Discovery of genes required for body axis and limb formation by global identification of conserved retinoic acid regulated enhancers and silencers

Marie Berenguer[1], Karolin F. Meyer[1], Jun Yin[2], and Gregg Duester[1,*]

[1]Development, Aging, and Regeneration Program

[2]Bioinformatics Core Facility

Sanford Burnham Prebys Medical Discovery Institute, 10901 N. Torrey Pines Road, La Jolla, CA 92037, USA

*Corresponding Author (duester@sbpdiscovery.org)

Keywords: Body axis formation; retinoic acid signaling; enhancer; silencer; *Aldh1a2* knockout; *Nr2f1*; *Nr2f2*

Short title: RNA-seq/ChIP-seq to find RA target genes

Abstract

Identification of target genes for transcription factors is hampered by the large number of genes whose expression changes when the factor is removed from a specific tissue and the numerous binding sites for the factor in the genome. Retinoic acid (RA) regulates transcription via RA receptors bound to RA response elements (RAREs) of which there are thousands in vertebrate genomes. Here, we combined ChIP-seq and RNA-seq on trunk tissue from wild-type and *Aldh1a2*-/- embryos lacking RA synthesis that exhibit body axis and forelimb defects. We identified a relatively small number of genes with altered expression when RA is missing that also have nearby RA-regulated deposition of H3K27ac (gene activation mark) or H3K27me3 (gene repression mark) associated with conserved RAREs. Such RAREs were identified in genes already known to be required for body axis and limb formation, plus many new candidate RA target genes were found. In addition to RARE enhancers, we identified many RARE silencers located near genes downregulated by RA. *Nr2f1*, *Nr2f2*, *Meis1*, and *Meis2* gene family members were identified as candidate RA target genes for embryonic trunk by our approach, and double knockouts of each family demonstrated previously unknown requirements for body axis and/or limb formation, thus validating our approach for identifying new RA target genes important for a specific tissue.

Author Summary

Studies that reveal how genes are normally regulated is fundamental to understanding how incorrect gene regulation causes disease. Genes are turned on and off by signaling factors that introduce epigenetic marks near specific genes. However, the number of genes regulated by a signaling factor is unknown as loss of the factor alters the expression of thousands of genes, most of which are secondary effects not due to altered epigenetic marks. We describe a significant advance in the ability to identify target genes for signaling factors in specific tissues. We developed a genetic approach for identifying target genes for retinoic acid (RA) signaling in which genomic studies were performed on mutant mouse embryos lacking RA signaling to identify genes whose expression changes when RA is missing that also have nearby RA-regulated changes in epigenetic marks. This approach was able to take a list of thousands of genes with altered expression when RA is lost, and reduce this to less than a hundred that are considered RA target genes.

Introduction

Retinoic acid (RA) is generated from retinol by the sequential activities of retinol dehydrogenase 10 (RDH10) [1] and aldehyde dehydrogenase 1A2 (ALDH1A2) [2, 3]. Knockout studies of these enzymes revealed an essential role for RA in many early developmental programs including those controlling hindbrain anteroposterior patterning, neuromesodermal progenitor (NMP) differentiation, spinal cord neurogenesis, somitogenesis, forelimb bud initiation, and heart anteroposterior patterning [4, 5]. RA functions as a ligand for nuclear RA receptors (RARs) that bind DNA sequences known as RA response elements (RAREs) as a heterodimer complex with retinoid X receptors (RXRs) [6]. Binding of RA to RAR alters the ability of RAREs to recruit nuclear receptor coactivators (NCOAs) that activate transcription or nuclear receptor corepressors (NCORs) that repress transcription [7]. Thus, RA functions are mediated by transcriptional activation or repression of key genes via RAREs.

Identification of genes that are transcriptional targets of RA has been difficult as loss or gain of RA activity alters the mRNA levels of thousands of genes in various cell lines or animals, perhaps most being indirect targets of RA or regulated post-transcriptionally. As RA target genes are dependent upon RAREs, identification of RAREs by RAR-binding studies, cell line transfection assays, and enhancer reporter transgenes in mouse or zebrafish has been used to identify RA target genes, but progress is slow as each gene is analyzed separately [5]. Genomic RAR chromatin immunoprecipitation (ChIP-seq) studies on mouse embryoid bodies reported ~14,000 potential RAREs in the mouse genome [8, 9], but it is unclear how many of these RAREs are required to regulate genes in any specific tissue, and many may not function in any tissue at any stage of development. Only a few RAREs have been shown to result in gene expression and developmental defects when subjected to deletion analysis in mouse, i.e. a RARE enhancer that activates *Hoxa1* in the hindbrain [10], a RARE enhancer that activates *Cdx1* in the spinal cord [11], and a RARE that functions as a silencer to repress caudal *Fgf8* in the developing trunk [7]. In one additional case, a RARE described within intron 2 of *Tbx5* that was suggested to be required for activation of *Tbx5* in the forelimb field based on a mouse enhancer reporter transgene [12] was found to be unnecessary for *Tbx5* activation and forelimb budding when subjected to CRISPR deletion analysis, suggesting *Tbx5* is not an RA target gene [13]. Many DNA control elements

(including RAREs) that exhibit appropriate tissue-specific expression in enhancer reporter transgene assays have been shown to not be required as an enhancer in vivo when deleted; this may be due to enhancer redundancy or because the control element is really not an enhancer but appeared to be when inserted as a transgene at a random location in the genome near a heterologous promoter [14]. Thus, additional methods are needed (preferably genome-wide) to locate functional RAREs in a particular tissue which can be used to identify new candidate RA target genes for a particular tissue.

