Developmental Biology 432 (2017) 151–164

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Hoxa1 targets signaling pathways during neural differentiation of ES cells and mouse embryogenesis

Bony De Kumar[a], Hugo J. Parker[a], Ariel Paulson[a], Mark E. Parrish[a], Julia Zeitlinger[a,b], Robb Krumlauf[a,c,*]

[a] Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[b] Department of Pathology, Kansas University Medical Center, Kansas City, KS 66160, USA
[c] Department of Anatomy and Cell Biology, Kansas University Medical Center, Kansas City, KS 66160, USA

ARTICLE INFO

Keywords:
Hoxa1
Hox genes
Gene regulation
Mouse ES cells
Target genes
Signaling pathways
Mouse embryos

ABSTRACT

Hoxa1 has important functional roles in neural crest specification, hindbrain patterning and heart and ear development, yet the enhancers and genes that are targeted by Hoxa1 are largely unknown. In this study, we performed a comprehensive analysis of Hoxa1 target genes using genome-wide Hoxa1 binding data in mouse ES cells differentiated with retinoic acid (RA) into neural fates in combination with differential gene expression analysis in *Hoxa1* gain- and loss-of-function mouse and zebrafish embryos. Our analyses reveal that Hoxa1-bound regions show epigenetic marks of enhancers, occupancy of Hox cofactors and differential expression of nearby genes, suggesting that these regions are enriched for enhancers. In support of this, 80 of them mapped to regions with known reporter activity in transgenic mouse embryos based on the Vista enhancer database. Two additional enhancers in *Dok5* and *Wls1* were shown to mediate neural expression in developing mouse and zebrafish. Overall, our analysis of the putative target genes indicate that Hoxa1 has input to components of major signaling pathways, including Wnt, TGF-β, Hedgehog and Hippo, and frequently does so by targeting multiple components of a pathway such as secreted inhibitors, ligands, receptors and down-stream components. We also identified genes implicated in heart and ear development, neural crest migration and neuronal patterning and differentiation, which may underlie major *Hoxa1* mutant phenotypes. Finally, we found evidence for a high degree of evolutionary conservation of many binding regions and downstream targets of Hoxa1 between mouse and zebrafish. Our genome-wide analyses in ES cells suggests that we have enriched for in vivo relevant target genes and pathways associated with functional roles of Hoxa1 in mouse development.

1. Introduction

Hox proteins are a highly conserved family of transcription factors that play key roles in the gene regulatory networks involved in specification of anterior-posterior (AP) patterning (Alexander et al., 2009; Carroll, 1995; Mallo et al., 2010; Pearson et al., 2005) and lineage-specific cellular differentiation (Alharbi et al., 2013; De Kumar et al., 2017a; Minoux and Rijli, 2010). Of the clustered *Hox* genes, *Hoxa1* displays the earliest expression during mouse embryogenesis (Hunt et al., 1991; Murphy and Hill, 1991) and is one of the most rapidly induced genes during retinoid-induced differentiation of murine ES cells (De Kumar et al., 2015; Lin et al., 2011). Upstream signals and regulators that control the dynamic expression of *Hoxa1* during mouse embryogenesis have been characterized and retinoids have been shown to play a direct role in the rapid activation of *Hoxa1,* in part by stimulating the release of paused polymerase (Alexander et al., 2009; De Kumar et al., 2015; Dupé et al., 1997; Lin et al., 2011; Parker et al., 2016; Parker and Krumlauf, 2017). *Hoxa1* has important functional roles in mice and humans in neural crest specification, hindbrain patterning and in heart and ear development (Bosley et al., 2008; Gavalas et al., 1998, 2001; Lufkin et al., 1991; Makki and Capecchi, 2010, 2011, 2012; Studer et al., 1998; Tischfield et al., 2005). In clinical studies, *Hoxa1* is implicated in etiology and prognosis of various cancers through altered rates of cell proliferation and metastasis (Bitu et al., 2012; Taminiau et al., 2016; Wardwell-Ozgo et al., 2014; Zha et al., 2012).

A recent study analyzing the genome-wide binding properties of Hoxa1 in differentiated ES cells has shown that bound regions are enriched for clusters of consensus binding motifs for Hoxa1, Pbx and Meis and nearly all Hoxa1-bound regions display co-occupancy with one or more TALE proteins (Pbx, Meis, Prep, TGIF) (De Kumar et al., 2017b). This is consistent with the idea that the TALE family of

* Correspondence to: Stowers Institute for Medical Research, 1000 E. 50th, Kansas City, MO 64110, USA.
E-mail address: rek@Stowers.org (R. Krumlauf).

http://dx.doi.org/10.1016/j.ydbio.2017.09.033
Received 7 August 2017; Received in revised form 27 September 2017; Accepted 28 September 2017
Available online 02 October 2017
0012-1606/

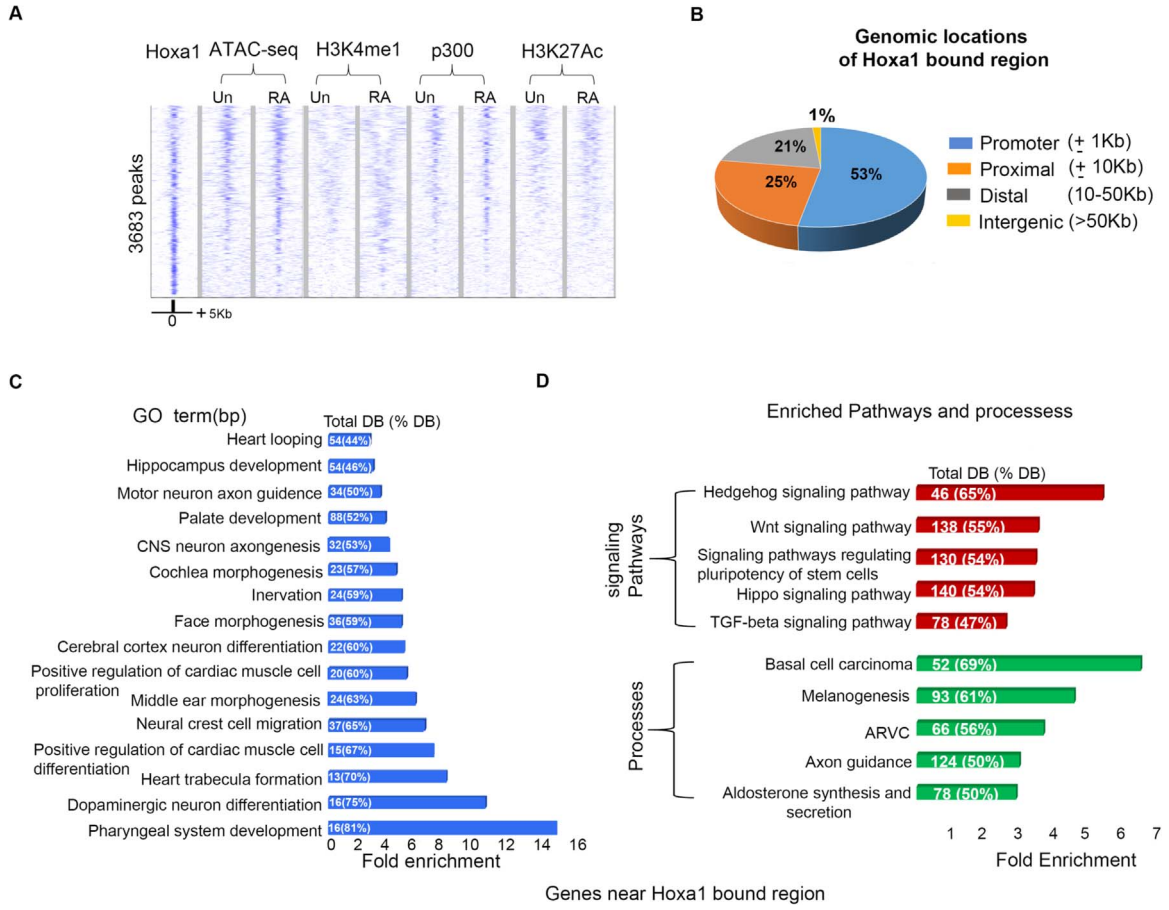


Fig. 1. Identification of Hoxa1-bound regions and associated target genes in differentiated ES cells. (A) Heatmap of genome-wide occupancy of epitope-tagged Hoxa1 identified by ChIP-seq in ES cells following 24 hrs of RA-induced differentiation. The presence of modified histone marks (H3K4me1 and H3K27Ac) characteristic of enhancers, occupancy of the p300 activator protein and accessibility of chromatin (ATAC-Seq) at these genomic loci are also shown for both uninduced (Un) and differentiated (RA) ES cells. Only positions of positive z-score are shown. Peaks are centered on the midpoint of Hoxa1 binding and 5 kb of flanking regions are shown. (B) Classification of Hoxa1-bound regions based on their distance to the nearest TSS and gene. (C) GO term analysis of genes near Hoxa1 bound regions in differentiated ES cells shows terms enriched for neurogenesis, heart and ear development which correlate with phenotypes in *Hoxa1* mutant mice. Top 16 terms with greater than 2-fold change with at least 20 genes associated with these terms with lowest adjusted p-values are shown in the figure. (D) KEGG pathway analyses show enrichment for signaling pathways among all genes near Hoxa1 bound regions. Bp stand for biological processes while DB terms indicates total number of genes associated with each GO term.

proteins are key cofactors for potentiating Hox activity (Merabet and Mann, 2016). Furthermore, there is evidence for extensive auto- and cross-regulatory interactions among the *Hoxa1* and *TALE* genes, indicating that they co-regulate each other by complex feedback loops to maintain appropriate patterns of expression (De Kumar et al., 2017b). For example, Hoxa1 participates in regulating the early expression of its paralogous gene, *Hoxb1* during mouse development. This is achieved through its binding, in partnership with Pbx and Meis, to a Hox-response element upstream of the *Hoxb1* gene, which then triggers an auto-regulatory feedback loop that maintains *Hoxb1* expression in rhombomere (r) 4 of the hindbrain (Pöpperl et al., 1995; Studer et al., 1998; Tvrdik and Capecchi, 2006). This regulatory relationship underlies the genetic synergy and shared functional roles for *Hoxa1* and *Hoxb1*, in patterning the hindbrain, cranial nerves and second pharyngeal arch (Gavalas et al., 1998, 2001; Rossel and Capecchi, 1999).

To explore downstream targets of *Hoxa1,* transcriptome analysis using microarrays has been utilized to compare differential gene expression in micro-dissected samples from prospective r3-r5 of the hindbrain in wild type and *Hoxa1* mutant embryos (Makki and Capecchi, 2011). This identified 137 downregulated and 162 upregulated genes. A complementary gain-of-function approach to identify novel targets was performed through overexpression of *Hoxa1* in r4 (Tvrdik and Capecchi, 2006). GO term analysis of the differentially expressed genes from these studies indicated enrichment of terms related to development of embryonic organs, hindbrain, inner ear, vasculature, hematopoietic and lymphoid organs and cardiac muscle tissues. There was also enrichment of terms related to differentiation of neurons and muscles, cell migration, regulators of apoptosis, retinol metabolism, Wnt and TGF-β signaling pathways (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006). Many of these differentially expressed genes and GO terms are consistent with characterized phenotypes of *Hoxa1* mutants. However, these studies do not distinguish between direct and indirect effects.

