI) transgenic line will be referred to as *KI:Hoxa13b-FLAG-GFP*.

To identify *in vivo* Hoxa13b binding sites we adapted the CUT&RUN procedure (Meers et al., 2019; Skene and Henikoff, 2017) for use in zebrafish. In order to produce Hoxa13b-FLAG in embryos without perturbing development from *HS:hoxa13b-FLAG-GFP*, we first established that heat shocking at 36°C produced just enough GFP to detect, but did not cause any phenotypic alterations to embryonic development. We then manually isolated tailbuds from embryos of both the inducible and knockin transgenic lines at the 15-18 somite stage, which is in the middle of the somite-forming stages, and then dissociated and permeabilized the cells. Based on the expression pattern in the knockin line (Figure 1F and Video S1), we estimate that approximately 50% of the cells in our dissected tailbuds produce Hoxa13b-FLAG protein. Anti-FLAG antibody was added, followed by Protein AG-Micrococcal nuclease (pAG/MNase), which cuts just adjacent to the FLAG-tagged transcription factor (Figure 1B). Once the genomic DNA is cut, the complex is released into solution and the associated DNA is subjected to DNA sequence analysis. In parallel, we isolated tailbuds from wild-type (WT) embryos and performed CUT&RUN using antibodies to H3K4me1 and H3K27ac in order to identify the regions of the genome in the tailbud associated with transcriptional activation (Calo and Wysocka, 2013). Finally, as a control, we performed CUT&RUN on WT embryos with the anti-FLAG antibody, which is not expected to bind any endogenous zebrafish transcription factors.

As CUT&RUN has not been previously performed in zebrafish, we first wanted to be sure that it could be done effectively. We evaluated the success of our CUT&RUN experiment by checking the size distribution of all aligned fragments for the histone modification samples. We observed a strong enrichment of mononucleosome-sized fragments, with a small amount of dinucleosome-sized fragments, which is a typical pattern for histone modification CUT&RUN

data from a successful experiment (Figure 2A,B, Meers et al., 2019; Skene and Henikoff, 2017). In contrast, the control antibody samples showed a random distribution of sizes. Correlation analysis of the samples from independent experiments showed a good reproducibility among biological replicates, and showed that the H3K27ac and H3K4me1 modifications were often found together, as expected (Figure 2C). Finally, we used heat maps to examine signal enrichment at specific locations, as used previously (Henikoff et al., 2020; Skene and Henikoff, 2017). Whereas DNA fragments produced from the control DNA showed no enrichment at the called-peak sites, DNA fragments produced using the histone modification antibodies showed central enrichment at these sites (Figure 2D,E).

Because the most comparable ChIP-seq data using the histone modifications was produced with 24 hour post-fertilization (hpf) whole embryos (Bogdanovic et al., 2012; de la Calle Mustienes et al., 2015) whereas we used posterior explants at 16-17 hpf, we were not able to directly compare through bulk analysis our CUT&RUN results to stage-matched ChIP-seq data. However, comparing a variety of posteriorly expressed genes that remain on at both time points such as *sox2* and *cdx4* showed similar results for the histone modifications in both methods (Figure 2I). The results of the histone modification analysis are provided in Supplemental Tables 1 and 2, and will be useful for identifying transcriptional regulatory regions for genes involved in the NMps, presomitic mesoderm formation and spinal cord neurogenesis. In summary, our results show that CUT&RUN can be effectively used in zebrafish. And whereas the previous ChIP-seq data required 1000 whole embryos per biological replicate (Bogdanovic et al., 2012; de la Calle Mustienes et al., 2015), our results were obtained with 30 tailbuds per replicate, opening up this type of genomic analysis to a variety of studies in zebrafish that were not previously possible such as analyzing chromatin signatures in small regions of the embryo or tissue specific fragments.

Identification of *in vivo* Hoxa13b binding sites in the tailbud

We next proceeded to use CUT&RUN on our transgenic lines expressing Hoxa13b-FLAG. Results from both lines showed clear central enrichment of signal at the called-peak regions (Figure 2F,G). Comparing the results using *HS:hoxa13b-FLAG-GFP* and *KI: hoxa13b-FLAG-GFP* we

found that the heat shock line gave better signal to noise and so we used that for an initial identification of Hoxa13b binding sites, and then used the data from the knockin line as a second level of analysis for peaks called from the inducible line to find sites identified in both transgenic lines. The reduced signal from the knockin line is not only due to the lower level of expression of Hoxa13b (based on relative GFP expression), but also because at lower levels of expression the Hoxa13b-FLAG is competing for binding against other Hox13 factors, and potentially all Hoxa9-13 proteins (see Discussion), making it more difficult to detect binding. Because the knockin data is likely to underrepresent endogenous Hoxa13b binding sites, as shown below for one of the enhancer elements in the *tbxta* gene, we have included peaks that were found with both the inducible and knockin lines in Supplementary Table 3 as well as peaks found only with the inducible line. This table also includes the associated histone marks discussed above for each of the peaks.

In total we identified 1871 Hoxa13b occupied peaks using *HS:hoxa13b-FLAG-GFP* embryos, with 689 of them overlapping peaks identified using the *KI: hoxa13b-FLAG-GFP* line. Only 12.6% of the 1871 binding sites are located within a 3 kb distance from the transcription start site of the nearest gene whereas more than 60% of the sites are in the distal intergenic regions (Figure 3A, Figure S2A). The same result was observed in the recent ChIP-seq analysis of the binding sites of multiple Hox factors in differentiated mouse embryonic stem cells (Bulajic et al., 2020). Based on the TxDb.Drerio.UCSC.danRer11.refGene annotation database, we mapped the 1871 binding sites identified from the *HS:hoxa13b-FLAG-GFP* analysis to the nearest neighboring gene, which identified 1380 genes. The 689 sites overlapping in the knockin data mapped to 576 neighboring genes (Supplementary table 3).

We analyzed the DNA peaks for common motifs using the programs MEME and DREME, and found that a Hox motif was the most common motif discovered (Figure 3B and S2B). The motif is centrally located in the peak region (Figure 3C), with footprinting analysis showing that it forms a 19-bp footprint (Figure 3D) due to the protection of the DNA from MNase cutting by the bound transcription factor (Liu et al., 2018; Skene and Henikoff, 2017). Intriguingly, at a very specific position in the motif (Figure 3B) approximately half the fragments have a T, which is the base bound by all Hox proteins (Hox1-13), whereas in the other half a C is detected, which is

bound specifically by the posterior Hox proteins (Hox9-13, Jolma et al., 2013). Thus, Hoxa13b binds both *in vitro* predicted motifs (TTTAT and TTTAC) in our system. We did not note a clear preferential association of T versus C with other enhancer marks. In addition, apart from the Hox motif, we did not consistently identify any obvious secondary motifs, suggesting that the posterior Hox proteins, unlike the more anterior Hox proteins that bind TALE co-factors (Mann et al., 2009), might instead rely on a diverse group of co-factors in order to achieve specific binding.

Identification of Hoxa13b bound enhancer elements in *tbxta*

We next examined the list of genes bound by Hoxa13b for targets that could explain how Hox13 proteins might support the Brachyury/Wnt loop. We were particularly intrigued to find a Hoxa13b binding site neighboring the *tbxta* gene, located 24.6 kb upstream of the start of transcription (Figure 4A), which we call Hox element 1. This site is associated with the active enhancer marks H3K4me1 and H3K27ac (Figure 4A), suggesting that this region is an active transcriptional enhancer in the tailbud. This was of particular interest since a major puzzle with *tbxta* in zebrafish is that a 2.1 kb fragment upstream of the transcription start site drives strong expression throughout the NMps during the gastrula stages, and later in the notochord, but does not produce expression in the NMps during the somitogenesis stages when the axis extends (Harvey et al., 2010), suggesting that a hitherto unidentified distal enhancer is required to promote the second phase of neuromesodermal *tbxta* expression. Indeed, in our previous work showing the importance of the Hox13 factors on the Brachyury/Wnt loop we speculated that the Hox proteins might work through a yet unidentified distal enhancer for *tbxta* (Ye and Kimelman, 2020).

When we used BLAST to compare this *tbxta* distal enhancer region including the Hoxa13 binding site to other fish, we found that an approximately 200 bp region was conserved in all ray-finned fish we could analyze, including reedfish from the genus *Polypterus*, which diverged from all other actinopterygians (ray finned fish) including zebrafish about 400 million years ago (Takeuchi et al., 2009), indicating a strong evolutionary pressure to retain this element (Figure

4B). The Hox binding motif is completely conserved among all these fish, and intriguingly all of them contain the C that is specific for binding by the posterior Hox proteins (Figure 4B, arrowhead).

We were also very interested to observe that Hox element 1 contains a perfect match to the consensus binding site for Tcf7 and Lef1 transcription factors, which are the transcription factors activated by the Wnt pathway (Cadigan and Waterman, 2012), since Wnt expression is required for *brachyury/tbxta* expression throughout early development (Arnold et al., 2000; Galceran et al., 1999; Martin and Kimelman, 2008; Vonica and Gumbiner, 2002; Yamaguchi et al., 1999). This Tcf7/Lef1 binding sequence is exactly the same one used in the Super TOPflash Wnt reporter (Veeman et al., 2003). The combination of a Hox13 binding site and a Tcf7/Lef1 binding sequence therefore suggests that Hox element 1 is likely to integrate Hox13 binding and activation by the Wnt pathway.

In examining the CUT&RUN data from the *HS:hoxa13b-FLAG-GFP* fish we saw a much weaker peak of Hoxa13b binding 28.9 kb upstream of the start of transcription start site in the *tbxta* gene that was not called as a peak by our algorithm, nor was a peak present in data from the *KI: hoxa13b-FLAG-GFP* fish (Figure 4A). When we examined this region using a BLAST search we found that this region is also highly conserved, with conservation as far as the lobe-finned fish coelacanth (Figure S3). Both the posterior-specific Hox motif and a consensus Tcf7/Lef1 site were found in this element, suggesting it might serve as a second enhancer element for *tbxta*. This site is also associated with the activating marks H3K4me1 and H3K27ac (Figure 4A), demonstrating that it is also an active enhancer in the tailbud. We therefore call the region at 28.9 kb Hox element 2.

Hox element 1 drives tailbud expression of *tbxta*

To test a role for the Hox elements we inserted them along with non-conserved flanking sequences in order to allow for DNA looping into a Tol2 vector that contained GFP and the 2.1 kb proximal promoter (Figure 5A). We noticed a site that had strong H3K4me1 and H3K27ac occupancy at -5.4 kb, and this also was conserved among different fish species, although no recognizable transcription factor motifs were found in this element (Figure S4). Therefore, we

included it in our initial analysis in case it was important for enhancing *tbxta* transcription. As shown previously (Harvey et al., 2010), transgenic zebrafish lines containing the 2.1 kb proximal promoter produced strong GFP expression during the gastrula stages and at later stages in the notochord, but very little GFP expression in the tailbud and posterior somites during somitogenesis stages (Figures 5B and S5). In contrast, a transgenic line that included the proximal promoter, the two Hox elements and the -5.4 kb element produced robust expression in the somitogenesis stage tailbud and posterior somites (Figures 5C and S5). We then made a new transgenic line that included just Hox element 1 and its flanking sequences, and found that this also produced strong tailbud expression (Figure 5D). These results demonstrate that the Hox element 1 is sufficient to confer somitogenesis stage tailbud expression upon the proximal *tbxta* promoter.

While in most fish these Hox elements are located far from the start of *tbxta* transcription, in channel catfish we noticed that they are located much closer (Table 1). We therefore isolated 5.7 kb of the catfish *tbxta* upstream region and derived transgenic zebrafish lines containing this construct. As shown in Figure 5E, the catfish promoter was able to drive strong tailbud expression in zebrafish embryos, demonstrating functional conservation of the *tbxta* upstream region despite it being much smaller in catfish. Interestingly, while the three elements shown in Figure 5 are conserved between zebrafish and catfish, we could not detect any conservation in the 2.1 kb proximal promoter region between these species, as was also seen previously in a comparison between zebrafish and medaka for the same promoter region (Harvey et al., 2010).