Based on analysis of the individual genes, there is evidence that *Cad6* and *EphA2* are examples of direct downstream targets of *Hoxa1*. *Cad6* shows overlapping expression patterns with *Hoxa1,* and *Hoxa1* mutant embryos exhibit rhombomeric and stage specific defects in expression of a *Cad6* reporter (Inoue et al., 1997). *Hoxa1* and *Hoxb1* show overlapping expression with *EphA2* in the primitive streak during gastrulation and, in later stages of development, *EphA2* and *Hoxb1* show restricted expression in r4 (Gale et al., 1996; Murphy and Hill, 1991). Hox-Pbx bipartite sites have been characterized in an *EphA2* r4-enhancer, and this regulatory region responds to trans-activation by Hoxa1/Hoxb1-Pbx heterodimers (Chen and Ruley, 1998). In addition, compound mutants of *Hoxa1* and *Hoxb1* show reduced *EphA2* expression, which provides further evidence that *EphA2* is a downstream target of Hoxa1 and Hoxb1.

In this study, we sought to identify direct downstream targets of Hoxa1 through integration of genome-wide Hoxa1 binding data and

differential gene expression data in differentiated murine ES cells. We then compared the Hoxa1 binding profiles from the ES cells to differential gene expression in *Hoxa1* mutant mouse embryos and in *Hoxa1* gain-of-function zebrafish embryos. Our analyses revealed that many bound regions function as enhancers and that a large number of them appear to regulate multiple components of major signaling pathways. This conclusion is corroborated by reporter assays and evidence for evolutionary conservation between mouse and zebrafish.

2. Results

2.1. Hoxa1 preferentially occupies regions that have properties of enhancers

To investigate genome-wide occupancy of Hoxa1 and identify downstream targets, we utilized KH2 ES cells, carrying an epitope-tagged variant of *Hoxa1*, and differentiated them into neural fates by treatment with retinoic acid (RA) for 24 h (De Kumar et al., 2015). This epitope-tagged version of Hoxa1 is under the tight control of a doxycycline-inducible promoter and we have previously demonstrated that under these conditions the levels of gene expression of the tagged variant were equivalent to endogenous *Hoxa1* (De Kumar et al., 2017b). We then performed chromatin immunoprecipitation with an α-M2 Flag antibody followed by deep sequencing (ChIP-seq) (Fig. 1A and De Kumar et al., 2017b). We identified 3683 Hoxa1 bound regions consistent between biological replicates. After mapping these binding sites to near adjacent genes using Ensembl 80 annotation, 53% are located within 1 kb of the TSS, and the remaining 47% are present in more distal intergenic or intragenic regions (Fig. 1B).

To test whether Hoxa1 bound regions are likely candidates for enhancers, we analyzed histone modifications associated with enhancers (H3K4me1 and H3K27Ac) and occupancy of co-activators (p300) by ChIP-seq. Accessible chromatin was also examined using ATAC-seq (Assay for Transposase-Accessible Chromatin with high throughput sequencing). These experiments were done on both uninduced and differentiated cells to monitor the dynamic changes during differentiation (Fig. 1A).

We found that many future Hoxa1 bound regions are already accessible in uninduced ES cells, have occupancy of p300 and display the presence of H3K4me1 and H3K27Ac that is associated with enhancers. This indicates that some of the future Hoxa1 binding sites are primed or even moderately active in ES cells. At 24 hrs of differentiation, the fraction of Hoxa1 binding peaks associated with at least one of these enhancer marks rises to three quarters and the majority of them have open chromatin and/or occupancy of p300. We also observe newly accessible regions that also acquire the H3K4me1 enhancer mark over time (Fig. 1A), suggesting that they represent latent enhancers (Ostuni and Natoli, 2013; Ostuni et al., 2013). The large fraction of Hoxa1 binding regions with enhancer marks suggests that Hoxa1 is enriched in occupancy in candidate enhancer regions.

2.2. Putative Hoxa1 targets are enriched for components of signaling pathways

To understand the function of the putative target genes, we analyzed which biological processes and pathways are enriched among the genes neighboring the Hoxa1 bound regions (Fig. 1C, D, Table S1). Among the enriched GO terms, many of the biological processes are associated with neural patterning and differentiation, ear development and heart morphogenesis, which directly correlates with previously characterized *Hoxa1*$^{-/-}$ mutant phenotypes in mice and humans (Bosley et al., 2008; Holve et al., 2003; Lufkin et al., 1991; Makki and Capecchi, 2010; Tischfield et al., 2005). For example, the enrichment of the GO term for neural crest cell migration correlates with previous phenotypic analysis showing *Hoxa1* collaborates with *Hoxb1* in modulating the ability of neural crest cells to migrate from r4 (Gavalas et al., 1998, 2001; Makki and Capecchi, 2010, 2012). Furthermore, heart defects observed in humans and mice with mutations in *Hoxa1* are associated with abnormalities in formation, migration and patterning of derivatives of cardiac neural crest cells (e.g. outflow tract). With respect to the GO term for heart looping, there is evidence from gene expression analysis and lineage tracing that *Hoxa1* is involved with other Hox genes in defining regional properties of the secondary heart field (Bertrand et al., 2011; Buckingham et al., 2005) and *HoxA* and *HoxB* genes cooperate in regulating heart looping (Soshnikova et al., 2013). The GO terms for cochlea morphogenesis and middle ear morphogenesis are related to previously shown functions for *Hoxa1* in formation and patterning of components of the external, middle and inner ear in mice and humans (Chisaka et al., 1992; Lufkin et al., 1991; Makki and Capecchi, 2010; Tischfield et al., 2005). One of the most enriched terms, pharyngeal system development, is associated with characterized roles for *Hoxa1* in the second pharyngeal arch (Gavalas et al., 1998; Rossel and Capecchi, 1999). Furthermore, several GO terms associated with neuronal differentiation and axon guidance appear related to functional roles of *Hoxa1* in these processes in mouse and humans (Gavalas et al., 2003; Helmbacher et al., 1998; Tischfield et al., 2005).

With respect to pathways, KEGG analysis revealed an enrichment for components of many signaling pathways (e.g. Wnt, Hippo, Hh and TGF-β) amongst genes near Hoxa1 bound regions (Fig. 1D, Table S1). This is interesting because an examination of the genes falling into many of the different and apparently unrelated GO term categories for biological processes indicate that a unifying theme is enrichment for targets in these signaling pathways. This suggests that *Hoxa1* may have direct regulatory input into modulation of genes in these key signaling pathways that underlies its role in regulating differentiation, morphogenesis and patterning.

In addition to near adjacent genes, we were interested in exploring genes potentially regulated by Hoxa1 bound regions through long range interaction. There is evidence that enhancers can frequently work at longer ranges to impact the expression of more distal genes. To explore this possibility, we generated Chromosome Conformational Capture (HiC) data, which maps long-range physical interactions between regions (Lieberman-Aiden et al., 2009), and integrated it with our Hoxa1 ChIP-seq data. (Fig. S1A). Interestingly, many Hoxa1 bound regions show long range interaction with distal genes. Of the bound regions, more than 10 kb away from nearest adjacent gene,10% interact only with the near adjacent gene, 69% make multiple contacts and interact with both nearest neighbor and distal genes while 15% only interacts with more distal genes (Fig. S1A, Table S2). The long-range interactions of Hoxa1 bound regions show a distribution with a medial distance around 150 kb and majority of this group of bound regions show connections with 1–2 other regions (Figs. S1B, 1C). These results are consistent with the emerging body of work from chromatin confirmation studies on promoter-enhancer interactions that indicate in general putative enhancers frequently make a majority of their contact with distal genes as opposed to only near adjacent genes (reviewed in Yao et al., 2015). It is interesting that analysis of enriched pathways that includes these distal genes identified from HiC analysis also shows over-representation of pathways for Wnt, TGF-β and focal adhesion (Fig. S1D, Table S3). Many of these connections may be preexisting and result from binding of various chromatin bound proteins (e.g. CTCF, Cohesin etc.). It will be to interesting to determine whether Hoxa1 binding contributes to the regulatory potential of these connections. Nonetheless it is intriguing that the genes associated with the connections bound by Hoxa1 show an enrichment for signaling pathways. This adds support to the idea that Hoxa1 bound regions may function as enhancers targeting both near adjacent and distal genes involved in signaling pathways.

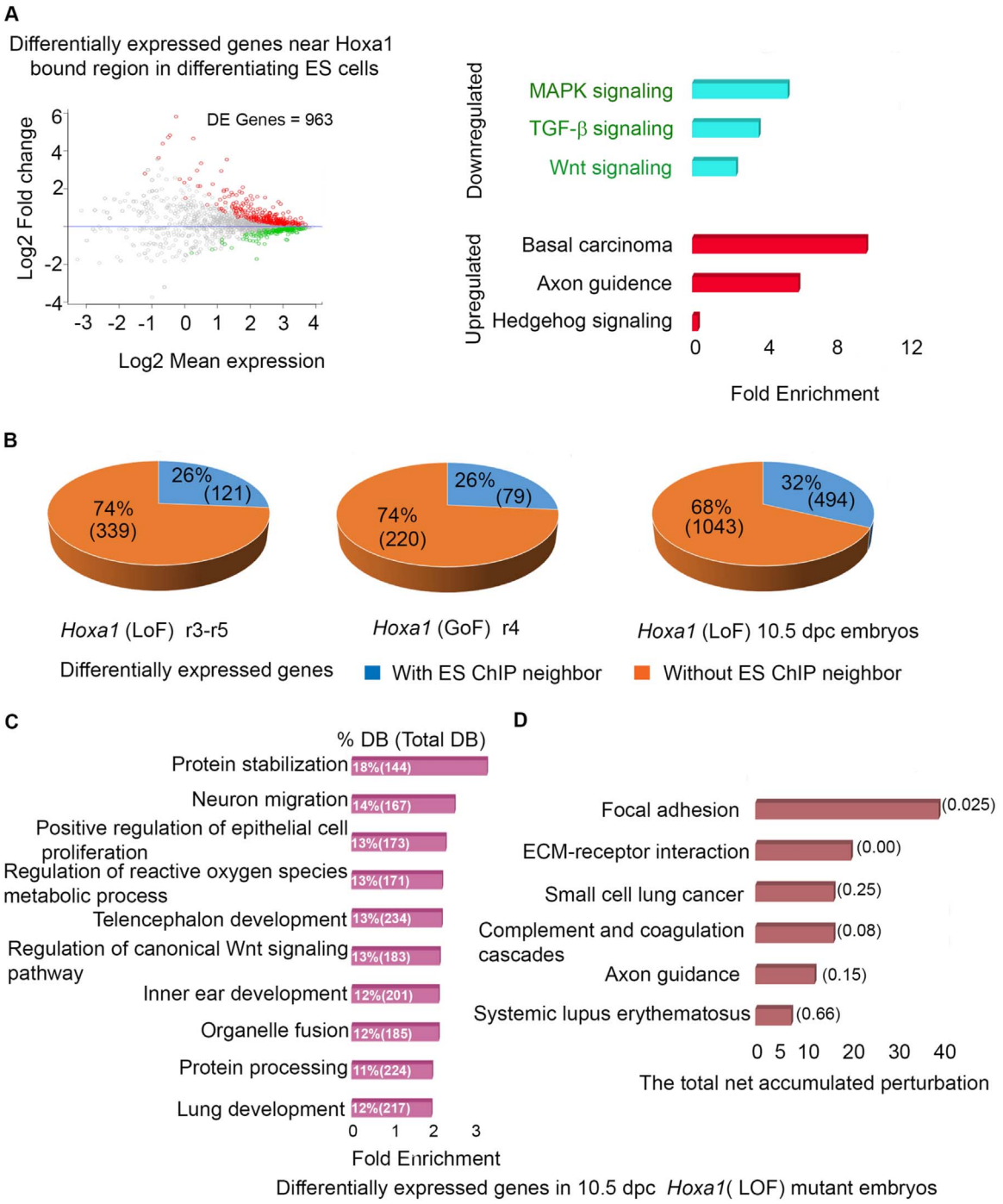


Fig. 2. Integration of Hoxa1 occupancy and differential gene expression. (A) MA plot showing differentially expressed genes in differentiated versus untreated ES cells (as red and green dots) near Hoxa1 bound regions. KEGG pathway analysis showing enrichment for signaling pathways among differentially expressed genes is also shown on right panel. (B) Comparison of genes near Hoxa1-bound regions in differentiated ES cells with differentially expressed genes in the r3-r5 region of *Hoxa1* LoF (Loss-of–Function) mutant embryos (Makki and Capecchi, 2011), in r4 of *Hoxa1* GoF(Gain-of-Function) embryos (Tvrdik and Capecchi, 2006) and in whole10.5 dpc LoF mutant embryos. In these cases, 26–32% of the differentially expressed genes have a nearby occupancy of Hoxa1 suggesting in vivo relevance of targets identified in differentiated ES cells. (C & D) GO term and KEGG pathway analysis of differentially expressed genes from 10.5 dpc whole mouse embryos. Top 10 specific terms with greater than 2-fold change with at least 10 genes associated with these terms and least adjusted p-values are shown in the figure. Bp stand for biological processes while DB terms indicates total number of genes associated with each GO term. DE means differentially expressed genes. For D, The P_{NDE} value is shown in bracket which represents the probability of getting set of DE genes on the given pathway.