Hox element 1 is necessary for normal body axis formation

While our data showed that Hox element 1 is sufficient to regulate *tbxta* expression in the posterior, we wanted to know if it is necessary. To examine this we used CRISPR with gRNAs that flanked Hox element 1 (Figure 6A), and isolated two transgenic lines that had deleted this element and bred these through two generations. Fish *Ptbxta*$^{\Delta 312}$ has a 312 bp deletion and *Ptbxta*$^{\Delta 314}$ has a 323 bp deletion and a 9 bp insertion, and both produced the same phenotype. In crosses of fish heterozygous for the deletion, one quarter of the fish had a shorter body

(Figure 6B,G; 25%, n=92). When a subset of the embryos were genotyped, all of the short bodies were homozygous for the deletion (100%, n=6) whereas the normal length bodies were heterozygous or WT (100%, n=6). Examining these shorter body fish revealed no change in the number of somites and instead we observed aberrant formation of the most posterior somites compared to WT (Figure 6H, Figure S6A). We also examined *tbxta* expression as we expected this to be changed. At 24 hpf, we identified a portion of embryos with strongly reduced *tbxta* (Figure 6C,D). Genotyping of these embryos showed that 100% were homozygous for the deletion (n=13). Since reduced *tbxta* expression would be expected to lead to reduced mesoderm formation, we examined the early mesodermal marker *tbx16* four hours later (28 hpf) and observed embryos with strongly decreased *tbx16* (Figure 6E,F). Genotyping of these embryos showed that 100% were homozygous for the deletion (n=14). Thus, deletion of Hox element 1 causes strongly reduced *tbxta* and *tbx16* expression and a shorter body.

Because the NMps are bipotential, when the amount of mesoderm decreases there is a corresponding increase in neural tissue (Garriock et al., 2015; Gouti et al., 2014; Martin and Kimelman, 2012; Tsakiridis et al., 2014; Turner et al., 2014). To see if this occurs in fish lacking Hox element 1, we examined the volume of the neural tube using the neural marker *sox2* with Fluorescent in situ hybridization (FISH). Whereas WT and heterozygous embryos had the same neural tube volume, embryos homozygous for the Hox element 1 had a markedly increased volume (Figure 6I, Figure S6B). Thus, when Hox element 1 is deleted from the *tbxta* gene the NMPs switch from producing mesoderm to neural tissue in the most posterior end of the embryo.

Rbpj contributes to *tbxta* expression

In addition to Hox and Tcf7/Lef1 binding sites in Hox element 1 we also found a perfect match to the consensus sequence for Rbpj (Figure 4, Johnson and Macdonald, 2011). Rbpj, which in *Drosophila* is known as Suppressor of Hairless, is best known for its transcription repressing roles in the Notch signaling pathway, but it also has important Notch-independent roles that include both activation and repression (reviewed in Johnson and Macdonald, 2011). Zebrafish have two duplicated *rbpj* ohnologs, *rbpja* and *rbpjb*, both of which have maternal and

ubiquitous expression through the first day of development (Echeverri and Oates, 2007; Sieger et al., 2003). Consistent with their role in the Notch pathway, knockdown of both ohnologs using morpholino oligonucleotides (MO) causes defects in cyclic gene expression in the presomitic mesoderm and somite boundary defects (Echeverri and Oates, 2007; Sieger et al., 2003). Interestingly, MO knockdown embryos have posterior defects, particularly in tail formation, although this was not examined in detail. In mouse, null mutants of *Rbpj* die by 10.5 days post fertilization and form less than six somites (Oka et al., 1995), which is reminiscent of the somite defects in *T/brachyury* null mutants (Beddington et al., 1992). Unfortunately, the expression of *T/brachyury* was not examined in these embryos, but there was a very strong reduction in the levels of the somite marker *Meox1*. These results show that Rbpj has roles that extend beyond just regulation of the Notch pathway in vertebrate embryos, and involve both formation of the mesoderm as well as cell growth and survival.

In order to determine if Rbpj binds Hox element 1, we made a new transgenic line *HS:rbpja-FLAG-GFP*, which allows conditional expression of a FLAG tagged version of Rbpja with a co-expressed GFP. Embryos expressing Rbpja-FLAG after heat shock did not show any defects using the standard temperature of 40°C. Embryos co-expressing GFP were subjected to CUT&RUN as described above, and central enrichment of signal at the called-peak regions was observed (Figure 2H). In agreement with the proposed role of Rbpj on Notch signaling, we observed peaks near the presomitic cycling genes *her1, her4, her7, hes6, hey1* and *hey2*, and also at the Notch targets *dlc* and *dld* (Supplementary table 4). In addition, we observed peaks near *cyp26a1*, which was shown to be strongly reduced in the tailbud when *rbpj* MOs were used (Echeverri and Oates, 2007). Importantly, we identified binding at the consensus Rbpj site in Hox element 1 (Figure S7). We also observed Rbpja binding to Hox element 2 (Figure S7), and while we could detect an Rbpj site close to Hox element 2 in all fish we examined, the position of the site was not conserved between species. In zebrafish, for example, the Rbpj site is 40 bp beyond the Tcf7/Lef1 site shown in Figure 4B.

Because of the well known issues with the use of morpholinos, we chose to take an alternate dominant-negative approach to examine the role of Rbpj in zebrafish by creating a transgenic line in which the Engrailed repressor domain is expressed at the N-terminus of Rbpja

(*HS:EnR-rbpja-FLAG-GFP*). This construct was very effective as EnR-Rbpja expressing embryos heat shocked at 38°C showed a strong posterior defect whereas *HS:rbpja-FLAG-GFP* embryos were unaffected (Figure 7A). Expression of *tbxta* was examined 5 hours after heat shock and observed to be markedly reduced in *HS:EnR-rbpja-FLAG-GFP* embryos in the tailbud, but not in WT or *HS:rbpja-FLAG-GFP* embryos (Figure 7B-D), whereas *cdx1a* and *snai1*, which are genes that do not have Rbpja binding sites (Supplementary table 4), were unaffected (Figure 7E-J). In contrast, notochord expression of *tbxta* was not inhibited by EnR-Rbpja (Figure 7B-D). We also examined apoptosis and saw no increase in apoptosis in *HS:EnR-rbpja-FLAG-GFP* embryos 5 hours post heat shock, demonstrating that the reduction of *tbxta* was due to a change in gene expression and not a loss of cells. These results show that Rbpja also contributes to the activation of *tbxta* expression.

DISCUSSION

Identification of novel Hox-binding *tbxta* enhancer elements

While our previous work provided genetic evidence for a role of the Hox13 proteins in sustaining the NMps, the direct targets of the Hox13 proteins were unknown. Using two new transgenic zebrafish lines and the CUT&RUN method for identifying *in vivo* binding targets, we have identified at least 576 candidate Hox13 direct target genes in the tailbud and potentially almost 1400 if all of the candidate Hoxa13b binding sites from the inducible transgenic line are included. We note that this is the first use of CUT&RUN in zebrafish, and our results demonstrate that this method will be very valuable since a high quality antibody for a particular transcription factor is not required as long as a small epitope tag can be added, and small amounts of tissue can be used, eliminating background signals from other regions of the embryo. While direct targets of specific transcription factors have been identified in zebrafish previously using ChIP-seq (Jahangiri et al., 2012; Morley et al., 2009), this used large numbers of whole embryos (5000 embryos per biological replicate). Even analysis of chromatin signatures required 1000 whole embryos per replicate (Bogdanovic et al., 2012; de la Calle Mustienes et al., 2015). In contrast, we used 30 tailbud explants per biological replicate, demonstrating that transcription factor binding data can be ascertained on very small embryonic tissue samples,

greatly expanding the ability to do specific genomic analysis on defined regions of the embryo. This same approach should also work on other embryos, particularly where transgenic approaches are feasible.

The data provided here should be of widespread use for researchers studying many aspects of posterior body formation, including transcriptional analysis of somite boundary formation, since we included analysis in WT embryos of two transcriptional activating epigenetic marks. Moreover, of the more than 500 Hoxa13b bound genes shown in Supplementary table 3, we only analyzed one of them, leaving many other interesting targets for future studies. Here we focused on a transcriptional activating enhancer of somitogenesis stage *tbxta* expression that we call Hox element 1, which was of great interest since in our previous work (Ye and Kimelman, 2020) we speculated that a distal *tbxta* enhancer could be a Hox13 target based on a failure of a proximal 2.1 kb promoter fragment to drive expression in the NMps after the end of gastrulation. This enhancer is the missing regulatory element that had been revealed in a previous study of the *tbxta* promoter in zebrafish (Harvey et al., 2010), and is of high practical value as it has previously not been possible to drive expression of constructs in the NMps in zebrafish beyond the gastrula stage, except for the notochord (Row et al., 2016).

The importance of Hox element 1, as well as Hox element 2, is supported by its conservation throughout all of the ray finned fish, spanning 400 million years of evolution. Interestingly, the distance from these elements to the start of transcription is highly variable (Table 1), demonstrating that much of the upstream region is not essential for *tbxta* expression. Indeed, we found that most of the *tbxta* upstream sequences are not conserved between fish species, nor is the proximal *tbxta* promoter, as previously noted (Harvey et al., 2010). While we were not able to find sequences obviously matching these Hox-binding elements in other vertebrates, relatively simple DNA rearrangements or changes in the sequences other than the Hox and Tcf7/Lef1 motifs will preclude finding homology using similarity searches. Especially because the Hox binding motif is so degenerate, our results show the importance of beginning with an *in vivo* identification of Hox binding sites.

The Hox binding elements as collaborative regions

While our previous loss of function studies (Ye and Kimelman, 2020), and overexpression studies in mouse and chicks (Aires et al., 2019; Denans et al., 2015; Young et al., 2009), have focused on the Hox13 factors, our observation that the *tbxta* Hox-binding elements, as well as binding sites in many other genes contain a motif specific to the posterior Hox factors *in vitro* (Jolma et al., 2013) suggests that these factors may act in combination to regulate posterior body formation. A very recent paper examining binding of HOXC9, HOXC10 and HOXC13 in an *ex vivo* system (motor neurons produced from embryonic stem cells) using induced epitope-tagged HOX proteins and chromatin immunoprecipitation (ChIP), revealed that HOXC13 binds many sites not bound by HOXC9 or HOXC10, although approximately 10% of the sites bound all three factors (Bulajic et al., 2020). It will therefore be interesting in the future to use CUT&RUN to determine which of the Hox13 *in vivo* targets in the tailbud are also bound by Hox9-12 to determine which genes are likely to be solely Hox13 regulated genes and which are more generally controlled by all of the posterior Hox genes. The posterior Hox proteins have always been somewhat confusing since unlike the more anterior Hox proteins that bind TALE cofactors such as Pbx and Meis to increase local DNA binding specificity, specific cofactors for the posterior Hox proteins have not been identified (Mann et al., 2009). With our large dataset of Hox13 binding sites we looked for a motif commonly associated with the Hoxa13b binding site and did not find anything specific. Potentially this implies that the posterior Hox factors might be more promiscuous in their interacting partners.

In our previous work we proposed that the Hox13 factors act as “guarantors” of gene expression in the NMPs, following an idea first proposed from studies in *C. elegans* by Chalfie and colleagues (Zheng and Chalfie, 2016; Zheng et al., 2015). In this view, the Hox13 factors may not be absolutely essential but instead provide robustness to the Brachyury/Wnt autoregulatory loop. Our proposal was based on the observation that zebrafish *hoxa13;hoxd13* mutants showed synergistic defects in formation of the body axis and in *tbxta* expression when Tbxta activity was reduced using a temperature sensitive mutation, but did not show an effect when Tbxta was fully active (Ye and Kimelman, 2020). How Hox13 proteins might act as guarantors is unclear. One possibility is that they enhance the binding of other factors such as

Tcf7/Lef1 and Rbpj through collaborative interactions creating what Mann and colleagues have called a “Hoxasome” (Mann et al., 2009). Alternatively, or in addition, the Hox13 proteins, and perhaps all of the posterior Hox proteins, could act as pioneer transcription factors that help to open the chromatin at the Hox elements based on very recent studies demonstrating HOX13 proteins as well as other posterior HOX factors change the chromatin landscape by opening up inaccessible regions (Amandio et al., 2020; Bulajic et al., 2020; Desanlis et al., 2020). While this has been well documented in the limb bud and genital tubercle in mice, the precise role of the Hox13 factors in regulating the *tbxta* enhancer elements awaits further study.

A model for regulation of *tbxta*

During the gastrula stages, the proximal *tbxta* promoter is sufficient to drive expression throughout the NMps through the use of two enhancer elements, one of which (E2) contains a binding site for the Nodal-response factor Foxh1 (Chen et al., 1997; Harvey et al., 2010). This proximal promoter also contains sequences necessary to drive *tbxta* expression in the notochord beyond the gastrula stages, although the enhancer element for this expression is not yet known (Harvey et al., 2010; Row et al., 2016). By the end of gastrulation, we propose that a switch occurs such that the distal Hox-binding elements are required to promote *tbxta* expression in the NMps, primarily through the Wnt signaling stabilizing β-catenin, which binds to Tcf7/Lef1 factors and activates transcription (Figure 8). We propose that Rbpj and the posterior Hox factors aid in this activation, with the posterior Hox factors acting to ensure robust activation of *tbxta* expression.

Elimination of Hox element 1 causes phenotypic defects only in the most posterior somites. This might be due to partial compensation from Hox element 2 since we observed *tbxta* expression in the tailbud of the mutants lacking Hox element 1 until fairly late in somitogenesis. However, previous experiments with a temperature sensitive Tbxta demonstrated that a strong reduction in Tbxta activity from the start of gastrulation onward causes severe phenotypic changes in posterior body formation, whereas the same reduction in Tbxta activity beginning at the end of gastrulation produced only mild phenotypic changes restricted to the most posterior somites (Kimelman, 2016a). This surprising minor requirement

for Tbxta function during the somitogenesis stages in zebrafish is potentially explained by a study that showed that while most of the neural and mesodermal cells in the posterior body are allocated from a NMp pool during the gastrula stages, a second NMp population remains resident in the tailbud and only contributes to the most caudal somites (Attardi et al., 2018). It is this second population that may critically depend on the post-gastrula expressed *tbxta* in zebrafish, and thus the requirement for Hox element 1 to maintain robust *tbxta* transcription until the end of somitogenesis. Interestingly, zebrafish have a relatively small number of somites (~32) whereas other fish, particularly teleosts such as eel, have many more (up to ~200, Ward and Brainerd, 2007), and in these more elongated body plans the role of the NMps and of post-gastrula *tbxta* expression driven by the upstream Hox elements may have a much larger role. While these other fish species are far less studied than zebrafish, it will be very interesting to study the NMps and the role of Tbxta in these species (see also Sambasivan and Steventon, 2020).

Why do the distal *tbxta* enhancer elements become required after the gastrula stages? In fish and frogs, Bmp signaling on the ventral side of the embryo is required for formation of the tailbud and posterior body development (Kimelman and Martin, 2012; Tuazon and Mullins, 2015), and Bmp regulates the gastrula stage ventral expression of *tbxta* in the cells that will become the NMps of the posterior body (Northrop et al., 1995), acting through one of the elements (E1) in the proximal 2.1 kb *tbxta* promoter (Harvey et al., 2010). However, after the gastrula stage Bmp signaling does not have a major role in axis formation (Pyati et al., 2005). Thus, we propose that during the gastrula stages Nodal, Wnt and Bmp act combinatorially to activate *tbxta* throughout all of the NMps of the body, but as gastrulation progresses and Bmp signaling declines, the distal Hox13-binding enhancer elements are required to sustain *tbxta* throughout the somitogenesis stages. While this model in detail is specific to zebrafish and potentially frog embryos, in light of the observation that Rbpj is essential for posterior body formation in mouse (Oka et al., 1995), it will be interesting to examine the role of Rbpj and posterior Hox factors in the regulation of Brachyury in other vertebrates.

METHODS

Zebrafish lines

Wild-type fish were an AB/WIK mixture. Fish were used for crosses between 3 months and 3 years old.

Constructs

Amplification of cDNA or genomic DNA used the primers listed below with using either Platinum Pfx DNA Polymerase (ThermoFisher) or Q5 DNA Polymerase (New England Biolabs) according to the manufacturer's conditions, with an annealing temperature of 60°C.

HS:hoxa13b-FLAG-GFP: the *hoxa13b* coding region was amplified from 15s zebrafish embryo cDNA using the primers *GGATCC*GCTATGACAGCGTCTTTACTCCTC and GTCAACAAGTACAAGGGCATCAGTT*ATCGAT*, with added BamH1 and ClaI sites shown in italics, and inserted into a vector such that the stop codons were removed and a viral 2A peptide (Provost et al., 2007) was placed immediately after the coding region. The GFP sequence was placed immediately after the 2A sequence. This sequence was placed in a Tol2-*hsp70* vector and the resulting plasmid was used together with Tol2 transposase to create stable transgenic lines as previously described (Kawakami, 2004). This line is designated w249.

KI: hoxa13b-FLAG-GFP: hoxa13b genomic DNA was amplified from the 5' UTR to a primer that removed the stop codon using the primers GTTGGATCCTGACGCACGC and GTCAACAAGTACAAGGGCATCAGTTATCGAT, with the naturally occurring BamH1 site underlined and the added ClaI site in italics. This was inserted into a vector such that following the stop codon is a sequence encoding three FLAG tags, followed by the viral 2A peptide and then GFP. This was then flanked on both sides by the *Mbait* sequence (Kimura et al., 2014). The plasmid was injected into one-cell stage embryos together with Cas9 and gRNAs targeting the *Mbait* sequence and 5' and 3' untranslated regions of the *hoxa13b* gene to produce a CRISPR/Cas9 mediated knockin fish. Fish were screened for GFP expression in the posterior region. This line is designated w250.

HS:rbpja-FLAG-GFP: the *rbpja* coding region was amplified from 15s zebrafish embryo cDNA using the primers *GGATCC*AAGATGGCGCCTGTTGTGACAG and CCTCTGCCATGTCCGTCTCCT*ATCGAT*, with the added BamH1 and ClaI sites shown in italics, and inserted into the same vector used for *hoxa13b* as described above that has the viral 2A sequence between the GFP and the Rbpja. A transgenic line was made with Tol2, and designated w251.

HS:EnR-rbpja-FLAG-GFP: an Engrailed repressor domain was inserted in frame in front of the *rbpja* coding region in the *HS:rbpja-FLAG-GFP* plasmid. A transgenic line was made with Tol2, and designated w252.

Embryos were heat shocked in a circulating water bath for 30 mins at 40°C, unless otherwise noted.

Zebrafish *tbxta* promoter reporter lines. A 2.1 kb fragment of the *tbxta* promoter was cloned from genomic DNA using *GAATTC*ATACAATTCCTTTGTGCTGTTGCAACAC, which places an EcoRl I site at the 5' end, and *CCATGG*TTCCGATCAAATAAAGCTTGAGATAAGTCCG, which places a Nco1 site at the ATG, and inserted into a vector containing GFP and Tol2 elements. Cloning of the upstream elements used the following primers: for Hox element 2 *GTCGAC*CGTTGTTTAAATAAAACGGCGAGATACATG and GTCAGATATGGACAAACCCATCCAT*CTCGAG*, using Sal1 and Xho1, respectively; for Hox element 1, *GTCGAC*TTTAACCTTGACAAGTGTGAATAGCTG and *ATCGAT*ACGAATAATTAACTAATAATTTTGATTTCAACTGTAC, using Sal1 and Cla1, respectively; for the element at -5.4 kb *ATCGAT*TCCTAAAGGTCTGACATGTACTGCG and *GAATTC*GGTTTTTCATTGTAGTTAACTGTGAGAGC,using Cla1 and EcoR1, respectively.

Catfish *tbxta* promoter reporter line. The 5.7 kb *tbxta* promoter was cloned with *GTCGAC*CGTCAACCTAGACACATTTAAATTTGGC, which places a Sal1 site at the 5' end and *CCATGG*TTCCAGTCTTATTGGGGGAAAAGCC, which places a Nco1 site at the ATG, and inserted into a vector containing GFP and Tol2 elements.

Hox element 1 deletion lines: these were produced using the gRNAs listed below together with Cas9. *Tbxta*$^{\Delta 312}$ has a 312 bp deletion: AAAGCTTGTCCTGTAGGGGGTGCTATΔGATCCTATACTGTGCTCCAGACTCCACAGA. *Tbxta*$^{\Delta 314}$ has a 323 bp deletion and 9 bp insertion (insertion in italics) GCTTGTCCTGTAGGGGGTΔ*CTGCTGGCC*ΔTCCTATACTGTGCTCCAGA. In both cases the Δ indicates the region deleted.
Embryos with the deletions were screened using the primers GCTGCACCCAAGAAAAGCAA and GTCACTTTGTTCAACTGTAGCGT using standard PCR conditions, and analyzed on 2% agarose gels.

All animal protocols used here were approved by the University of Washington Institutional Animal Care and Use Committee.

gRNAs

The gRNAs used to make the knockin line are GAGCTGCTGGGCTCCATGTA and GGGCTTGATATTGGTGGTAT. The *Mbait* gRNA is GGCTGCTGCGGTTCCAGAGG. The *tbxta* gRNAs to remove Hox element 1 are CCTGTAGGGGGTGCTATCTC and GAGCACAGTATAGGATCTGG.

CUT&RUN

The basic CUT&RUN procedure followed that described in Skene and Henikoff (2017) except that a pAG/MNase (Meers et al., 2019) was used instead of the original pA/MNase. Briefly, zebrafish tailbuds were dissected at the boundary of the third newly formed somite from groups of 30 embryos at the 15 to 18s stage, with the epidermis first removed prior to cutting as previously described (Manning and Kimelman, 2015). Explants were then dissociated

into single cells with wash buffer (20 mM HEPES, pH7.5, 150 mM NaCl, 0.5 mM spermidine and a protease inhibitor (Roche; one complete tablet per 50 ml added fresh) by gentle pipetting up and down 50 times using a 1-ml pipette. Tubes were pre-coated with 1% BSA overnight at 4°C to avoid sticking of cells to the wall. The dissociated single cells were then washed twice, bound to activated Concanavalin A-coated magnetic beads, and suspended in 250 µl antibody buffer (wash buffer containing 0.03% [wt/vol] digitonin and 2 mM EDTA). Antibody was added immediately to a final concentration of 1:100 and cells were incubated overnight at 4°C on a nutator. After incubation, beads/cells were washed three times with Dig-wash buffer (wash buffer containing 0.03% [wt/vol] digitonin) to get rid of the unbound antibody, followed by resuspension and incubation of beads/cells in 150 µl Dig-wash buffer with pAG/MNase (1:100 dilution), at 4°C for 1 hour. After two washes with Dig-wash buffer, beads/cells were resuspended in 100 µl Dig-wash buffer and chilled to 0°C. The pAG/MNase was activated by adding 2 µl 100 mM $CaCl_2$ and the cutting reaction was allowed to continue for 1 hour at 0°C. Equal volume (100 µl) of 2X stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% Digitonin, 100 µg/mL RNase A and 50 µg/mL glycogen) was added to stop the reaction, followed by 30 min incubation in a 37°C water bath to release the cut DNA fragments from the cells. Beads/cells were placed on a magnetic stand and the liquid containing the cut DNA fragments was transferred to a new tube. DNA was extracted using the phenol-chloroform-extraction method.

The CUT&RUN experiment was done in duplicates for *HS:hoxa13b-FLAG-GFP* and histone mark H3K4me1 and H3K27ac, and triplicates for *KI:hoxa13b-FLAG-GFP*. WT embryos with the anti-FLAG antibody was included as the control for each replicate CUT&RUN experiment. Antibodies used were: anti-FLAG (Millipore F3165), anti-H3K4me1 (Abcam ab8895) and anti-H3K27ac (Abcam ab4729).

Raw data from the CUT&RUN analysis is available from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with the BioProject ID PRJNA719435.

DNA sequencing and CUT&RUN data analysis

Libraries for next generation sequencing were prepared using the NEBNext Ultra II DNA Library Prep Kit, NEB (Ipswich, USA) for transcription factors following the protocol of Liu (https://dx.doi.org/10.17504/protocols.io.wvgfe3w), with some modifications for histone marks due to larger Cut&Run fragment size (Meers et al., 2019). Modifications for histone mark samples include: 1) In step 2, PCR program for End Prep 20°C for 30 min, 58°C for 45 min and hold at 4°C; 2) In step 8, 1.5 X AMPure XP beads size selection was applied after adaptor ligation; 3) In step 19, 0.65 X AMPure XP beads size selection was applied for the first round size selection to remove PCR products larger than 500 bp.

Libraries were pooled at equimolar concentrations and paired-end (PE) 150 bp sequencing was performed on a Illumina HiSeq 4000 platform by Novogene Corporation Inc. (California, USA). Bioinformatics analysis was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. Paired reads were quality checked and trimmed using Trim Galore version 0.4.4_dev (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Only reads longer than 20 bp after trimming were kept. Trimmed PE reads were aligned to zebrafish genome (danRer11 masked) using Bowtie2 version 2.3.4.1 (Langmead and Salzberg, 2012) with options: --local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33. The resulting bam files containing the alignment information were converted to bed format using Samtools 1.9 (Li et al., 2009) and Bedtools 2.29.2 (Quinlan and Hall, 2010), or bigwig format using a custom R script (R version 3.6.2) and visualized with the Integrated Genomics Viewer (IGV version 2.7.2, Robinson et al., 2011). Assessment of mapped fragment size distribution, replicate reproducibility and heatmap of reads over the peak regions were plotted following the scripts of Henikoff et al. (2020). Fragments of 24-121 bp and 149-500bp were used for peak calling and downstream analysis of transcription factors and histone marks, respectively. MACS2 version 2.1.1.20160309 (Zhang et al., 2008) was used for peak calling with parameters: callpeak -t

input_file -q 1e-2 -f BEDPE -keep-dup all -n output_file_name. The broadpeak option was used for histone mark peak calling.