2.3. Many putative Hoxa1 targets are regulated in ES cells and mouse embryos

To obtain evidence that putative targets near Hoxa1 bound regions may be regulated by Hoxa1, we compared the patterns of differential gene expression between ES cells and those differentiated with RA (De Kumar et al., 2015), with genes near Hoxa1 bound regions Of the total of 3683 regions bound by Hoxa1, 963 (26.1%) display differential expression after 24 hrs of RA treatment compared to uninduced ES cells (Fig. 2A). These represent candidates for direct target genes of Hoxa1. Up-regulated genes (red) are significantly enriched for pathways related to axon guidance and Hedgehog signaling, while down-regulated genes (green) are enriched for MAPK, TGF-β and Wnt signaling pathway components (Fig. 2A). A similar fraction of genes

bound directly by Hoxa1 in our ES cell analysis was found among previously published microarray profiling of micro-dissected hindbrain segments from *Hoxa1* mutant mice (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006). Genes from loss-of function (LoF) mutants show 26% (121 of 460) overlap with Hoxa1 bound regions, and those from gain-of-function (GoF) 26% (79 of 299) overlap (Fig. 2B). To further identify putative Hoxa1 targets in mouse embryos, we performed transcriptional profiling (RNA-seq) of whole *Hoxa1* mutant and wild type embryos at 10.5 dpc and identified 1537 differentially expressed genes (Fig. 2B, Table S4). Among these genes, 32% (494) have near adjacent Hoxa1 binding. Comparing this whole embryo *Hoxa1* mutant analysis with the LoF study on micro-dissected r3-r5 (Makki and Capecchi, 2011), we find that 10% (25) of the genes overlap and of these 40% have near adjacent Hoxa1 bound regions.

GO term analysis of all differentially expressed genes from the 10.5 dpc embryos shows enrichment of pathways and biological processes relevant to some established sites, tissues and processes of Hoxa1 function (Fig. 2C, D, Table S5). For example, the GO term for inner ear development is related to the function for Hoxa1 in formation and patterning of components of the external, middle and inner ear (Chisaka et al., 1992; Lufkin et al., 1991; Makki and Capecchi, 2010; Tischfield et al., 2005). The terms for neuron migration and axon guidance appear related to functional roles of Hoxa1 in these processes in mouse and humans (Bosley et al., 2008; Gavalas et al., 2003; Helmbacher et al., 1998; Holve et al., 2003; Tischfield et al., 2005). There is also an intriguing link between the enrichment for the term small cell lung cancer and *Hoxa1*. Cancer studies have indicated that Hoxa1 acts on TGF-β pathways to increase metastasis and cell invasion (Wardwell-Ozgo et al., 2014). These GO terms, based on Hoxa1 mutant embryos, correlate very well with those observed in Fig. 1C, D based solely on near adjacent genes to Hoxa1 bound regions. It is worth noting that differentially expressed genes near Hoxa1 bound regions in all of these different analyses were both up- and down-regulated, consistent with the idea that Hoxa1 is capable of exerting both positive and negative regulatory inputs into target genes. These different in vivo comparisons were made using Hoxa1 binding data from RA-treated ES cells, but they illustrate that it is likely that the ES cell system is suitable for the identification of a variety of in vivo relevant direct downstream targets of Hoxa1 in mouse embryos.

2.4. *Hoxa1 bound regions function as enhancers*

To obtain direct functional evidence that Hoxa1 bound regions may be associated with regions that function as enhancers, we extracted the genomic coordinates of fragments that were validated as enhancers based on reporter assays in mouse embryos from the VISTA Enhancer Browser database (http://enhancer.lbl.gov/) and compared them with occupancy of Hoxa1. This analysis reveals that 105 of the 2870 previously tested mouse regions in the database have Hoxa1 occupancy. Table S6 gives a complete list of genomic coordinates for the 105 Hoxa1 bound regions along with links to access their respective reporter expression data in mouse embryos. Notably, 80 of these regions mediate reporter gene activity in mouse embryos, including many with expression in a variety of brain and cranio-facial structures in developing embryos, consistent with the characterized tissues in which *Hoxa1* is active (Fig. 3). It is worth noting that many of these 80 enhancers also mediate reporter activity in regions where *Hoxa1* is not expressed (e.g. forebrain and midbrain). This indicates that these enhancers are likely to have a variety of other *cis*-elements that integrate complex regulatory inputs independent of or in combination with Hoxa1. Hence, Hoxa1 is likely to reflect only a portion of the total regulatory input and it may have both positive and negative regulatory inputs to modulate enhancer activity. These regions are therefore likely to represent some enhancers with direct input from Hoxa1.

We independently tested 17 Hoxa1 bound regions for enhancer activity using transgenic reporter assays in zebrafish and mice (Fig. 4 and De Kumar et al., 2017b). Two examples are shown in Fig. 4. One region is in the 5th intron of *Dok5,* while the other resides ~50 kb upstream of *Wls1*. *Dok5* is an example of a target gene that encodes a membrane protein which serves as an adapter involved in signal transduction. Dok5 interacts in the membrane with phosphorylated receptor tyrosine kinases to stimulate the MAP kinase pathway and modulate the outgrowth of neurons (Grimm et al., 2001; Pan et al., 2013; Wen et al., 2009). *Wls1* is an example of a potential target gene near components of the Wnt signaling pathway. *Wls1* encodes for a protein that controls the trafficking, secretion and subcellular location of Wnt proteins and has been shown to have a function in AP patterning during development (Fu et al., 2009; Yu et al., 2010; Zhang et al., 2016). In both cases, the peaks of Hoxa1 binding overlap with occupancy of p300, assessable chromatin (ATAC-seq) and have characteristic flanking enhancer marks of H3K27Ac and H3K4me1. To enhance the identification of Hoxa1-bound regions at single base pair resolution, we also performed ChIP-nexus experiments (He et al., 2015). This analysis revealed that single peaks identified by ChIP-seq often represent clusters of multiple binding events. For example, three distinct binding regions were identified in the *Dok5* region (Fig. 4A) and two separate sites were found in the *Wls1* region (Fig. 4B). Since some of these motifs resemble consensus binding sites for Pbx and Meis, which can serve as cofactors for Hox proteins (Merabet and Mann, 2016), we confirmed the occupancy of Pbx and Meis by ChIP-seq experiments.

When we tested the regulatory potential of these two Hoxa1 bound regions, we found that reporter constructs carrying the core binding regions along with flanking areas (~800 bp) mediated restricted reporter expression in the developing hindbrain of developing mouse and zebrafish embryos (Fig. 4C, D). The *Dok5* enhancer directs strong expression in r4 of mouse embryos but there is also reporter expression detected in neural crest adjacent to the r4 and r6 region, somites and the posterior growth zone in the tailbud (Fig. 4C). In zebrafish, mosaic expression is detected in F_0 embryos in the r2-r5 region. The region from *Wls1* mediates reporter expression in both mouse and zebrafish embryos predominantly in r4 and the mid/hindbrain junction (Fig. 4D). In zebrafish, there is also some ectopic reporter staining in the forebrain region, often observed with the HLC vector (De Kumar et al., 2017b; Parker et al., 2014). Beyond these two regions, in a recent study, we found that some of the Hoxa1 bound regions mapped in the vicinity of genes encoding its TALE cofactors (De Kumar et al., 2017b). Furthermore, we demonstrated that 11 out of 13 of these regions near *TALE* genes functioned as enhancers and more detailed analysis of these enhancers from *Meis2* and *Meis3* indicate the respond to ectopic Hoxa1 expression and the Hoxa1 binding sites are required for regulatory activity (De Kumar et al., 2017b).

Together our analyses showing epigenetic marks of enhancers, assessable chromatin, occupancy of cofactors, differential expression of nearby genes and regulatory potential indicate that Hoxa1-bound regions in differentiated ES cells are enriched for enhancer activities. Hence, many of these nearby genes are potential direct targets of Hoxa1 and may receive regulatory input from *Hoxa1* in mouse development.

2.5. *Hoxa1 frequently targets multiple signaling pathway components*

Targets of Hoxa1 identified from our ChIP-seq and differential gene expression studies indicate an enrichment for major signaling pathways (Figs. 1 and 2). As noted earlier, genes falling into many of the different and apparently unrelated GO term categories for biological processes are related through enrichment for targets shared signaling pathways. A hallmark of this analysis is that we found multiple inputs of Hoxa1 into genes involved in these pathways, as there are binding sites near genes encoding secreted inhibitors, ligands, receptors and

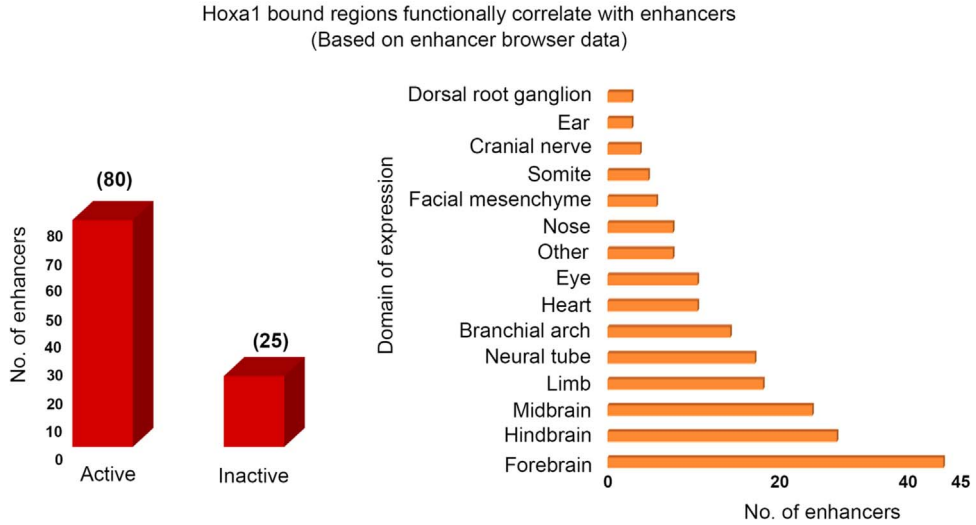


Fig. 3. Analysis of Hoxa1-bound regions that mapped to Vista enhancers. 105 of the Hoxa1-bound regions were previously analyzed for enhancer activity as described in the Vista enhancer database. Of these, 80 display restricted expression in transgenic mouse embryos (left side). A large fraction is expressed in area overlapping with Hoxa1 activity (right side). Supplementary Table 1 gives a complete list of genomic coordinates for the 105 Hoxa1 bound regions along with links to access their respective reporter expression data in mouse embryos.

down-stream effectors among the *Wnt*, *TGF-β* and *Hedgehog* pathways (Figs. 5 and 6). For example, Hoxa1 binds to genes associated with diverse components of canonical and non-canonical Wnt pathways (Fig. 5A). This includes the Wnt modulator *Wls1*, which we validated above (Fig. 4B and D), as well as other Wnt pathway members (*Axin2*, *Dkk* and *Wnt11*), which display the typical occupancy of Hox cofactors (Pbx and Meis) and other enhancer-related properties (Fig. 5B).

Similarly, there is evidence for multiple inputs into the TGF-β signaling pathways by *Bmp7, Chordin* and *Smad5* (Fig. 6A) and the Hedgehog signaling pathway by *Ptch1* and *Su(fu)* (Fig. 6B). Direct input into the TGF-β pathway is interesting since clinical studies have correlated *HOXA1* with enhanced tumor growth and metastasis in association with modulation of the TGF-β pathway (Wardwell-Ozgo et al., 2014). Our data suggests this may be a direct input by HOXA1. This underscores the biological relevance of the downstream targets and signaling pathways uncovered by the analysis in differentiated ES cells.

2.6. Gain-of-function analysis reveals conservation of Hoxa1 targets between zebrafish and mouse

To explore the evolutionary conservation of Hoxa1 targets between zebrafish and mouse, we independently identified Hoxa1-dependent target genes in zebrafish embryos using a *Hoxa1* gain-of-function assay. We utilized a zebrafish line carrying an *mCherry* reporter integrated near the *egr2b (krox20)* locus, which mediates expression in r3 and r5 (Distel et al., 2009), and crossed it with a line in which a mouse *Hoxb1* enhancer drives eGFP in r4 in a *Hoxb1*- and *Hoxa1*-dependent manner (Parker et al., 2014; Pöpperl et al., 1995; Zhang et al., 1994). Upon ectopic expression of *Hoxa1,* by injection of either wild-type or epitope-tagged mouse *Hoxa1* mRNA, the mCherry expression in r3 is lost and the r4-restricted expression of GFP is expanded into anterior regions (Fig. 7A). These alterations are consistent with the established ability of ectopic *Hoxa1* to induce transformations of anterior neural tissue to an r4-like state (Alexandre et al., 1996; Zhang et al., 1994). We then used transcriptional profiling (RNA-seq) of the *Hoxa1*-injected embryos to identify differentially expressed zebrafish genes (Fig. 7B, C, Table S7). The results from wild-type and epitope-tagged *Hoxa1* injections showed very similar results (correlation co-efficient = 0.96), providing further evidence for the functional equivalency of these two RNAs (Fig. 7 D).

In total, we identified 409 differentially expressed genes in zebrafish embryos upon *Hoxa1* injection (Table S5), 294 of which have identifiable mouse homologs. Among these mouse genes, 121 (41%) have nearby Hoxa1 binding sites in RA-treated ES cells (Fig. 7E). They include some of the known targets of Hoxa1 such as *Meis3, Wnt8a* and *Hoxb1a* (Fig. 7F). To address whether evolutionarily conserved enhancer sequences may be responsible for the Hoxa1 targets common between mouse and zebrafish, we performed phylogenetic sequence analysis on the Hoxa1-bound regions near this set of 121 genes. This revealed extensive sequence conservation as compared to random sequences, with PhylolP values indicating that Hoxa1-bound regions are under positive selection (Fig. S2). Furthermore, 18 of these showed extensive conservation and enhancer activity when tested in the VISTA Enhancer Browser database (Fig. 3). This database utilized a high degree of evolutionary conservation of sequences as a basis for selecting many of the regions to test for enhancer activity. To further illustrate the degree of conservation, we also examined several (> 10) of these Hoxa1 bound regions selected at random using Vista alignment plots to compare human, mouse, chicken and zebrafish sequences. Conservation plots showing alignments along with the position of Hoxa1 binding and assessable regions (ATAC-seq) are shown for *Auts2, BC0502040, Dnmt1* and *Elp4* (Fig. 8). In these four examples, the conserved regions contain multiple binding site motifs for Hox proteins and their TALE cofactors. Together, these diverse comparisons indicate that there is a high degree of evolutionary conservation of many binding regions and downstream targets of Hoxa1 between mammals and bony fish.

3. Discussion

In this study, we have identified and characterized regions bound by Hoxa1 on a genome-wide basis in differentiating mouse ES cells, utilizing a series of genomic approaches. These bound regions display chromatin signatures characteristic of enhancers and are located near genes with dynamic expression during neuroectodermal differentiation of ES cells. Many of the Hoxa1 bound regions (80) map to enhancers identified in the VISTA Enhancer Browser database. Transgenic reporter assays of other selected regions in zebrafish and mouse revealed that they have conserved regulatory activity and function as enhancers to mediate neural expression in developing embryos. Furthermore, many of these bound regions in differentiating ES cells are adjacent to genes whose expression has been shown to be altered in vivo in *Hoxa1* loss-of-function mouse mutants and in *Hoxa1* gain-of-

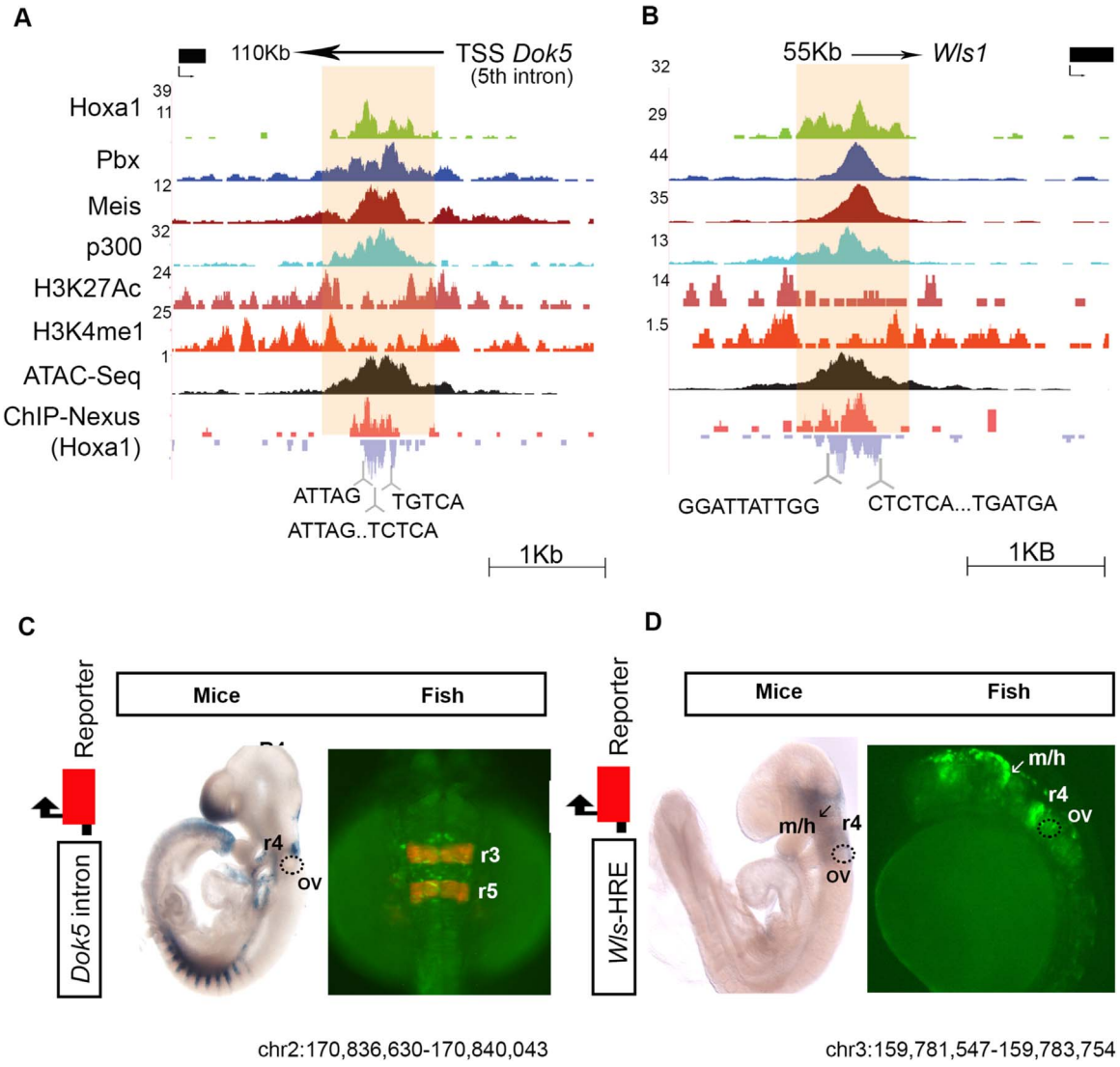


Fig. 4. Hoxa1-bound regions function as enhancers. (A, B) Browser shots of two Hoxa1 bound regions (A, *Dok5*) and (B, *Wls1*) along with occupancy of TALE cofactors (Pbx and Meis), co-activator p300, modified histone marks characteristic of enhancers (H3K4me1 and H3K27Ac) and open chromatin states (ATAC-seq) are shown. ChIP-nexus data reveals the presence of multiple binding motifs (bottom) for Hoxa1 and its cofactors in the protected regions. (C, D) Regulatory analysis of the (A, *Dok5*) and (B, *Wls1*) Hoxa1-bound regions along with 250 bp flanking sequences in mouse and zebrafish embryos using *LacZ* and *GFP* transient transgenic reporter assays, respectively. In C, zebrafish embryos with mCherry inserted in the endogenous *egr2b* locus (*krox20*), mark r3 and r5 with reporter expression (Distel et al., 2009). The bound region from each gene mediates neural-specific expression in mouse and zebrafish embryos, indicating conservation of activity. OV, otic vesicle, m/h, mid/hindbrain boundary and r, rhombomere.

function assays in mouse and zebrafish (Makki and Capecchi, 2011; Tvrdik and Capecchi, 2006). The Hoxa1 bound regions make contact with near adjacent and distal genes that are enriched for components of major signaling pathways (e.g. Wnt, TGF-β, Hedgehog and Hippo) and are associated with biological processes, such as heart and ear development, neural crest migration and neuronal patterning and differentiation that underlies major *Hoxa1* mutant phenotypes (Lufkin et al., 1991; Makki and Capecchi, 2010). Together, our findings imply that through these experiments in ES cells we have identified on a genome-wide basis and enriched set of in vivo relevant Hoxa1 targets and this allows us to further explore the nature of these regions.

In addition to roles in fundamental developmental processes, many putative targets of Hoxa1 are associated with cancers (e.g. basal cell carcinoma and small cell lung cancer). This is interesting because studies on various human cancers (squamous cell, breast, melanoma and hepatocellular carcinoma) have implicated roles for *HOXA1* in the etiology and prognosis of these tumors (Bitu et al., 2012; Taminiau et al., 2016; Wardwell-Ozgo et al., 2014; Zha et al., 2012). High levels of *HOXA1* correlate with poor patient outcomes and shorter times to metastatic events. In a melanoma cell culture model, expression of *HOXA1* drives increased metastasis and cell invasion through activation of the TGF-β pathway and represses genes involved in melanocyte differentiation (Wardwell-Ozgo et al., 2014). Furthermore, the role of *HOXA1* in promoting a pro-invasion phenotype requires the TGF-β pathway. Relevant to these findings, our analysis reveals that melanogenesis is one of the most highly enriched biological processes, whereby 61% of 93 genes associated with this pathway have a nearby Hoxa1 bound region (Fig. 1D). Similarly, the TGF-β signaling pathway is also highly enriched, whereby 47% of 78 genes associated with this pathway have a Hoxa1 binding peak (Figs. 1D and 6A). In conjunction with the clinical data, our results suggest that HOXA1 directly regulates key genes that control the process of melanocyte differentiation and directly modulates TGF-β signaling during melanoma tumorigenesis.