Peak filtering and Annotation

Overlapping peaks between the duplicates of each group, or between the triplicates for *KI:hoxa13b,* were kept and used for subsequent analysis. The overlap analysis was performed using the overlapsAny function in the GenomicRanges R package (Lawrence et al., 2013), with the standard of overlapping set at a minimum of 1 bp. The resulting peaks of the *HS:hoxa13b* sample were thought to be the direct targets of Hoxa13b and they are further annotated by overlapping with the *KI:hoxa13b* peaks to add another level of confidence, since these two lines have different advantages in terms of Hoxa13b target profiling. The potential target genes near the peaks were annotated based on the TxDb.Drerio.UCSC.danRer11.refGene annotation database using a custom R script and ChIPseeker (Yu et al., 2015) was used for genome ontology analysis to generate Figure 3A and Figure S2A.

Motif discovery

Sequences of the ± 200 bp summit center region of the Hoxa13b peaks were used for comprehensive motif analysis including *de novo* motif discovery (MEME and DREME), similarity to known motifs (Tomtom) and central enrichment analysis (CentriMo) using the MEME-ChIP program, which integrates the above mentioned programs and automatically group significant motifs by similarity (Machanick and Bailey, 2011).

Footprint

The top-rated motif discovered was the Hoxa13b motif and therefore the footprint analysis was done using this motif. Single-base resolution cutting probability of the MNase around the peak regions containing the Hoxa13b motif was calculated using a custom R script (R version 3.6.2) and plotted using a R package ggplot2 (Wickham et al., 2016)to reveal the general footprint of Hoxa13b.

In situ hybridization and immunofluorescence

Alkaline phosphate in situ hybridization used standard conditions (https://wiki.zfin.org/display/prot/Thisse+Lab+-+In+Situ+Hybridization+Protocol+-+2010+update). Fluorescent in situ hybridization use the Lauter protocol (Lauter et al., 2011). Immunofluorescence was performed using the MF20 antibody (Developmental Studies Hybridoma Bank) for somite muscle staining and the anti-FLAG antibody (Millipore) for revealing Hoxa13b-FLAG expression in the *KI:Hoxa13b-FLAG-GFP* embryos. A standard immunofluorescence stain protocol was used. Briefly, fixed embryos stored in MeOH were rehydrated into PBST and blocked for 1 hour at RT, then incubated with the first antibody (1:50 dilution for MF20 and 1:200 for anti-FLAG) overnight at 4°C. After 3 washes with PBST (15 mins each time), embryos were incubated with the secondary antibody (Abcam ab150113) for 2 hrs at room temperature, then washed with PBST and stepped into 70% glycerol for later imaging.

Imaging and measurement

Brightfield Images for in situ hybridization and for body length measurement were taken using a Nikon AZ100 microscope with white light illumination. Wholemount fluorescence images shown in Figure 5B-E (first and second rows) were taken with the same microscope using a fluorescence illumination. Confocal images shown in Figure 1C-F, Figure S5, Figure S6 and the zoomed-in images shown in Figure 5B-E (third row) were obtained using an Olympus Fluoview 1200 scanning confocal microscope with a 40X oil lens for fixed embryos mounted on a slide or a 40X dipping lens for live embryos. The 3D image stacks were taken at a step-size of 2 microns. Three-D reconstruction of the stacked images was done using Imaris 9.2 (Oxford Instruments) and the posterior somite and neural tube volume data was generated by Imaris after manual creation of objectives based on the fluorescent signal.

Acknowledgements

We thank Derek Janssens and Steve Hennikoff for kindly supplying Protein AG/MNase and valuable technical advice, Colin Kenny and Hannah Arbach for helpful discussions, Lauren Saunders for the help with Tapestation analysis, Alex Chitsazan for bioinformatics help, and Rex Dunham and Baofeng Su for providing catfish genomic DNA. D.K. and A.W. were supported by grants from the National Institutes of Health (RO1GM079203 to DK and RO1NS099124 to AW).

REFERENCES

Aires, R., de Lemos, L., Novoa, A., Jurberg, A.D., Mascrez, B., Duboule, D., and Mallo, M. (2019). Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. Dev Cell *48*, 383-395 e388.

Amandio, A.R., Lopez-Delisle, L., Bolt, C.C., Mascrez, B., and Duboule, D. (2020). A complex regulatory landscape involved in the development of mammalian external genitals. Elife *9*.

Arnold, S.J., Stappert, J., Bauer, A., Kispert, A., Herrmann, B.G., and Kemler, R. (2000). Brachyury is a target gene of the Wnt/beta-catenin signaling pathway. Mech Dev *91*, 249-258.

Attardi, A., Fulton, T., Florescu, M., Shah, G., Muresan, L., Lenz, M.O., Lancaster, C., Huisken, J., van Oudenaarden, A., and Steventon, B. (2018). Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. Development *145*.

Beddington, R.S., Rashbass, P., and Wilson, V. (1992). Brachyury--a gene affecting mouse gastrulation and early organogenesis. Dev Suppl, 157-165.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J., and Gomez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res *22*, 2043-2053.

Bouldin, C.M., Manning, A.J., Peng, Y.-H., Farr, G.H.I., Hung, K.L., Dong, A., and Kimelman, D. (2015). Wnt signaling and tbx16 form a bistable switch to commit bipotential progenitors to mesoderm Development *142*, 2499-2507.

Bulajic, M., Srivastava, D., Dasen, J.S., Wichterle, H., Mahony, S., and Mazzoni, E.O. (2020). Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns. Development *147*.

Cadigan, K.M., and Waterman, M.L. (2012). TCF/LEFs and Wnt signaling in the nucleus. Cold Spring Harb Perspect Biol *4*.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature *389*, 85-89.

de la Calle Mustienes, E., Gomez-Skarmeta, J.L., and Bogdanovic, O. (2015). Genome-wide epigenetic cross-talk between DNA methylation and H3K27me3 in zebrafish embryos. Genom Data *6*, 7-9.

Denans, N., Iimura, T., and Pourquie, O. (2015). Hox genes control vertebrate body elongation by collinear Wnt repression. Elife *4*.

Desanlis, I., Kherdjemil, Y., Mayran, A., Bouklouch, Y., Gentile, C., Sheth, R., Zeller, R., Drouin, J., and Kmita, M. (2020). HOX13-dependent chromatin accessibility underlies the transition towards the digit development program. Nat Commun *11*, 2491.

Echeverri, K., and Oates, A.C. (2007). Coordination of symmetric cyclic gene expression during somitogenesis by Suppressor of Hairless involves regulation of retinoic acid catabolism. Dev Biol *301*, 388-403.

Galceran, J., Farinas, I., Depew, M.J., Clevers, H., and Grosschedl, R. (1999). Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes Dev *13*, 709-717.

Garriock, R.J., Chalamalasetty, R.B., Kennedy, M.W., Canizales, L.C., Lewandoski, M., and Yamaguchi, T.P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. Development *142*, 1628-1638.

Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. PLoS Biol *12*, e1001937.

Harvey, S.A., Tumpel, S., Dubrulle, J., Schier, A.F., and Smith, J.C. (2010). no tail integrates two modes of mesoderm induction. Development *137*, 1127-1135.

Henikoff, S., Henikoff, J.G., Kaya-Okur, H.S., and Ahmad, K. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. Elife *9*.

Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. Development *142*, 2864-2875.

Jahangiri, L., Nelson, A.C., and Wardle, F.C. (2012). A cis-regulatory module upstream of deltaC regulated by Ntla and Tbx16 drives expression in the tailbud, presomitic mesoderm and somites. Dev Biol *371*, 110-120.

Johnson, J.E., and Macdonald, R.J. (2011). Notch-independent functions of CSL. Curr Top Dev Biol *97*, 55-74.

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G.*, et al.* (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327-339.

Jurberg, A.D., Aires, R., Novoa, A., Rowland, J.E., and Mallo, M. (2014). Compartment-dependent activities of Wnt3a/beta-catenin signaling during vertebrate axial extension. Dev Biol *394*, 253-263.

Kawakami, K. (2004). Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. Methods Cell Biol *77*, 201-222.

Kimelman, D. (2016a). A novel cold-sensitive mutant of ntla reveals temporal roles of brachyury in zebrafish. Dev Dyn *245*, 874-880.

Kimelman, D. (2016b). Tales of tails (and trunks): forming the posterior body in vertebrate embryos. Curr Topics in Dev Biol *116*, 517-536.

Kimelman, D., and Martin, B.L. (2012). Anterior-Posterior patterning in early development: Three stratagies. WIREs Dev Biol *1*, 253-266.

Kimura, Y., Hisano, Y., Kawahara, A., and Higashijima, S. (2014). Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Sci Rep *4*, 6545.

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

Lauter, G., Soll, I., and Hauptmann, G. (2011). Multicolor fluorescent in situ hybridization to define abutting and overlapping gene expression in the embryonic zebrafish brain. Neural Dev *6*, 10.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput Biol *9*, e1003118.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, N., Hargreaves, V.V., Zhu, Q., Kurland, J.V., Hong, J., Kim, W., Sher, F., Macias-Trevino, C., Rogers, J.M., Kurita, R.*, et al.* (2018). Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch. Cell *173*, 430-442 e417.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696-1697.

Mallo, M. (2020). The vertebrate tail: a gene playground for evolution. Cell Mol Life Sci *77*, 1021-1030.

Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. Curr Top Dev Biol *88*, 63-101.

Manning, A.J., and Kimelman, D. (2015). Tbx16 and Msgn1 are required to establish directional cell migration of zebrafish mesodermal progenitors. Dev Biol *406*, 172-185.

Martin, B.L., and Kimelman, D. (2008). Regulation of canonical Wnt signaling by Brachyury is essential for posterior mesoderm formation. Dev Cell *15*, 121-133.

Martin, B.L., and Kimelman, D. (2012). Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. Dev Cell *22*, 223-232.

Meers, M.P., Bryson, T.D., Henikoff, J.G., and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. Elife *8*.

Morley, R.H., Lachani, K., Keefe, D., Gilchrist, M.J., Flicek, P., Smith, J.C., and Wardle, F.C. (2009). A gene regulatory network directed by zebrafish No tail accounts for its roles in mesoderm formation. Proc Natl Acad Sci U S A *106*, 3829-3834.

Northrop, J., Woods, A., Seger, R., Suzuki, A., Ueno, N., Krebs, E., and Kimelman, D. (1995). BMP-4 regulates the dorsal-ventral differences in FGF/MAPKK-mediated mesoderm induction in Xenopus. Dev Biol *172*, 242-252.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell *133*, 1277-1289.

Oka, C., Nakano, T., Wakeham, A., de la Pompa, J.L., Mori, C., Sakai, T., Okazaki, S., Kawaichi, M., Shiota, K., Mak, T.W.*, et al.* (1995). Disruption of the mouse RBP-J kappa gene results in early embryonic death. Development *121*, 3291-3301.

Provost, E., Rhee, J., and Leach, S.D. (2007). Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. Genesis *45*, 625-629.

Pyati, U.J., Webb, A.E., and Kimelman, D. (2005). Transgenic zebrafish reveal stage-specific roles for Bmp signaling in ventral and posterior mesoderm development. Development *132*, 2333-2343.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Row, R.H., Tsotras, S.R., Goto, H., and Martin, B.L. (2016). The zebrafish tailbud contains two independent populations of midline progenitor cells that maintain long-term germ layer plasticity and differentiate in response to local signaling cues. Development *143*, 244-254.

Sambasivan, R., and Steventon, B. (2020). Neuromesodermal Progenitors: A Basis for Robust Axial Patterning in Development and Evolution. Front Cell Dev Biol *8*, 607516.

Sieger, D., Tautz, D., and Gajewski, M. (2003). The role of Suppressor of Hairless in Notch mediated signalling during zebrafish somitogenesis. Mech Dev *120*, 1083-1094.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*.

Steventon, B., and Martinez Arias, A. (2017). Evo-engineering and the cellular and molecular origins of the vertebrate spinal cord. Dev Biol *432*, 3-13.

Takeuchi, M., Okabe, M., and Aizawa, S. (2009). The genus Polypterus (bichirs): a fish group diverged at the stem of ray-finned fishes (Actinopterygii). Cold Spring Harb Protoc *2009*, pdb emo117.

Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S.*, et al.* (2014). Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. Development *141*, 1209-1221.

Tuazon, F.B., and Mullins, M.C. (2015). Temporally coordinated signals progressively pattern the anteroposterior and dorsoventral body axes. Semin Cell Dev Biol *42*, 118-133.

Turner, D.A., Hayward, P.C., Baillie-Johnson, P., Rue, P., Broome, R., Faunes, F., and Martinez Arias, A. (2014). Wnt/beta-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells. Development *141*, 4243-4253.

Veeman, M.T., Slusarski, D.C., Kaykas, A., Louie, S.H., and Moon, R.T. (2003). Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. Curr Biol *13*, 680-685.

Vonica, A., and Gumbiner, B.M. (2002). Zygotic Wnt activity is required for Brachyury expression in the early Xenopus laevis embryo. Dev Biol *250*, 112-127.

Ward, A.B., and Brainerd, E.L. (2007). Evolution of axial patterning in elongate fishes. Biological Journal of the Linnean Society *90*, 97-116.