Direct inputs into signaling pathways appear to be an important feature of Hox proteins. Genome-wide binding studies revealed that canonical *Wnt* signaling is a target of *Hoxa2* in cranial neural crest cells, consistent with the role for *Hoxa2* in patterning neural crest derived structures in the head (Donaldson et al., 2012). They also showed that expression *fzd4* and *Wnt-β-catenin* is lost in *Hoxa2* mutant mice. The prominence of altered signaling pathways in

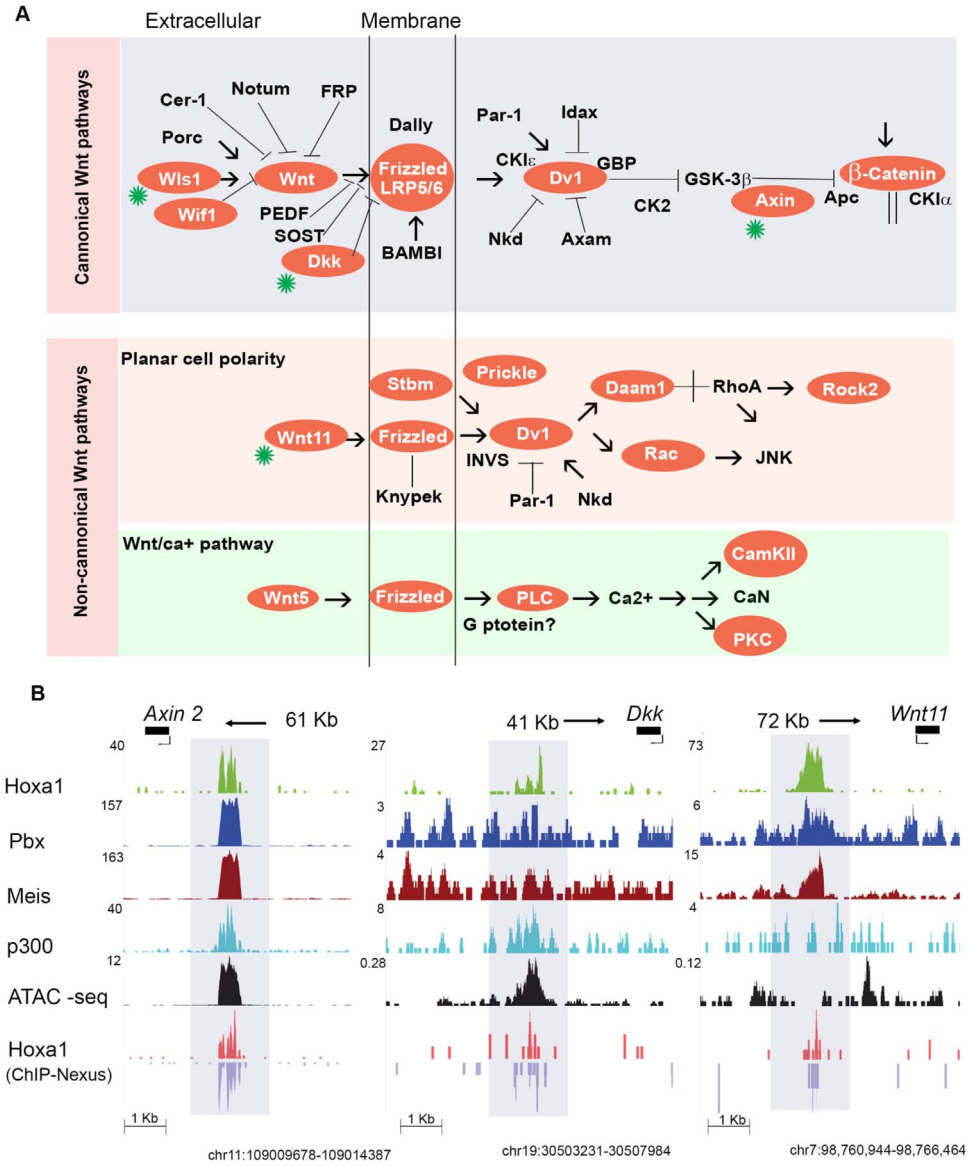


Fig. 5. KEGG analysis indicates enrichment in Hoxa1 bound regions in differentiated ES cells for components of the Wnt signaling pathway. (A) Diagram of components of Wnt signaling pathways, with genes containing nearby occupancy of Hoxa1 shaded in orange. Both the canonical and the non-canonical pathways are shown. The green star indicates the specific examples shown in (B) along with Wls1 shown in Fig. 4. (B) UCSC genome browser shots for the *Axin2, Dkk* and *Wnt11* loci showing binding of Hoxa1 (ChIP-seq and ChIP-nexus) and occupancy of the Pbx and Meis cofactors. The occupancy of the p300 co-activator and accessibility of chromatin (ATAC-Seq) at these genomic loci are also shown. Genomic coordinates (mm10) are indicated below each locus.

differential gene expression analyses of many Hox mutants (e.g. Hoxd10, Hoxc6, Hoxc8) (Hedlund et al., 2004; Lei et al., 2006, 2005; McCabe et al., 2008) adds indirect support for the idea that control of signaling pathways is likely to reflect a direct regulatory input of Hox proteins associated with their functional roles in patterning and differentiation.

Beyond inputs into signaling cascades, we recently found that *Hoxa1* and *TALE* genes are involved in extensive cross regulatory interactions and that altering the expression level of *Hoxa1* changes the levels of *TALE* genes (De Kumar et al., 2017b). This is significant because TALE proteins are important cofactors for a variety of Hox proteins and other transcription factors (reviewed in Merabet and Mann, 2016). In a genome-wide study of Hoxa2 binding peaks, Hox and Hox-Pbx motifs are highly enriched (Donaldson et al., 2012). Hence, Hoxa1 may have a role in regulating the level and cohort of TALE proteins available to partner with a wide variety of Hox proteins and other transcription factors, impacting their binding abilities or specificity.

Aspects of Hox gene regulatory networks in the neurectoderm, including key roles in hindbrain segmental patterning, have been shown to be conserved across diverse vertebrate species (reviewed in Alexander et al., 2009; Krumlauf, 2016; Parker et al., 2016; Parker and Krumlauf, 2017). Comparative genomics approaches in vertebrates have revealed thousands of highly conserved enhancer elements that are shared between jawed vertebrates and are associated with genes that have roles in transcription and development (Pennacchio et al., 2006; Woolfe et al., 2005). Hox-Pbx transcription factor binding site motifs were found to be enriched within these conserved sequences, leading to the proposal that these sites may be crucial Hox inputs into gene regulatory networks that are shared across jawed vertebrates (Grice et al., 2015; Parker et al., 2011). Our mouse-zebrafish comparative data further supports this model, providing evidence for the binding of Hoxa1 to such conserved elements. We have found that many of the shared Hoxa1 target genes between mouse and zebrafish have nearby Hoxa1 bound peaks in mouse cells. These Hoxa1 bound sequences show significant phylogenetic sequence conservation and

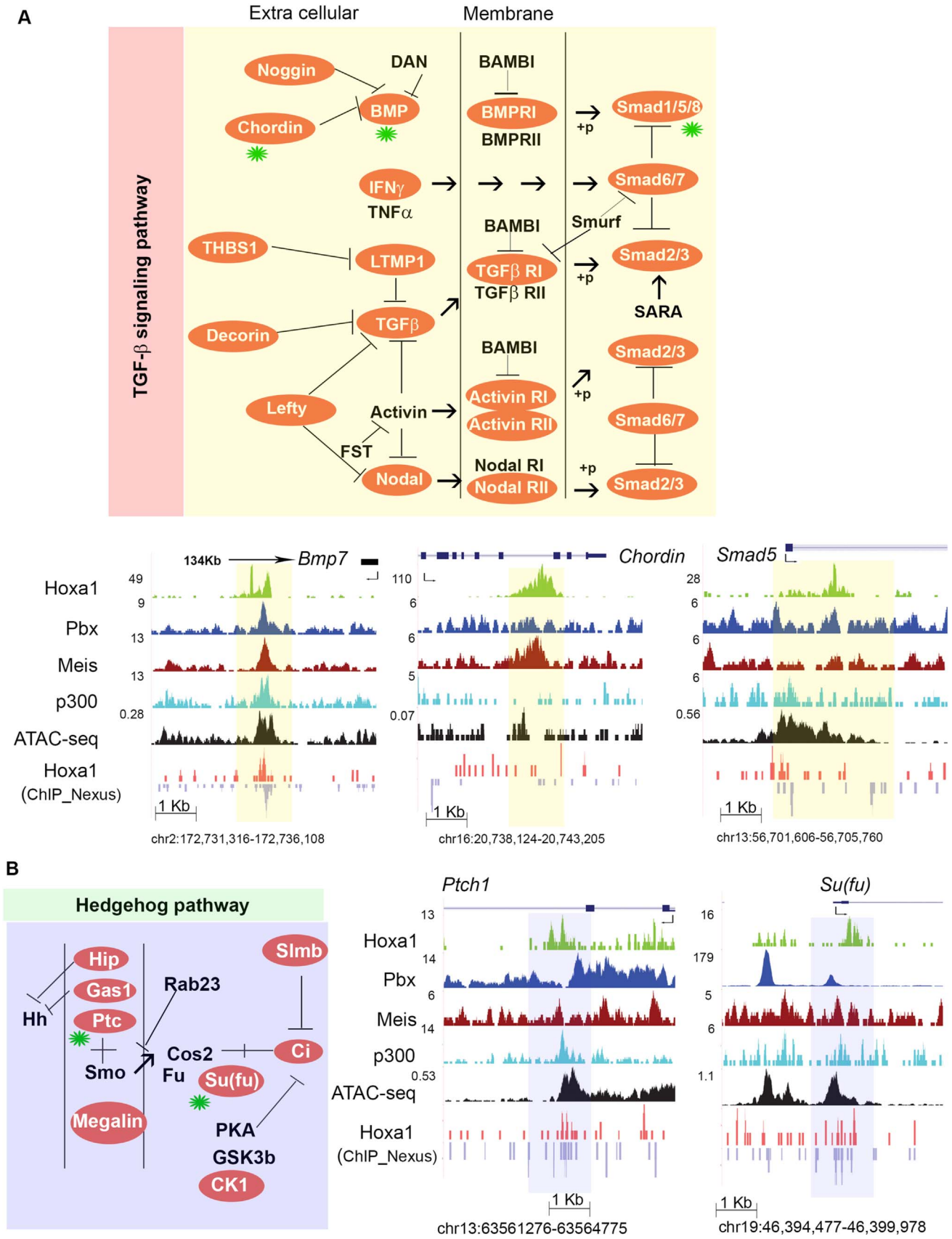
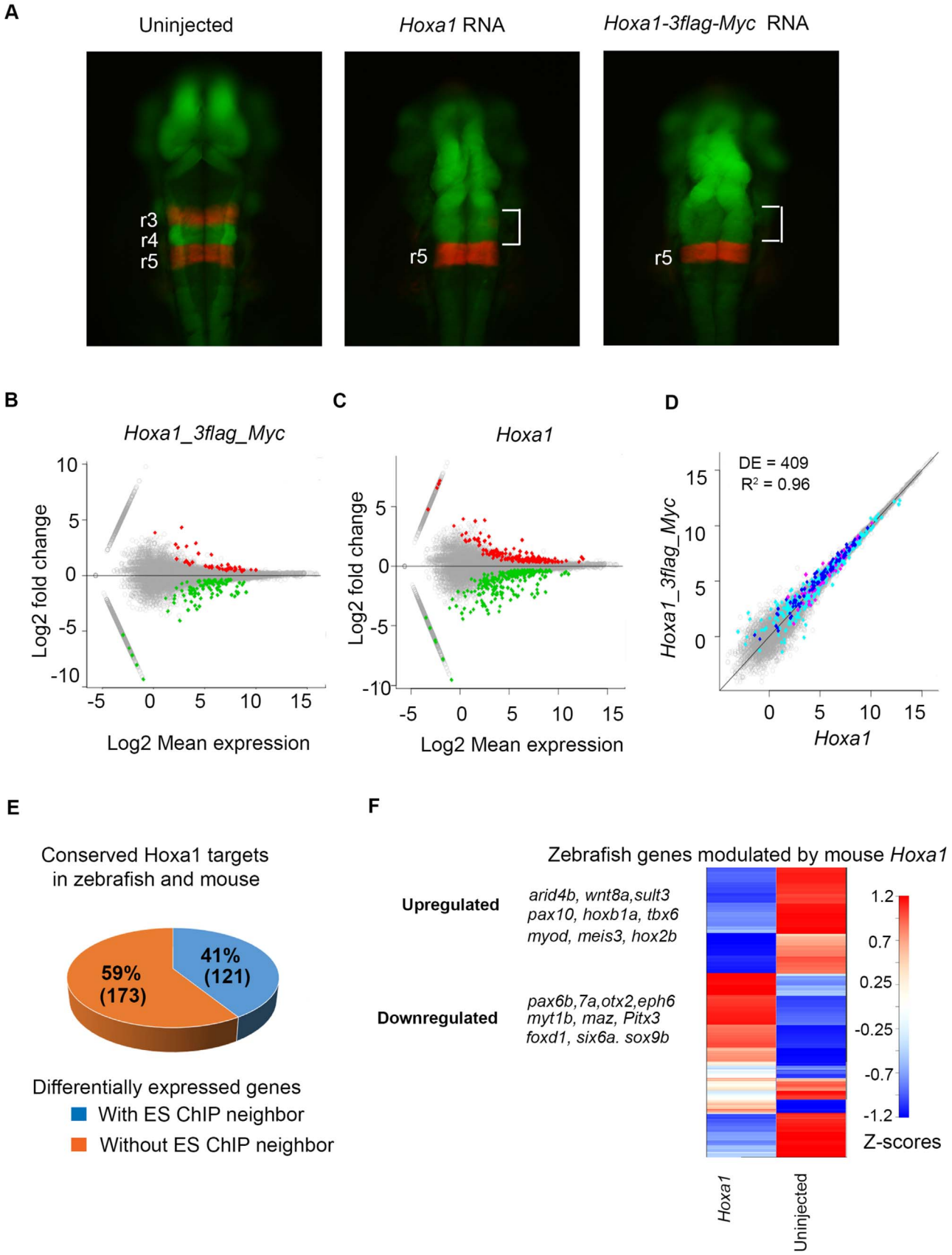


Fig. 6. KEGG analysis indicates enrichment in Hoxa1-bound regions in differentiating ES cells for components of the TGF-β and Hedgehog signaling pathways. (A) Diagram of components of TGF-β with genes containing nearby occupancy of Hoxa1 shaded in orange. The green star at *Bmp7, Chordin* and *Smad5* indicates that these loci are shown in more detail as UCSC genome browser shots at the bottom. (B) Diagram of components of Hedgehog signaling pathways, with genes containing nearby occupancy of Hoxa1 shaded in orange. The green star at *Ptch1* and *Su (fu)* indicated that these loci are shown as UCSC genome browser shots at the bottom. The binding of Hoxa1 (ChIP-seq and ChIP-nexus), occupancy of the Pbx and Meis cofactors, occupancy of the p300 co-activator and accessibility of chromatin (ATAC-Seq) at these genomic loci are also shown. Genomic coordinates (mm10) are indicated below each locus.



harbor transcription factor binding site motifs for Hox and TALE factors, suggesting that they represent ancient Hoxa1-TALE-responsive *cis*-regulatory elements that have been conserved between mammals and fish. Together this data provides a resource for elucidating ancient downstream targets of Hox patterning that have been conserved across vertebrates.

Fig. 7. Evolutionary conservation of Hoxa1 targets. (A) Ectopic expression of Hoxa1 in zebrafish. In control embryos, Red, mCherry reporter expression in rhombomeres 3 and 5 (r3 & r5) and green, GFP reporter expression under control of a Hoxb1-ARE enhancer mediates spatially-restricted expression in r4 and the rostral brain. Injection with Hoxa1 or epitope-tagged Hoxa1 RNAs repressed mCherry-specific expression in r3 and anteriorizes r4 expression of GFP (brackets). (B, C) MA plots showing differentially expressed (DE) genes (as red and green dots) in zebrafish after injection with mouse *Hoxa1_3flag_myc* (B) or mouse *Hoxa1* (C) mRNA. (D) A scatterplot of gene expression values (log2-fold-change) from the two assays show minimal differences (R = 0.96). DE genes are shown in color. (E) Pi chart showing fraction of mouse orthologous genes modulated by *Hoxa1* in zebrafish. Nearly half of the orthologous genes have a nearby Hoxa1 binding site, suggesting evolutionary conservation of Hoxa1 targets between mouse and zebrafish. (F) Transcriptional profiling of zebrafish embryos comparing control and mouse *Hoxa1* injected embryos (20 hpf). Differential gene expression at FDR ≤ 0.05 are shown as a heatmap with values capped at ± 3 for visibility. At the left are examples of up and downregulated genes.

In summary, our findings help to elucidate target genes and pathways that underlie the functional roles for Hoxa1 in patterning and differentiation during development and disease.

4. Methods

4.1. Generation of epitope tagged HOXA1 cell line and RA induction

KH2 ES cells contain an engineered insertion site in the *Col1A1* locus under control of a doxycycline inducible promoter (Beard et al., 2006). We previously utilized this line to insert a cDNA encoding a Hoxa1-triple flag-Myc (De Kumar et al., 2017b). Clones were confirmed by PCR genotyping and karyotyping was done using FACS calibur by analysis of DNA content. The induction kinetics of the *Hoxa1-triple Flag-Myc* transgene were titrated such that the levels of expression of the transgene were equivalent to endogenous *Hoxa1* expression (De Kumar et al., 2017b). These levels were validated by three independent methods including western hybrizdization after RA and Dox treatment for 6–48 h. The length of RA treatment (24 hrs), was selected based on previous analyses of a timecourse of the KH2 ES cell differentiation program to optimize a neural ectoderm like fates (De Kumar et al., 2015).

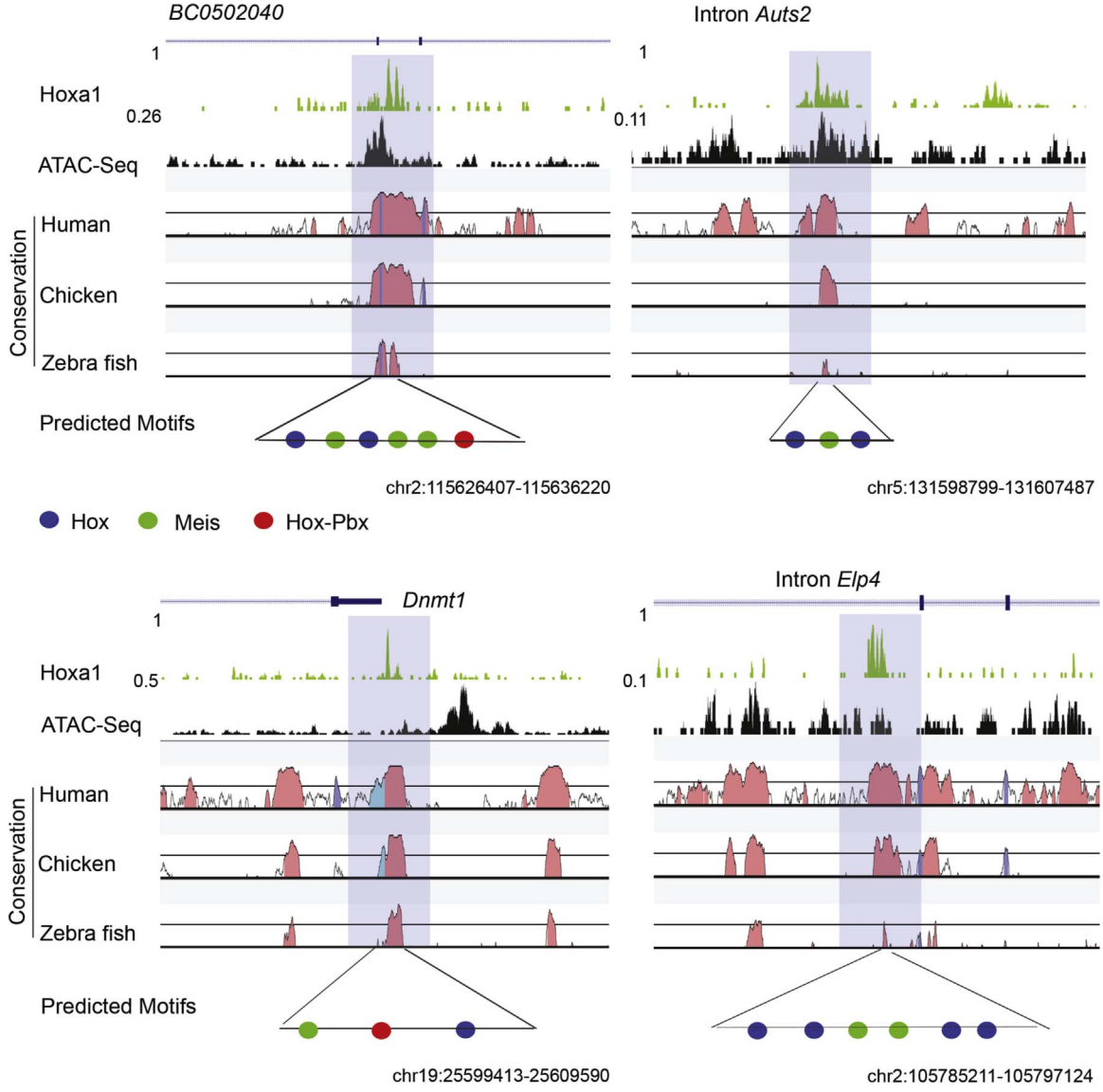


Fig. 8. Evolutionary conservation of binding regions and downstream targets of Hoxa1. Four examples showing evolutionarily conservation of Hoxa1-bound regions between mice, human, chicken and zebrafish. Browser shots show Hoxa1 binding and open chromatin state (ATAC-seq), along with Vista plots indicating level of sequence conservation. High conservation is observed in Hoxa1-bound regions and these regions contain binding motifs for Hox and cofactor proteins.

4.2. ChIP-seq and ChIP-nexus

ChIP-seq experiments were performed using modified upstate protocol (De Kumar et al., 2017b; Smith et al., 2010). ChIP-nexus was performed using a protocol developed by He and colleagues (He et al., 2015). Hoxa1 ChIP-seq and ChIP-nexus were performed using anti-flag M2 antibody (F1804, Sigma-aldrich). Other antibodies used were: Meis1/2 Antibody H-80 (SC-25412; Santacruz), Pbx 1/2/3 Antibody C-20 (sc-888; Santacruz), Anti-Histone H3K4me3 antibody (Ab1012; Abcam), Anti-Histone H3K4me1 antibody (ab8895; Abcam) and p300 Antibody C-20 (sc-585; Santacruz). were used against ChIP-seq of respective proteins and modifications. Data for ChIP-seq and ChIP-nexus are deposited in NCBI Sequence Read archive (SRA; http://www.ncbi.nim.nih.gov/sra/) under BioProject accession numbers PRJNA341679 and PRJNA335616. Raw reads were aligned to the UCSC mm10 mouse genome with bowtie2 2.2.0 (Langmead and Salzberg, 2012). Primary reads from each bam were normalized to reads-*per*-million and bigWig tracks visualized at the UCSC genome browser (https://genome.ucsc.edu/). Peaks were called with MACS2 2. 1.0. (Zhang et al., 2008), parameters "-g mm -p 0.25 –m 5 50". From each replicate, top 100,000 peaks based on p-value were compared with IDR 1.7.0 for reproducibility (https://sites.google.com/site/anshulkundaje/projects/idr) and valid pairs with IDR p-value ≤ 0.01 were taken as the peak list. For ChIP-nexus, alignments were separated by strand, and peaks were called on strandwise BAM files using MACS2, parameters "-g mm –nomodel", and no input.