Wickham, H., Navarro, D., and Pedersen, T.L. (2016). ggplot2: Elegant Graphics for Data Analysis (New York, USA: Spring-Verlag New York).

Wymeersch, F.J., Wilson, V., and Tsakiridis, A. (2021). Understanding axial progenitor biology in vivo and in vitro. Development *148*.

Yamaguchi, T.P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A.P. (1999). T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification. Genes Dev *13*, 3185-3190.

Ye, Z., and Kimelman, D. (2020). Hox13 genes are required for mesoderm formation and axis elongation during early zebrafish development. Development *147*, 1-12.

Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.N.*, et al.* (2009). Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. Dev Cell *17*, 516-526.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382-2383.

Zheng, C., and Chalfie, M. (2016). Securing Neuronal Cell Fate in C. elegans. Curr Top Dev Biol *116*, 167-180.

Zheng, C., Jin, F.Q., and Chalfie, M. (2015). Hox Proteins Act as Transcriptional Guarantors to Ensure Terminal Differentiation. Cell Rep *13*, 1343-1352.

FIGURE LEGENDS

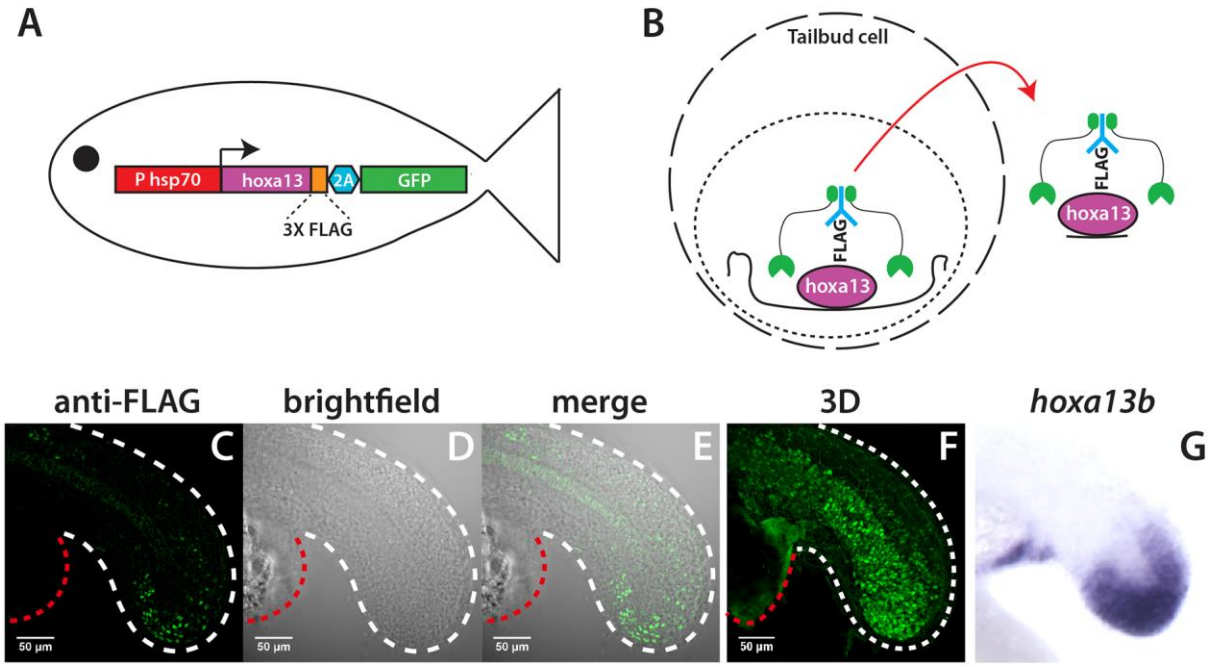


Figure 1 Identification of *in vivo* Hoxa13b targets with CUT&RUN

A) Schematic of *HS:hoxa13b-FLAG-GFP* transgenic line. B) CUT&RUN procedure. Cells isolated from dissected tailbuds are separated and permeabilized. First an antibody is added (in this case anti-FLAG), then a Protein A/G-Micrococcal nuclease (green), which cuts the genomic DNA locally, allowing transcription factor-DNA complexes to exit the cell where they are isolated and analyzed by DNA sequencing. C-F) Expression of Hoxa13b-FLAG in *KI:Hoxa13b-FLAG-GFP* embryos as shown in a 20s stage embryo with an anti-FLAG antibody. Shown are the lateral view of a middle slice of the tail (C-E) and an image from the 3D reconstruction (F). The white dotted line shows the posterior body border and red line shows the yolk extension. See also Video S1 for a 360° view of the tail. G) In situ hybridization showing *hoxa13b* transcripts in the tailbud of a 20s stage WT embryo for comparison.

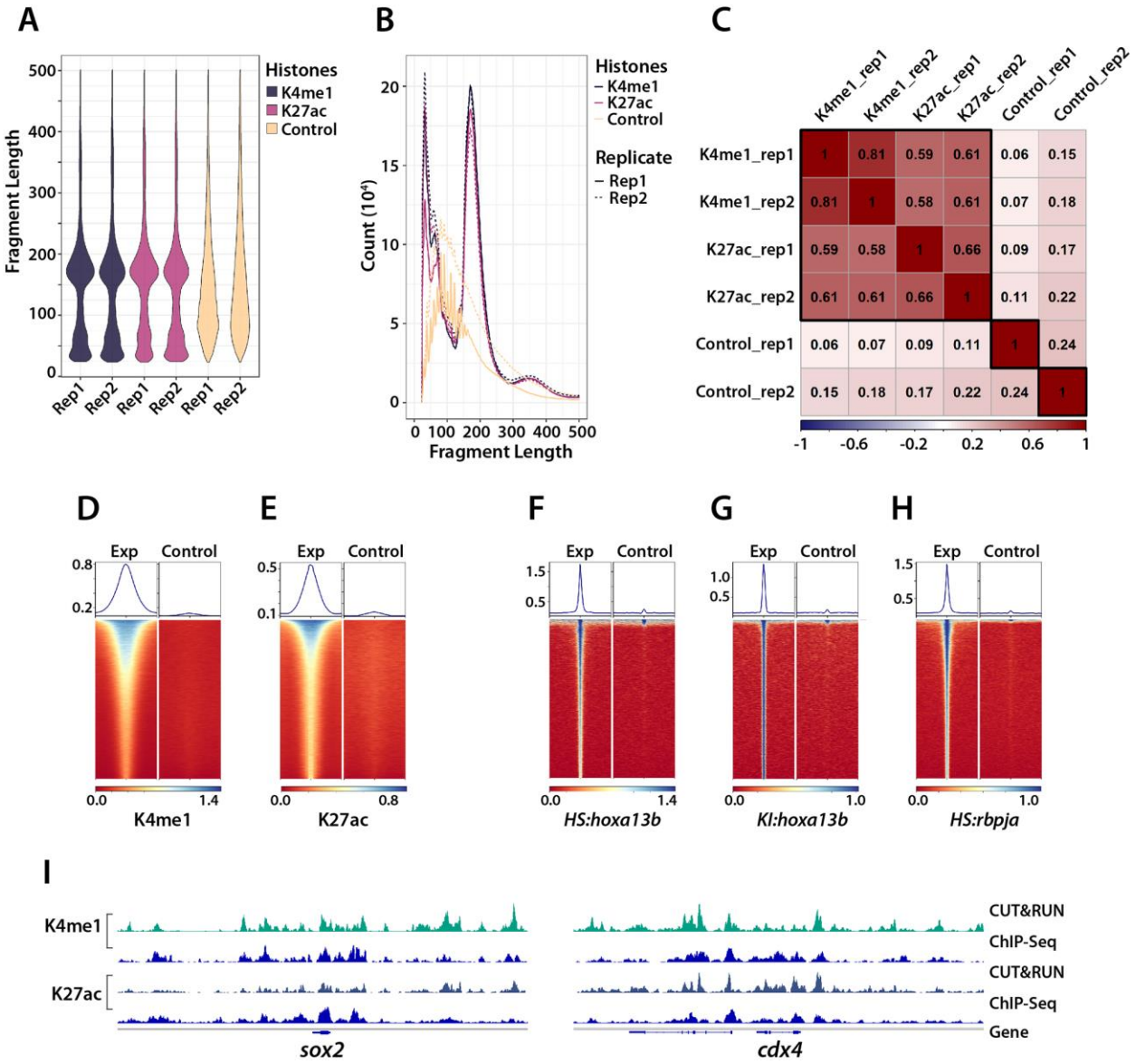


Figure 2 Quality assessment of the CUT&RUN

A,B) CUT&RUN fragment size distribution. Plots showing the proportion of DNA for each fragment size (A) and raw number of fragments by size (B). Both biological replicates (rep) are shown in each panel. The histone mark samples show strong enrichment of mono-nucleosome sized fragments whereas the control samples show a random size distribution. C) Assessment of replicate reproducibility. The genome was subdivided into 500 bp bins and the reads for each sample in each bin was assessed, and then subjected to correlation analysis between the samples. The high correlation between the replicates of the histone modification samples

shows the reproducibility of the CUT&RUN experiment. Note that the activating modifications (H3K27ac and H3K4me1) were generally well correlated. D-H) Heatmaps (bottom) and plots (top) showing the signal within ± 1 kb of the CUT&RUN peak summit center for both the experimental (Exp) and control samples. I) comparison of CUT&RUN results (light green and dark green tracks) to previous ChIP-Seq data (blue tracks) for two posteriorly expressed genes, *sox2* and *cdx4*.

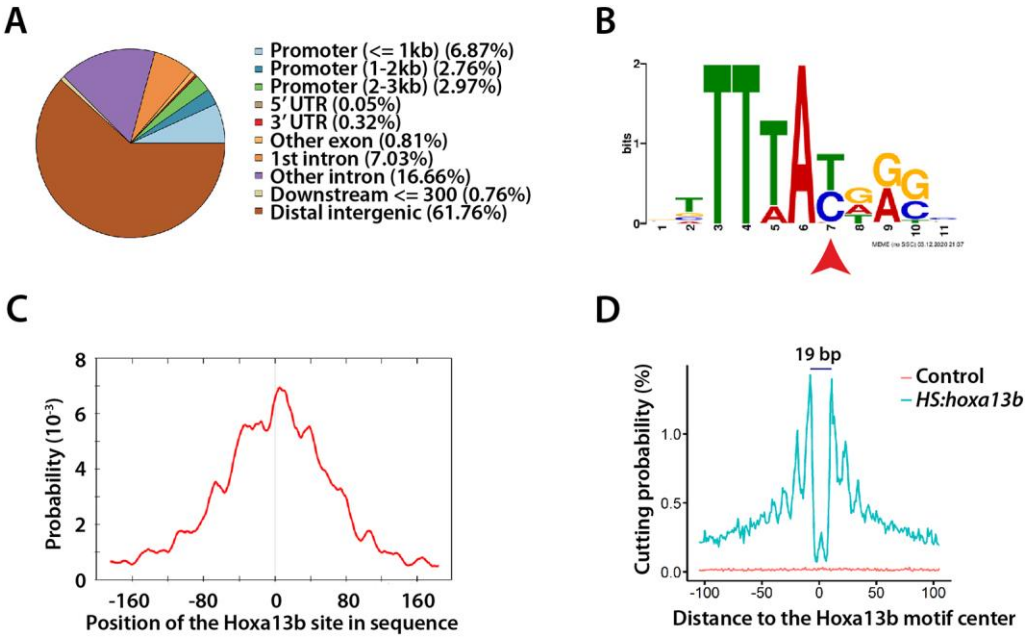


Figure 3 Hoxa13b binding sites and motif analysis

A) Distribution of the Hoxa13b target sites in the genome. The large majority of the sites are in distal intergenic regions. B) Hoxa13b binding motif discovered from analysis of the CUT&RUN data. The red arrowhead indicates a base that as a T is bound by all Hox proteins whereas as a C it is bound by posterior Hox proteins based on *in vitro* studies. C) The Hoxa13b motif shown in panel B is centrally located among all the CUT&RUN DNA fragments that contain a Hox motif, as expected. D) The Hoxa13b forms a 19 bp footprint at the Hoxa13b motif sites in the *HS:hoxa13b* sample (teal) while in the control sample the DNA cutting by MNase was random and rare at the same sites (red).