4.3. Zebrafish reporter assay

PCR-purified genomic regions were cloned into the HLC vector (Parker et al., 2014) using the Gibson Assembly Master Mix (NEB) (Gibson et al., 2009). To test for regulatory activity injections were performed in embryos from Slusarski AB (wild type) or a line, (*egr2b:KalTA4BI-1xUASkCherry),* which has mCherry inserted in the endogenous *egr2b* locus (*krox20*) conveniently marking r3 and r5 with reporter expression as a reference (Distel et al., 2009). Zebrafish transgenesis was performed as previously described (Fisher et al., 2006). At least 100 embryos were injected for each construct. Embryos (F_0) were screened for GFP reporter expression at approximately 24 hpf and 48 hpf with a Leica M205FA microscope and images were captured for fluorescent and bright-field signals with a Leica DFC360FX camera using LAS AF imaging software. Images were cropped and altered for brightness and contrast using Adobe Photoshop CS6.

4.4. Injection of Hoxa1 mRNA

Hoxa1 and *Hoxa1-3Xflag-Myc* mRNA were cloned into the pCS2+ vector using Gibson assembly. Constructs were linearized with Not1-HF enzyme, gel extracted and used as templates for generation of synthetic capped mRNAs by the Ambion mMessage mMachine kit (AM1340, ThermoFisher), with the resulting mRNA being resuspended in water. For zebrafish embryo injections, 1 nl of mRNA at either 25 or 50 ng/μl, with 0.05% phenol red as a tracer, was injected into the cytoplasm of one-cell stage embryos. *Hoxa1* mRNA injections were performed on embryos arising from crosses of two pre-existing transgenic lines:*egr2b:KalTA4BI-1xUASkCherry,* which has mCherry inserted in the endogenous *egr2b* locus (*krox20*) conveniently marking r3 and r5 with reporter expression (Distel et al., 2009) and a line in which a mouse *Hoxb1* enhancer drives eGFP in r4 in a *Hoxb1*- and *Hoxa1*-dependent manner (Parker et al., 2014; Pöpperl et al., 1995; Zhang et al., 1994). At least 150 GFP positive embryos were collected in Trizol. RNA was extracted using Directzol (Zymo Research, R2070) as per manufacturer's instruction.

4.5. Microinjection of mouse embryos

Genomic regions to be tested for enhancer activity were cloned into the p1229BGZ40 vector (Yee and Rigby, 1993). Enhancer constructs were diluted to 4 ng/μl in microinjection buffer (10 mM Tris-HCl pH 7.5, 100uM EDTA pH 8.0, 100 mM NaCl, 1x Polyamine mix) and microinjected into the pronucleus of mouse C57B6/JxCBA-F1 single cell embryos. Microinjected fertilized eggs were cultured overnight in KSOM media. Two cell embryos were then transplanted into pseudo-pregnant C57/Bl6 or CD1 females. Embryos were harvested at E 9.5 and stained for β-galactosidase activity.

4.6. Ethics and animal experiments

All experiments involving zebrafish (Protocol ID: 2015-0149) and mice (Protocol ID: 2016-0164) were performed under approved protocols issued to REK as the PI by the Institutional Animal Care and Use Committee of the Stowers Institute for Medical Research.

4.7. HiC chromosome conformational capture assay

KH2 ES cells were differentiated using 3.3uM RA as with ChIP-seq samples. HiC experiments were carried out as described by van Berkum and colleagues using differentiated ES cells (van Berkum et al., 2010). Libraries were sequenced on the Illumina HiSeq 2500–51 bp, paired-end. Each end was aligned independently to UCSC mm10 genome with Bowtie2 (Langmead and Salzberg, 2012). Only trans-aligning pairs, and cis-aligning pairs > 1 kb apart, were analyzed. 1000 bp was selected as the noise threshold, as this was the saddle point of the cis-aligned mate-pair distance distribution, which was sharply bimodal (not shown). Read pairs connecting Hoxa1 peaks and Ensembl 80 gene regions were tabulated. Gene regions included 10 kb upstream and 1 kb downstream. Genes with Hoxa1 peak connections, or with Hoxa1 peaks within the actual gene region, were analyzed for pathway enrichment using the SPIA package in R (Tarca et al., 2009).

4.8. Analysis of enriched pathways and functional terms

For each peak set, all Ensembl 80 protein-coding nearest-neighbor genes upstream and downstream were identified and analyzed for functional enrichments. GO terms (http://geneontology.org/) downloaded May 2016 were compared between gene lists versus the rest of the genome. Terms over-enriched in neighbor genes by Fisher Exact Test with p-value ≤ 0.05 (BH-adjusted) and with at least 3 genes were accepted. Enrichment of pathways and processes were calculated by SPIA package in R (Tarca et al., 2009) from Bioconductor.

4.9. RNA-seq transcriptional profiling

10.5 dpc embryos were harvested and washed with ice cold PBS. Tissue was homogenized and RNA was isolated using Directzol (Zymogen) kit as per manufacturer's instruction. Only samples with RIN greater than 9 were used. Poly-A selected directional Illumina TruSeq libraries were prepared according to manufacturer's specifications. Libraries were sequenced on the Illumina HiSeq 2500–51 bp, single-end. Reads were aligned to UCSC mm10 genome with Tophat 2.1.10 (Kim et al., 2013) and Ensembl 80 gene models quantitated with HTSeq-count (Anders et al., 2015). Counts were analyzed in R and differentially-expressed genes selected with the EdgeR package, default methods (Robinson et al., 2010).

4.10. Sequence conservation alignments

60-way vertebrate PhyloP conservation scores for each position in the mm10 genome were downloaded from UCSC (ftp://hgdownload.soe.ucsc.edu/goldenPath/mm10/phyloP60way/

mm10.60way.phyloP60way/). Mean PhyloP scores per peak were compared for two peak sets: 1) Hoxa1 peaks which were the 3′ or 5′ neighbors of a differentially-expressed gene, and 2) randomly-selected intervals with the same width distribution as the Hoxa1 peaks.

4.11. ATAC-seq assay for open chromatin

ATAC-seq was done using 50,000 feeder free uninduced and differentiated ES cells as described by Buenrostro and colleagues (Buenrostro et al., 2015). ATAC-seq libraries were sequenced on the Illumina NextSeq 500, as 76 bp paired-end sequences. For sequence alignment and visualization, samples were processed like treated similar to ChIP-seq samples.

4.12. Data access

All Raw sequencing data is submitted at NCBI as SRA BioProject accession number PRJNA397296. All other original source data underlying this manuscript are deposited in the Stowers Original Data Repository at the time of publication and can be assessed at http://odr.stowers.org/websimr/.

Acknowledgements

We thank Stowers Institute (SIMR) Tissue Culture, Molecular Biology, Aquatics Facility, members of the Krumlauf and Zeitlinger labs and Leanne Wiedemann for valuable discussions and feedback. This research was supported by funds from the Stowers Institute to REK (Grant 1001).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ydbio.2017.09.033.

References

Alexander, T., Nolte, C., Krumlauf, R., 2009. Hox genes and segmentation of the hindbrain and axial skeleton. Annu. Rev. Cell Dev. Biol. 25, 431–456.

Alexandre, D., Clarke, J., Oxtoby, E., Yan, Y.-L., Jowett, T., Holder, N., 1996. Ectopic expression of *Hoxa-1* in the zebrafish alters the fate of the mandibular arch neural crest and phenocopies a retinoic acid -induced phenotype. Development 122, 735–746.

Alharbi, R.A., Pettengell, R., Pandha, H.S., Morgan, R., 2013. The role of HOX genes in normal hematopoiesis and acute leukemia. Leukemia 27, 1000–1008.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., Jaenisch, R., 2006. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis 44, 23–28.

Bertrand, N., Roux, M., Ryckebusch, L., Niederreither, K., Dolle, P., Moon, A., Capecchi, M., Zaffran, S., 2011. Hox genes define distinct progenitor sub-domains within the second heart field. Dev. Biol. 353, 266–274.

Bitu, C.C., Destro, M.F., Carrera, M., da Silva, S.D., Graner, E., Kowalski, L.P., Soares, F.A., Coletta, R.D., 2012. HOXA1 is overexpressed in oral squamous cell carcinomas and its expression is correlated with poor prognosis. BMC Cancer 12, 146.

Bosley, T.M., Alorainy, I.A., Salih, M.A., Aldhalaan, H.M., Abu-Amero, K.K., Oystreck, D.T., Tischfield, M.A., Engle, E.C., Erickson, R.P., 2008. The clinical spectrum of homozygous HOXA1 mutations. Am. J. Med. Genet. A 146A, 1235–1240.

Buckingham, M., Meilhac, S., Zaffran, S., 2005. Building the mammalian heart from two sources of myocardial cells. Nat. Rev. Genet. 6, 826–835.

Buenrostro, J.D., Wu, B., Chang, H.Y., Greenleaf, W.J., 2015. ATAC-seq: a Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. 109, 21–29.

Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates. Nature 376, 479–485.

Chen, J., Ruley, H.E., 1998. An enhancer element in the EphA2 (Eck) gene sufficient for rhombomere-specific expression is activated by HOXA1 and HOXB1 homeobox proteins. J. Biol. Chem. 273, 24670–24675.

Chisaka, O., Musci, T.S., Capecchi, M.R., 1992. Developmental defects of the ear, cranial nerves and hindbrain resulting from targeted disruption of the mouse homeobox gene *Hox-1.6*. Nature 355, 516–520.

De Kumar, B., Parker, H.J., Parrish, M.E., Lange, J.J., Slaughter, B.D., Unruh, J.R., Paulson, A., Krumlauf, R., 2017a. Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-ectodermal differentiation of ES cells. Proc. Natl. Acad. Sci. USA 114, 5838–5845.

De Kumar, B., Parker, H.J., Paulson, A., Parrish, M.E., Pushel, I., Singh, N.P., Zhang, Y., Slaughter, B.D., Unruh, J.R., Florens, L., Zeitlinger, J., Krumlauf, R., 2017b. HOXA1 and TALE proteins display cross-regulatory interactions and form a combinatorial binding code on HOXA1 targets. Genome Res. 27, 1501–1512.

De Kumar, B., Parrish, M.E., Slaughter, B.D., Unruh, J.R., Gogol, M., Seidel, C., Paulson, A., Li, H., Gaudenz, K., Peak, A., McDowell, W., Fleharty, B., Ahn, Y., Lin, C., Smith, E., Shilatifard, A., Krumlauf, R., 2015. Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. Genome Res. 25, 1229–1243.

Distel, M., Wullimann, M.F., Koster, R.W., 2009. Optimized Gal4 genetics for permanent gene expression mapping in zebrafish. Proc. Natl. Acad. Sci. USA 106, 13365–13370.

Donaldson, I.J., Amin, S., Hensman, J.J., Kutejova, E., Rattray, M., Lawrence, N., Hayes, A., Ward, C.M., Bobola, N., 2012. Genome-wide occupancy links Hoxa2 to Wnt-β-catenin signaling in mouse embryonic development. Nucleic Acids Res 40, 3990–4001.

Dupé, V., Davenne, M., Brocard, J., Dollé, P., Mark, M., Dierich, A., Chambon, P., Rijli, F., 1997. *In vivo* functional analysis of the *Hoxa1* 3′ retinoid response element (3′ RARE). Development 124, 399–410.

Fisher, S., Grice, E.A., Vinton, R.M., Bessling, S.L., Urasaki, A., Kawakami, K., McCallion, A.S., 2006. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nat. Protoc. 1, 1297–1305.

Fu, J., Jiang, M., Mirando, A.J., Yu, H.M., Hsu, W., 2009. Reciprocal regulation of Wnt and Gpr177/mouse Wntless is required for embryonic axis formation. Proc. Natl. Acad. Sci. USA 106, 18598–18603.

Gale, N.W., Holland, S.J., Valenzuela, D.M., Flenniken, A., Pan, L., Ryan, T.E., Henkemeyer, M., Strebhardt, K., Hirai, H., Wilkinson, D.G., Pawson, T., Davis, S., Yancopoulos, G.D., 1996. Eph receptors and ligands comprise two major specificity subclasses and are reciprocally compartmentalized during embryogenesis. Neuron 17, 9–19.

Gavalas, A., Ruhrberg, C., Livet, J., Henderson, C.E., Krumlauf, R., 2003. Neuronal defects in the hindbrain of Hoxa1, Hoxb1 and Hoxb2 mutants reflect regulatory interactions among these Hox genes. Development 130, 5663–5679.

Gavalas, A., Studer, M., Lumsden, A., Rijli, F.M., Krumlauf, R., Chambon, P., 1998. *Hoxa1* and *Hoxb1* synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. Development 125, 1123–1136.

Gavalas, A., Trainor, P., Ariza-McNaughton, L., Krumlauf, R., 2001. Synergy between Hoxa1 and Hoxb1: the relationship between arch patterning and the generation of cranial neural crest. Development 128, 3017–3027.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345.

Grice, J., Noyvert, B., Doglio, L., Elgar, G., 2015. A simple predictive enhancer syntax for hindbrain patterning is conserved in vertebrate genomes. PLoS One 10, e0130413.

Grimm, J., Sachs, M., Britsch, S., Di Cesare, S., Schwarz-Romond, T., Alitalo, K., Birchmeier, W., 2001. Novel p62dok family members, dok-4 and dok-5, are substrates of the c-Ret receptor tyrosine kinase and mediate neuronal differentiation. J. Cell Biol. 154, 345–354.

He, Q., Johnston, J., Zeitlinger, J., 2015. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. Nat. Biotechnol. 33, 395–401.

Hedlund, E., Karsten, S.L., Kudo, L., Geschwind, D.H., Carpenter, E.M., 2004. Identification of a Hoxd10-regulated transcriptional network and combinatorial interactions with Hoxa10 during spinal cord development. J. Neurosci. Res. 75, 307–319.

Helmbacher, F., Pujades, C., Desmarquet, C., Frain, M., Rijli, F.M., Chambon, P., Charnay, P., 1998. *Hoxa1* and *Krox20* synergize to control the development of rhombomere 3. Development 125, 4739–4748.

Holve, S., Friedman, B., Hoyme, H.E., Tarby, T.J., Johnstone, S.J., Erickson, R.P., Clericuzio, C.L., Cunniff, C., 2003. Athabascan brainstem dysgenesis syndrome. Am. J. Med. Genet. A 120A, 169–173.

Hunt, P., Gulisano, M., Cook, M., Sham, M.H., Faiella, A., Wilkinson, D., Boncinelli, E., Krumlauf, R., 1991. A distinct Hox code for the branchial region of the vertebrate head. Nature 353, 861–864.

Inoue, T., Chisaka, O., Matsunami, H., Takeichi, M., 1997. Cadherin-6 expression transiently delineates specific rhombomeres, other neural tube subdivisions, and neural crest subpopulations in mouse embryos. Dev. Biol. 183, 183–194.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Krumlauf, R., 2016. Hox Genes and the Hindbrain: a Study in Segments. Curr. Top. Dev. Biol. 116, 581–596.

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

Lei, H., Juan, A.H., Kim, M.S., Ruddle, F.H., 2006. Identification of a Hoxc8-regulated transcriptional network in mouse embryo fibroblast cells. Proc. Natl. Acad. Sci. USA 103, 10305–10309.

Lei, H., Wang, H., Juan, A.H., Ruddle, F.H., 2005. The identification of Hoxc8 target genes. Proc. Natl. Acad. Sci. USA 102, 2420–2424.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293.

Lin, C., Garrett, A.S., De Kumar, B., Smith, E.R., Gogol, M., Seidel, C., Krumlauf, R., Shilatifard, A., 2011. Dynamic transcriptional events in embryonic stem cells

mediated by the super elongation complex (SEC). Genes Dev. 25, 1486–1498.
Lufkin, T., Dierich, A., LeMeur, M., Mark, M., Chambon, P., 1991. Disruption of the *Hox-1.6* homeobox gene results in defects in a region corresponding to its rostral domain of expression. Cell 66, 1105–1119.
Makki, N., Capecchi, M.R., 2010. Hoxa1 lineage tracing indicates a direct role for Hoxa1 in the development of the inner ear, the heart, and the third rhombomere. Dev. Biol. 341, 499–509.
Makki, N., Capecchi, M.R., 2011. Identification of novel Hoxa1 downstream targets regulating hindbrain, neural crest and inner ear development. Dev. Biol. 357, 295–304.
Makki, N., Capecchi, M.R., 2012. Cardiovascular defects in a mouse model of HOXA1 syndrome. Hum. Mol. Genet. 21, 26–31.
Mallo, M., Wellik, D.M., Deschamps, J., 2010. Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. 344, 7–15.
McCabe, C.D., Spyropoulos, D.D., Martin, D., Moreno, C.S., 2008. Genome-wide analysis of the homeobox C6 transcriptional network in prostate cancer. Cancer Res. 68, 1988–1996.
Merabet, S., Mann, R.S., 2016. To be specific or not: the critical relationship between Hox And TALE proteins. Trends Genet 32, 334–347.
Minoux, M., Rijli, F.M., 2010. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. Development 137, 2605–2621.
Murphy, P., Hill, R.E., 1991. Expression of the mouse labial-like homeobox-containing genes, Hox 2.9 and Hox 1.6, during segmentation of the hindbrain. Development 111, 61–74.
Ostuni, R., Natoli, G., 2013. Lineages, cell types and functional states: a genomic view. Curr. Opin. Cell Biol. 25, 759–764.
Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., Natoli, G., 2013. Latent enhancers activated by stimulation in differentiated cells. Cell 152, 157–171.
Pan, Y., Zhang, J., Liu, W., Shu, P., Yin, B., Yuan, J., Qiang, B., Peng, X., 2013. Dok5 is involved in the signaling pathway of neurotrophin-3 against TrkC-induced apoptosis. Neurosci. Lett. 553, 46–51.
Parker, H.J., Bronner, M.E., Krumlauf, R., 2014. A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates. Nature 514, 490–493.
Parker, H.J., Bronner, M.E., Krumlauf, R., 2016. The vertebrate Hox gene regulatory network for hindbrain segmentation: evolution and diversification: coupling of a Hox gene regulatory network to hindbrain segmentation is an ancient trait originating at the base of vertebrates. BioEssays 38, 526–538.
Parker, H.J., Krumlauf, R., 2017. Segmental arithmetic: summing up the Hox gene regulatory network for hindbrain development in chordates. WIREs Dev. Biol. e286. http://dx.doi.org/10.1002/wdev.286, [Epub ahead of print]. PMID: 28771970.
Parker, H.J., Piccinelli, P., Sauka-Spengler, T., Bronner, M., Elgar, G., 2011. Ancient Pbx-Hox signatures define hundreds of vertebrate developmental enhancers. BMC Genom. 12, 637.
Pearson, J.C., Lemons, D., McGinnis, W., 2005. Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. 6, 893–904.
Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B.L., Couronne, O., Eisen, M.B., Visel, A., Rubin, E.M., 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444, 499–502.
Pöpperl, H., Bienz, M., Studer, M., Chan, S., Aparicio, S., Brenner, S., Mann, R., Krumlauf, R., 1995. Segmental expression of *Hoxb1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/Pbx*. Cell 81, 1031–1042.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.
Rossel, M., Capecchi, M.R., 1999. Mice mutant for both *Hoxa1* and *Hoxb1* show extensive remodeling of the hindbrain and defects in craniofacial development. Development 126, 5027–5040.
Smith, K.T., Martin-Brown, S.A., Florens, L., Washburn, M.P., Workman, J.L., 2010. Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. Chem. Biol. 17, 65–74.
Soshnikova, N., Dewaele, R., Janvier, P., Krumlauf, R., Duboule, D., 2013. Duplications of hox gene clusters and the emergence of vertebrates. Dev. Biol. 378, 194–199.
Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F.M., Chambon, P., Krumlauf, R., 1998. Genetic interactions between Hoxa1 and Hoxb1 reveal new roles in regulation of early hindbrain patterning. Development 125, 1025–1036.
Taminiau, A., Draime, A., Tys, J., Lambert, B., Vandeputte, J., Nguyen, N., Renard, P., Geerts, D., Rezsohazy, R., 2016. HOXA1 binds RBCK1/HOIL-1 and TRAF2 and modulates the TNF/NF-kappaB pathway in a transcription-independent manner. Nucleic Acids Res..
Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.S., Kim, C.J., Kusanovic, J.P., Romero, R., 2009. A novel signaling pathway impact analysis. Bioinformatics 25, 75–82.
Tischfield, M.A., Bosley, T.M., Salih, M.A., Alorainy, I.A., Sener, E.C., Nester, M.J., Oystreck, D.T., Chan, W.M., Andrews, C., Erickson, R.P., Engle, E.C., 2005. Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. Nat. Genet 37, 1035–1037.
Tvrdik, P., Capecchi, M.R., 2006. Reversal of hox1 gene subfunctionalization in the mouse. Dev. Cell 11, 239–250.
van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., Lander, E.S., 2010. Hi-C: a method to study the three-dimensional architecture of genomes. J. Vis. Exp..
Wardwell-Ozgo, J., Dogruluk, T., Gifford, A., Zhang, Y., Heffernan, T.P., van Doorn, R., Creighton, C.J., Chin, L., Scott, K.L., 2014. HOXA1 drives melanoma tumor growth and metastasis and elicits an invasion gene expression signature that prognosticates clinical outcome. Oncogene 33, 1017–1026.
Wen, J., Xia, Q., Wang, C., Liu, W., Chen, Y., Gao, J., Gong, Y., Yin, B., Ke, Y., Qiang, B., Yuan, J., Peng, X., 2009. Dok-5 is involved in cardiomyocyte differentiation through PKB/FOXO3a pathway. J. Mol. Cell Cardiol. 47, 761–769.
Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y.J., Cooke, J.E., Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3, e7.
Yao, L., Berman, B.P., Farnham, P.J., 2015. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. Crit. Rev. Biochem. Mol. Biol. 50, 550–573.
Yee, S.-P., Rigby, P.W.J., 1993. The regulation of *myogenin* gene expression during the embryonic development of the mouse. Genes Dev. 7, 1277–1289.
Yu, H.M., Jin, Y., Fu, J., Hsu, W., 2010. Expression of Gpr177, a Wnt trafficking regulator, in mouse embryogenesis. Dev. Dyn. 239, 2102–2109.
Zha, T.Z., Hu, B.S., Yu, H.F., Tan, Y.F., Zhang, Y., Zhang, K., 2012. Overexpression of HOXA1 correlates with poor prognosis in patients with hepatocellular carcinoma. Tumour Biol. 33, 2125–2134.
Zhang, M., Kim, H.J., Marshall, H., Gendron-Maguire, M., Lucas, D.A., Baron, A., Gudas, L.J., Gridley, T., Krumlauf, R., Grippo, J.F., 1994. Ectopic *Hoxa-1* induces rhombomere transformation in mouse hindbrain. Development 120, 2431–2442.
Zhang, P., Zhou, L., Pei, C., Lin, X., Yuan, Z., 2016. Dysfunction of Wntless triggers the retrograde Golgi-to-ER transport of Wingless and induces ER stress. Sci. Rep. 6, 19418.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, (R137).