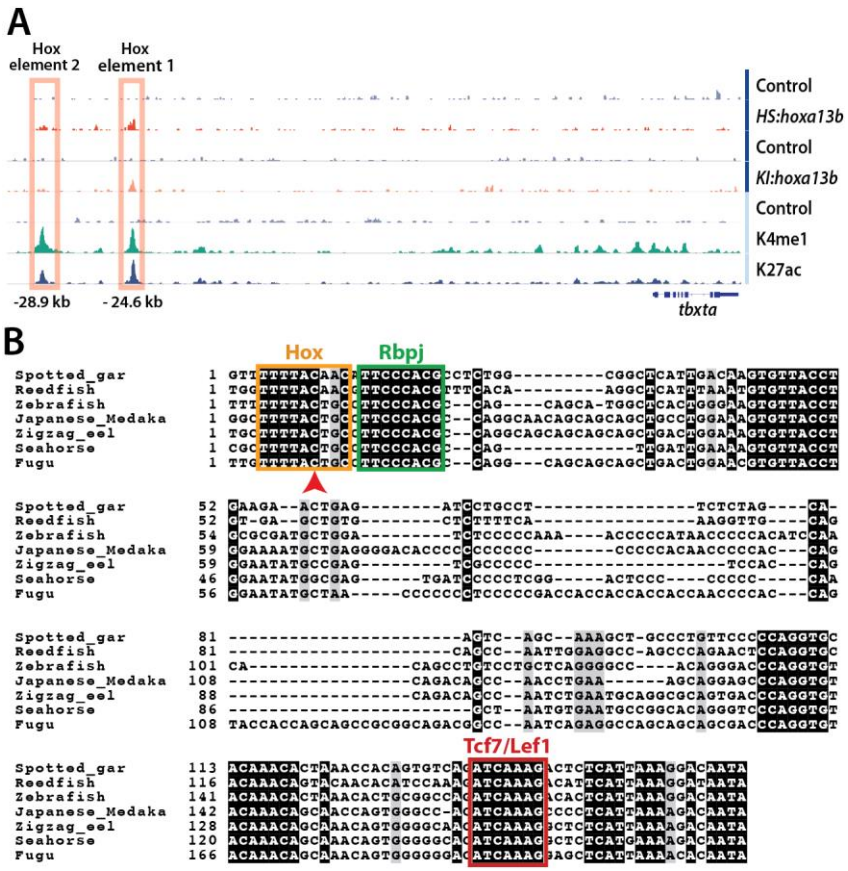


Figure 4 Hox elements upstream of the *tbxta* transcriptional start site

A) IGV tracks showing the two Hoxa13b binding elements discovered using CUT&RUN. A peak appears at the element 1 site in both the *HS:hoxa13b-FLAG* and *KI:hoxa13b-FLAG* lines, while a weaker signal exists at the element 2 site in the *HS:hoxa13b-FLAG* sample but not with *KI:hoxa13b-FLAG*. Both elements have peaks for H3K4me1 and H3K27ac. B) The alignment of Hox element 1 in a variety of species is shown, with consensus sites for Hox, Rbpj and Tcf7/Lef1 binding shown. In the Hox site is a C at the nucleotide position specifically recognized by the posterior Hox proteins (arrowhead).

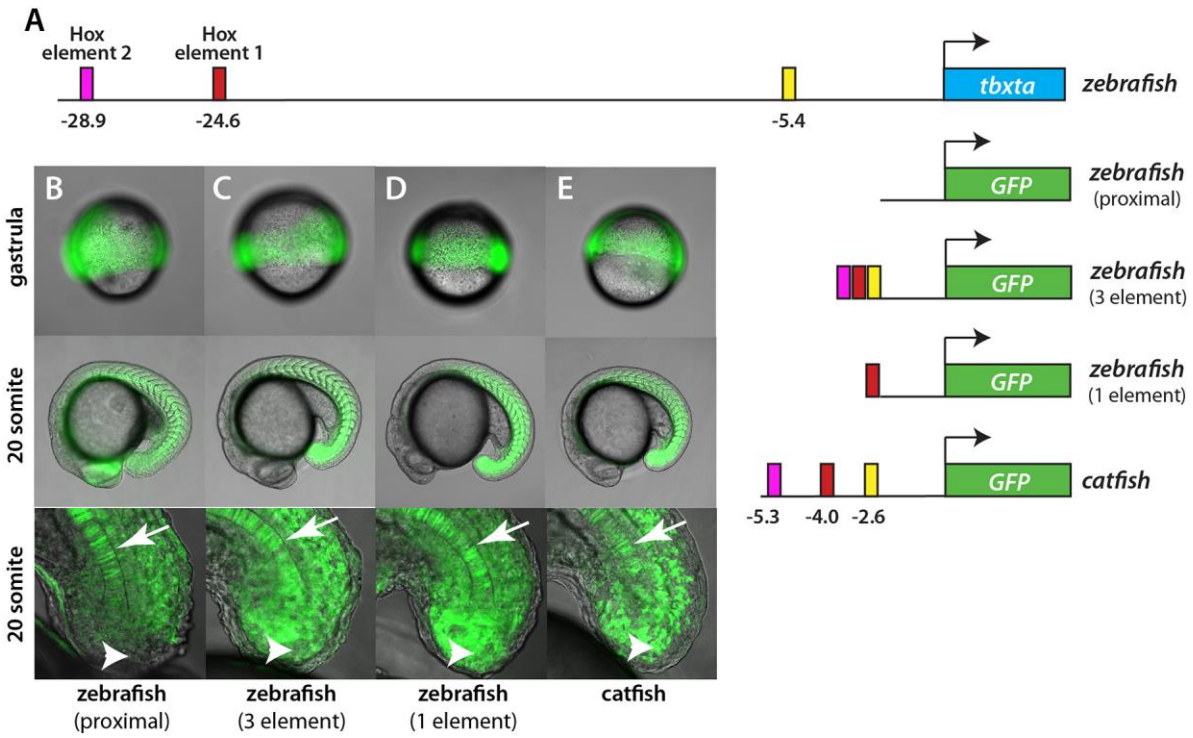


Figure 5 Hoxa13b binding elements regulate tailbud expression

A) Schematic of constructs used to make transgenic fish. The top line shows the position of the various elements in the zebrafish genome. B-E) GFP expression in transgenic lines at the gastrula and 20 somite stage embryos. The bottom panels show a closeup view of the tailbud, focused on the midline. In embryos expressing just the proximal 2.1 kb promoter, expression is strong in the gastrula and later in the notochord (arrows), but absent from the NMps (arrowheads) and mesoderm. With upstream elements added to the proximal promoter, or with the 5.7 kb catfish promoter, expression is strong in the gastrula stage and during somitogenesis, particularly in the somites (C-E). Some expression is observed in the neural tube because GFP expressed in the NMps perdures. Figure S5 shows a closeup of expression in the posterior somites and a comparison to marker gene expression in the NMps.

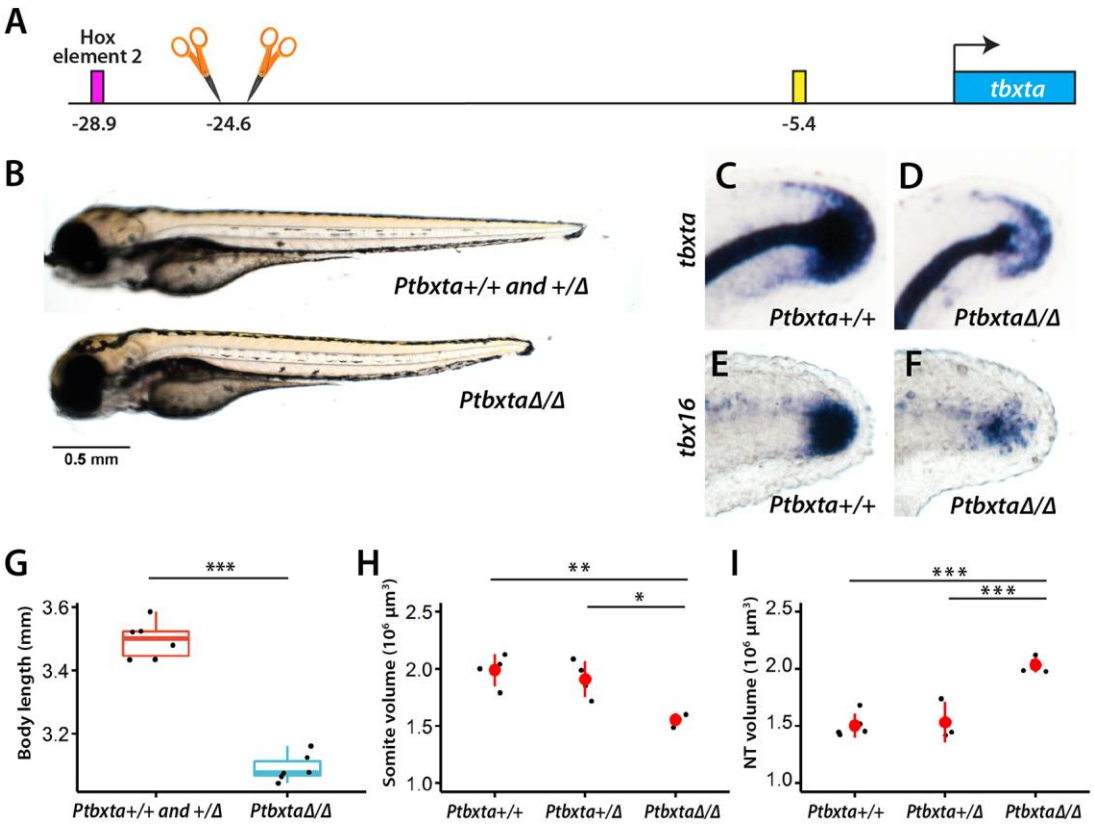


Figure 6 Hox element 1 deletion embryos have a shorter body and reduced *tbxta* expression

A) Two gRNAs flanking Hox element 1 were used to delete the element. B, G) A cross of heterozygous fish yielded embryos with short and longer bodies. The longer body embryos were genotyped to be either WT or heterozygous, whereas the shorter body embryos were homozygous for the deletion. The total length of embryos from both genotypes was compared (panel G, n=6 for each genotype; *** = $p < 0.001$, One-way ANOVA). C,D) Embryos were subjected to in situ hybridization with a *tbxta* or *tbx16* probe and then subsequently genotyped. In embryos that are homozygous for the deletion of Hox element 1 there is reduced expression in the NMps and tailbud mesoderm of both markers (D,F) compared to embryos without the deletion (C, E), but normal *tbxta* expression in the notochord. Note that in the homozygous embryos the posterior tip of the notochord is visible (D) whereas in WT embryos the posterior end of the embryo is hidden by the tailbud *tbxta* expression (C). The *tbxta* and *tbx16* embryos are at 24 and 28 hpf, respectively. Lateral view, dorsal at top.

H) Homozygous embryos show defects in the formation of posterior somites as shown by staining with the muscle antibody MF20 (see Figure S6A for image). The volume of the posterior somites starting from somite 16 is shown (n=4 for WT and heterozygous, and 3 for homozygous; * = $p < 0.05$, ** = $p < 0.01$, Tukey HSD test). I) Homozygous embryos show an enlarged neural tube (NT) as shown by *sox2* FISH (see Figure S6B for image). The volume of the posterior neural tube starting from the position of the 17^{th} somite is shown (n=5 for WT, 3 for heterozygous, and 4 for homozygous; *** = $p < 0.001$, Tukey HSD test).

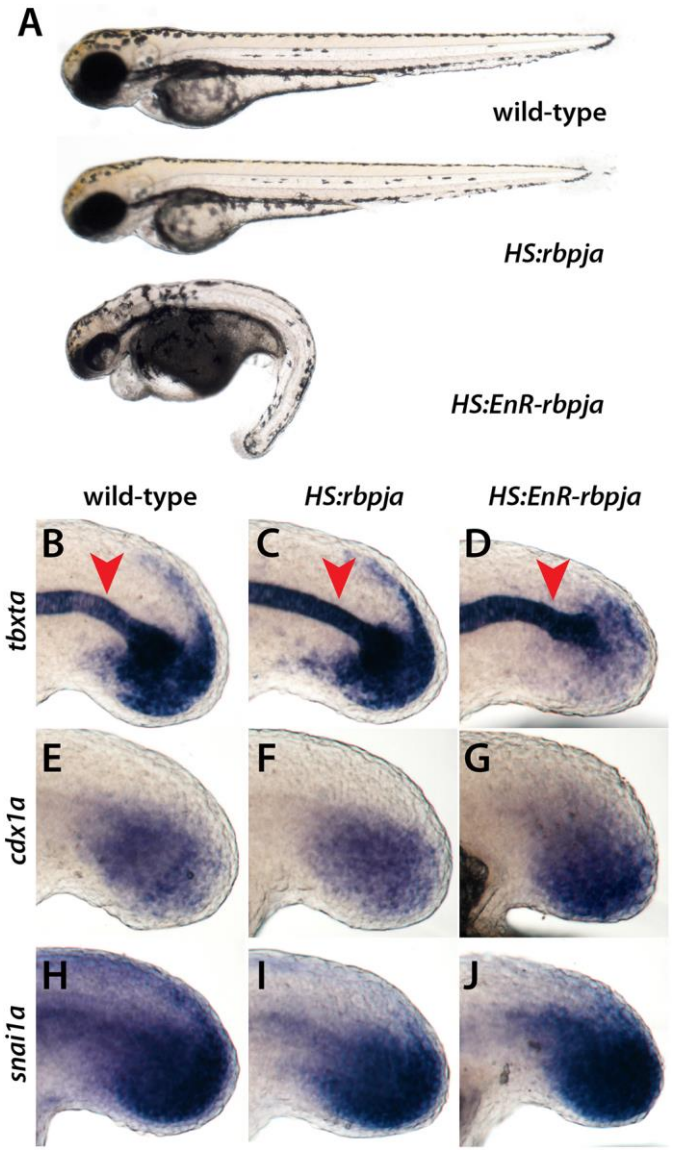


Figure 7 A dominant-negative Rbpj blocks expression of *tbxta*

A) Embryos were heat shocked at the 10s stage and left to develop. The dominant-negative (EnR-Rbpja) causes posterior defects. B-J) Embryos were heat shocked at the 10s stage and fixed 5 hours post-heat shock. Expression of *tbxta* in the tailbud is inhibited by EnR-Rbpja (D; 100%, n=19) but not by Rbpja (B, C; 100% for each, n=20 WT and 16 Rbpja), whereas the notochord expression is not inhibited (red arrowhead). The expression of *cdx1a* and *snai1a* are not inhibited by EnR-Rbpja (G, J; 100% in all cases, n= minimum of 14 embryos of each type).

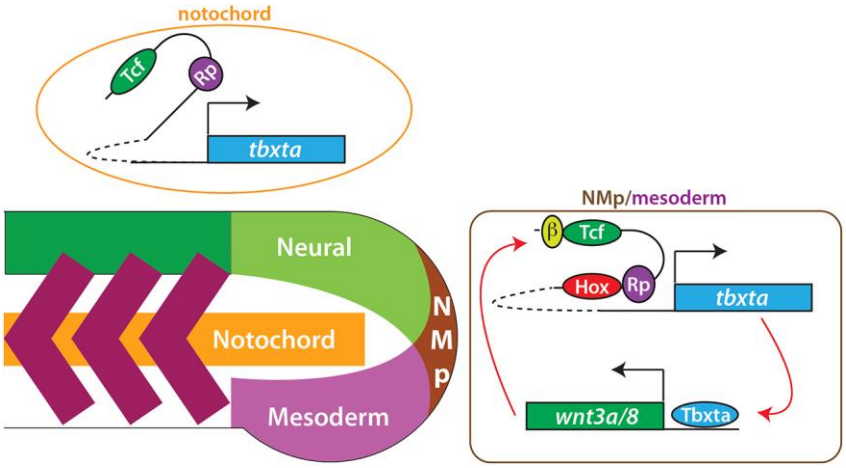


Figure 8 Model of posterior Hox regulation of *tbxta* expression

Tbxta and Wnt exist in a positive autoregulatory loop, with Tbxta directly activating *wnt3a* and *wnt8* expression (Martin and Kimelman, 2008). We propose that Wnts activate *tbxta* expression during the somitogenesis stages through the Tcf7/Lef1 (Tcf) sites in Hox binding element 1 and 2, located far upstream from the *tbxta* transcription start site, in collaboration with Hox proteins. While our data specifically shows binding of Hoxa13b to these elements, we suggest that all of the posterior Hox genes (Hox9-13) might act through these elements. In addition, Rbpj (Rp) binds to these elements to aid in transcriptional activation. In the notochord, the upstream elements are not necessary for expression during the somitogenesis stages.

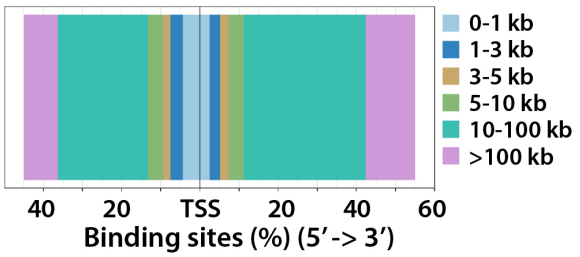
Species	Distance to Hox element 1 (kb)	Distance to Hox element 2 (kb)
Zebrafish	24.6	28.9
Fugu	3.8	4.3
Japanese medaka HNI	15.9	16.7
Channel catfish	4.0	5.3
Spotted gar	10.5	13.4
Tigertail seahorse	5.4	6.4
Zigzag eel	7.5	9.2
Reedfish	8.5	30.0

Table 1
Distance from the *tbxta* transcriptional start site to the two Hox elements is shown.

Figure S1 Comparison of heat shock transgenic lines. Refers to Figure 1.

Embryos were heat shocked at the 15s stage for 30 mins at 40°C and then left to develop at the standard temperature of 29°C. Whereas wildtype siblings show no phenotypic effects due to the heat shock, both transgenic lines produced the same posterior defects as described earlier (Ye and Kimelman, 2020), demonstrating that the FLAG tag does not alter Hoxa13b activity. Note that for both transgenic lines the fluorescent protein is produced separately from the Hoxa13b due to the 2A peptide that separates the two proteins.

A



B

		Rate (%)	E-value	Known motifs
MEME	1	78.8	1.2e-381	cad (MA0216.2), HOXC13 (MA0907.1), HOXD13 (MA0909.1)
	2	5.9	1.3e-030	BPC1 (MA1404.1), BPC6 (MA1402.1), RAMOSA1 (MA1416.1)
	3	11.9	1.8e-022	MSC (MA0665.1), Tcf12 (MA0521.1), hlh-1 (MA0545.1)
DREME	1	70.4	2.3e-225	HOXC10 (MA0905.1), Abd-B (MA0165.1), HOXD11 (MA0908.1)
	2	36.6	5.6e-083	HOXB13 (MA0901.1), HOXA13 (MA0650.1), HOXD13 (MA0909.1)
	3	20.8	2.8e-021	No

Figure S2 Hoxa13b binding targets genomic location and motif discovery. Refers to Figure 3.

A) Distribution of Hoxa13b targets according to their distance to the transcription start site (TSS) of the nearest gene. B) Top 3 motifs discovered by MEME and DREME. Rate shows the percentage of peaks that contributed to each motif. The 2nd ranked motif for MEME is composed of plant proteins, and the 3rd ranked motif has proteins are not expressed in the zebrafish tailbud (MSC), expressed at an extremely low level (Tcf12) or not homologous to any zebrafish protein (Hlh-1). Note that the DREME motif 2 is also partially included in motif 1, which is why the combined rate exceeds 100%. The first motif from MEME and the first and second motifs from DREME belong to the same cluster that has high similarity to known posterior Hox motifs.

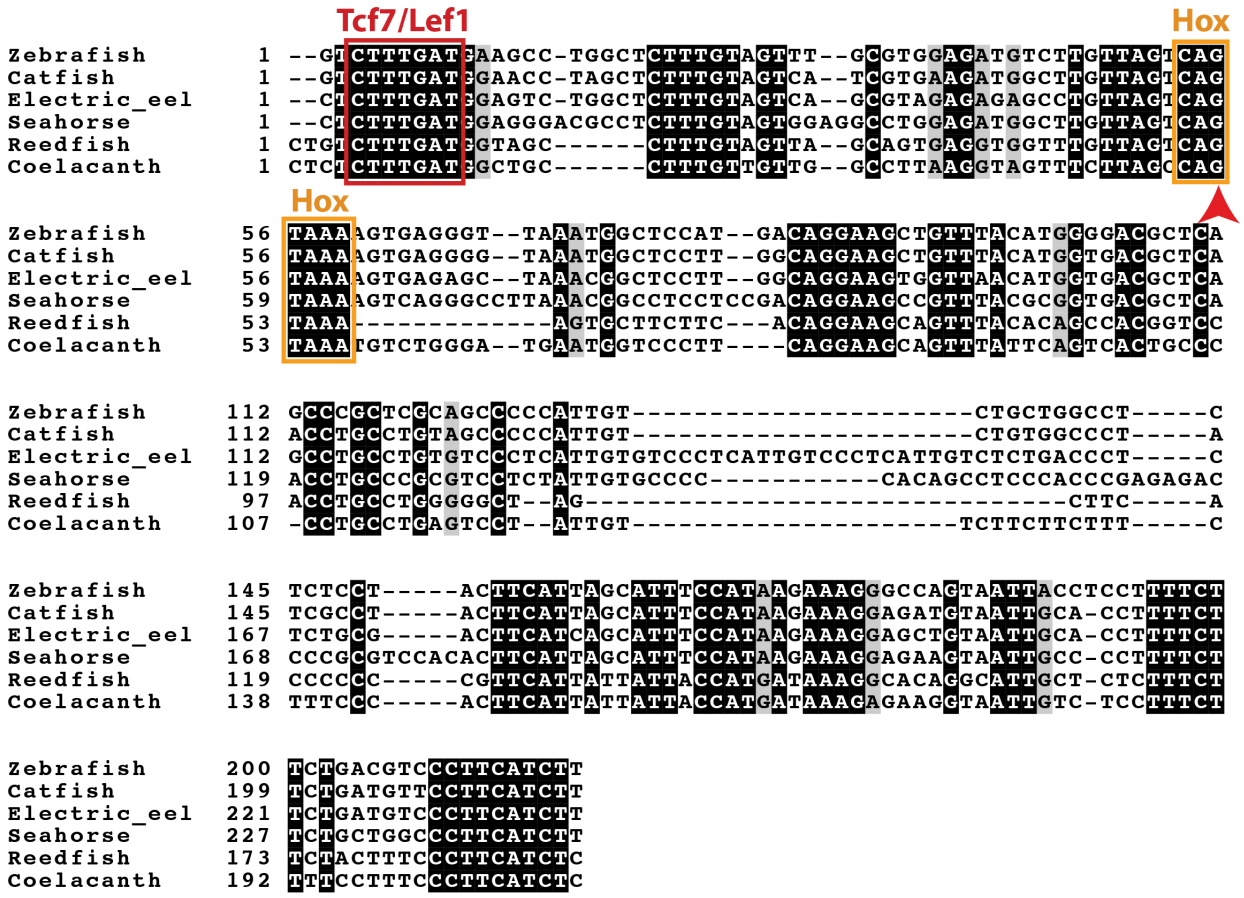


Figure S3 Hox element 2. Refers to Figure 4

The alignment of Hox element 2 in a variety of species is shown, with consensus sites for Hox and Tcf7/Lef1 binding shown. The Hox site has a C at the nucleotide position indicative of posterior Hox binding (arrowhead).

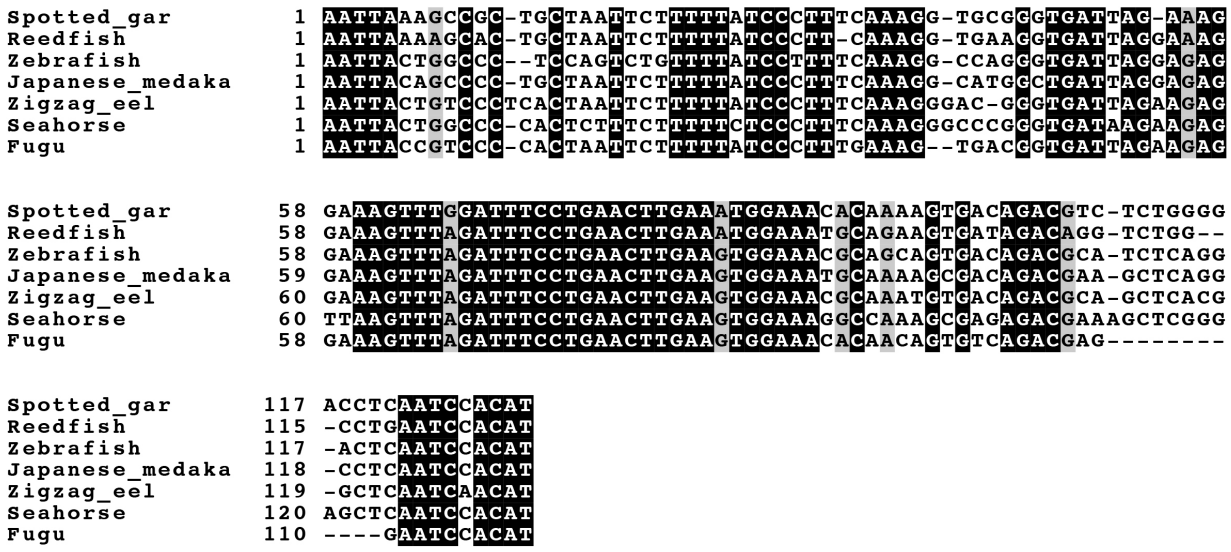


Figure S4 -5.4 kb element. Refers to Figure 5

The alignment of the conserved element located at approximately -5.4 kb in zebrafish relative to the same sequence in a variety of species is shown.

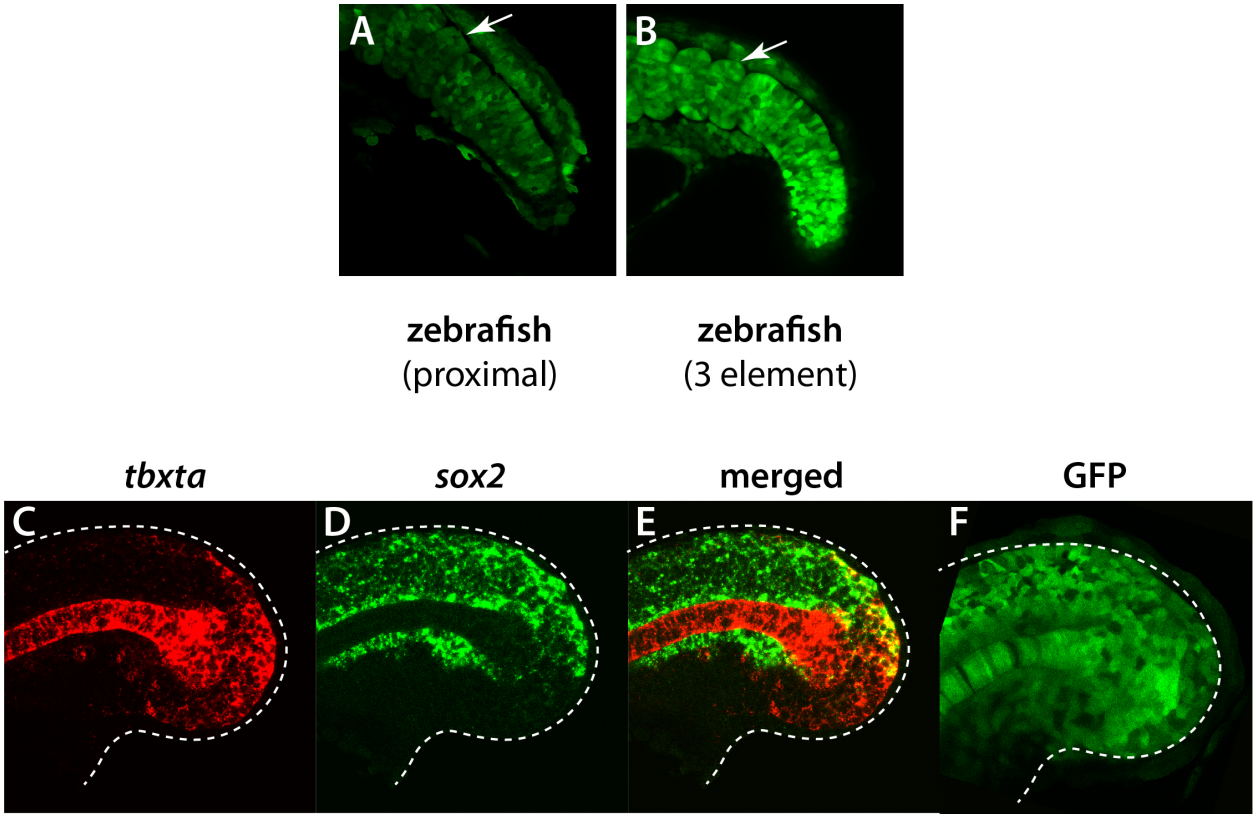


Figure S5 Hoxa13b binding elements regulate expression in the posterior somites and NMps. Refers to Figure 5.

A, B) GFP expression in the posterior somites of a 20s embryo from transgenic lines containing just the proximal *tbxta* promoter (A) or the proximal *tbxta* promoter with the three upstream elements (B). Note the strong somite GFP expression in panel B compared to A (arrows). (C-F) Comparison of GFP expression (F) from the transgenic line containing the *tbxta* promoter with the three elements to fluorescent in situ hybridization for *tbxta* (C) and *sox2* (D). The NMps are yellow in the merged panel (E).

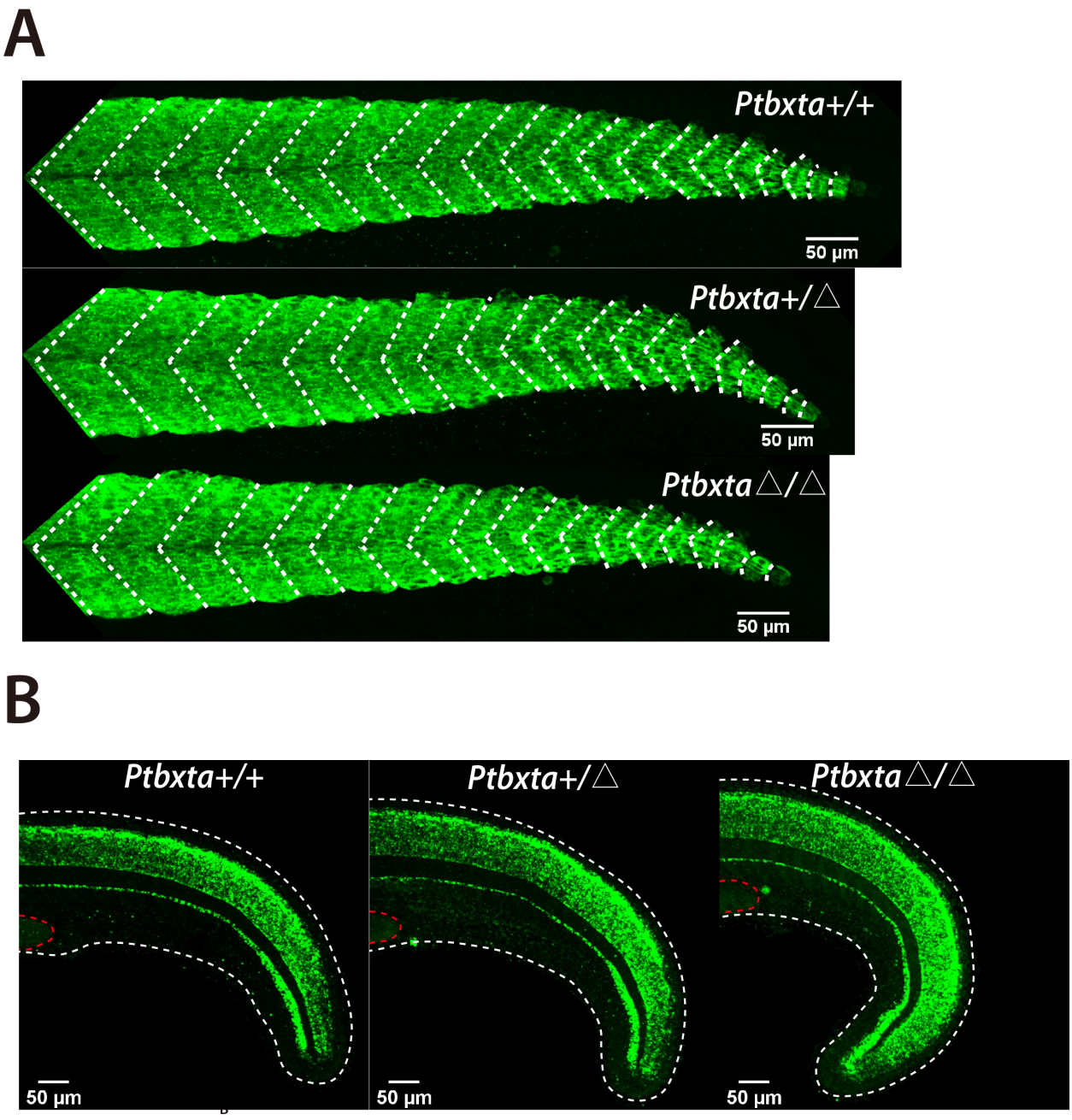


Figure S6 Hox element 1 deletion embryos have reduced posterior somites and an enlarged posterior neural tube. Refers to Figure 6.

A) Homozygous embryos show defects in the formation of posterior somites as shown by staining with the muscle antibody MF20 (for quantification of volume of the posterior somites see Figure 6H). Shown are snapshots of the 3D reconstructions in a lateral view, with dotted lines showing the somite borders. B) Homozygous embryos show an enlarged neural tube as

shown by *sox2* FISH. (for quantification of volume of the posterior neural tube see Figure 6I). Shown are snapshots of the 3D reconstructions in a lateral view, with white dotted lines showing the body border and red dotted lines showing the yolk protrusion.

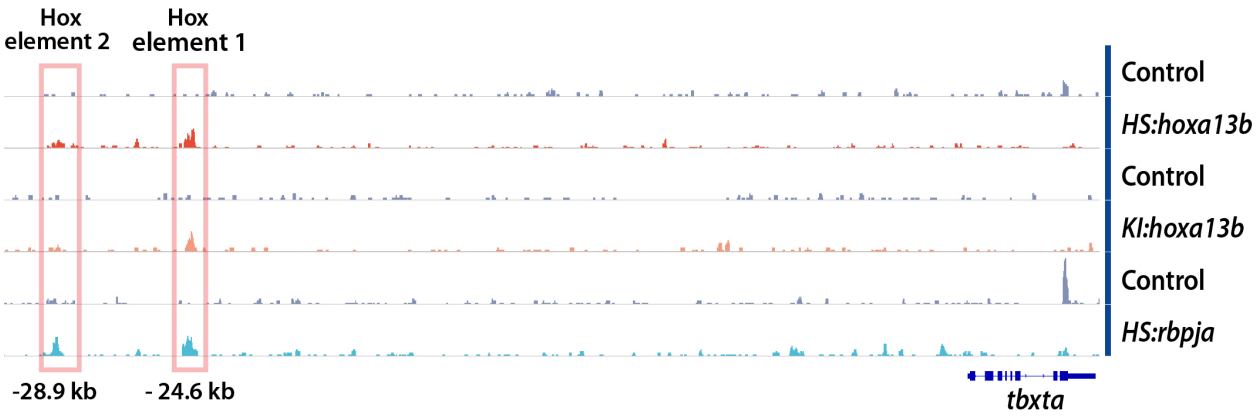


Figure S7 Rbpja binds to Hox element 1 and Hox element 2. Refers to Figure 7.

Rbpja CUT&RUN Peaks were found at Hox element 1 and element 2 sites, which also have the H3K4me1 and H3K27ac histone marks, as shown in Figure 4.

Table S1 H3K4me1 peak annotation.
Peaks were annotated to the nearest gene based on the org.Dr.eg.db annotation database.

Click here to download Table S1

Table S2 H3K27ac peak annotation.
Peaks were annotated to the nearest gene based on the org.Dr.eg.db annotation database.

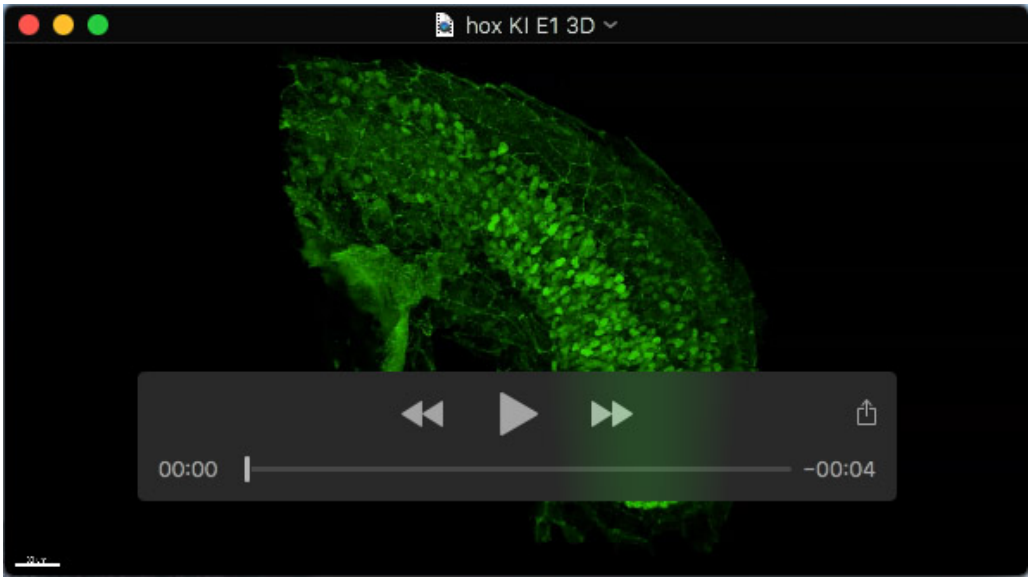
Click here to download Table S2

Table S3 Hoxa13b targets and annotation.
Shown are the CUT&RUN peaks from the *HS:hoxa13b-FLAG-GFP* line, with overlapping peaks from the *KI:hoxa13b-FLAG-GFP* line shown. Relevant Histone modification data from Tables S1 and S2 are included. Peaks were annotated to the nearest gene based on the org.Dr.eg.db annotation database.

Click here to download Table S3

Table S4 Rbpja targets and annotation.
Peaks were annotated to the nearest gene based on the org.Dr.eg.db annotation database.

Click here to download Table S4



Movie 1