Epigenetic studies have found that histone H3 K27 acetylation (H3K27ac) associates with gene activation and histone H3 K27 trimethylation (H3K27me3) associates with gene repression [15, 16]. Here, we performed genomic ChIP-seq (H3K27ac and H3K27me3) and RNA-seq studies on E8.5 mouse embryonic trunks from wild-type and *Aldh1a2*-/- mouse embryos lacking RA synthesis to globally identify RA target genes for embryonic trunk. Candidate targets are defined as genes whose mRNA levels are decreased or increased by genetic loss of RA that also have nearby RA-regulated epigenetic marks associated with RAREs. This approach was able to identify many previously reported RA target genes known to control embryonic trunk development (including all three known RA target genes from RARE knockout studies: *Hoxa1*, *Cdx1*, and *Fgf8*), plus we identified numerous new candidate RA target genes that may control trunk development. CRISPR knockout studies on several of these new candidate genes validated them as RA target genes required for body axis and/or limb formation. Our approach is generally applicable to determine tissue-specific target genes for any transcriptional regulator that has a knockout available.

Results

Comparison of RNA-seq and H3K27ac/H3K27me3 ChIP-seq for *Aldh1a2*-/- trunk tissue

We performed RNA-seq analysis comparing E8.5 trunk tissue from wild-type embryos and *Aldh1a2*-/- embryos that lack the ability to produce RA [3]. This analysis identified 4298 genes whose mRNA levels in trunk tissue are significantly decreased or increased when RA is absent (FPKM>0.5; a cut-off of log2 <-0.85 or >0.85 was employed to include *Sox2* known to be activated by RA; data available at GEO under accession number GSE131584).

We performed ChIP-seq analysis for H3K27ac and H3K27me3 epigenetic marks comparing

E8.5 trunk tissue from wild-type and *Aldh1a2*-/- embryos. This analysis identified 314 RA-regulated ChIP-seq peaks for H3K7ac (located within or near 214 genes) using a log2 cut-off of <-0.51 or >0.51 to include a RA-regulated peak near *Sox2* known to be activated by RA [17, 18]. We also identified 262 RA-regulated peaks for H3K27me3 (located within or near 141 genes) using a log2 cut-off of <-0.47 or >0.47 to include a RA-regulated peak near *Fst* known to be repressed by RA [19]; all ChIP-seq data available at GEO under accession number GSE131624. The much smaller number of RA-regulated peaks found with ChIP-seq compared to the very large number of genes found to have altered mRNA levels with RNA-seq suggests that for most genes loss of RA is not affecting mRNA levels directly via transcription.

In order to identify genes that are good candidates for being transcriptionally activated or repressed by RA (RA target genes), we compared our ChIP-seq and RNA-seq results to identify ChIP-seq peaks where nearby genes have significant changes in expression in wild-type vs *Aldh1a2*-/- based on RNA-seq. We found 73 RA-regulated peaks for H3K27ac near 63 genes with significant changes in expression when RA is lost (Table S1), plus 46 RA-regulated peaks for H3K27me3 near 41 genes with significant changes in expression when RA is lost (Table S2). As some genes have more than one nearby RA-regulated peak for H3K27ac or H3K27me3, and some genes have nearby RA-regulated peaks for both H3K27ac and H3K27me3 (*Rarb*, *Dhrs3*, *Fgf8*, *Cdx2*, *Fst*, *Meis1*, *Meis2*, *Nr2f2*, *Foxp4*, *Ptprs*, and *Zfhx4*), a total of 93 genes have nearby RA-regulated peaks for H3K27ac and/or H3K27me3 when RA is lost, thus identifying them as candidate RA target genes for E8.5 trunk development (Tables S1-S2; Fig. S1A).

Among the 93 candidate RA target genes for trunk development identified with our approach are included many examples of genes previously reported to be regulated by RA in the trunk based on studies of *Aldh1a2*-/- embryos [5, 20] or RA-treated NMPs [19]; this includes *Hoxa1*, *Cdx1*, *Rarb*, *Crabp2*, *Sox2*, *Dhrs3*, and *Pax6* whose expression is increased by RA, plus *Fgf8*, *Cdx2*, and *Fst* whose expression is decreased by RA (Table 1). H3K27ac peaks near *Cdx1, Rarb*, *Crabp2*, *Sox2*, *Dhrs3*, and *Pax6* are reduced in *Aldh1a2*-/- trunk consistent with these being RA-activated genes, whereas H3K27ac peaks near *Fgf8*, *Cdx2*, and *Fst* are increased in *Aldh1a2*-/- consistent with these being genes repressed by RA. Conversely, H3K27me3 peaks near *Fgf8*, *Cdx2*, and *Fst* are decreased in *Aldh1a2*-/-, whereas H3K27me3 peaks near *Rarb*, *Hoxa1*, and *Dhrs3* are increased in

Aldh1a2-/-, consistent with the former being genes repressed by RA and the latter being genes activated by RA (Table 1). In addition to these 10 well-established RA target genes, we also identified 83 additional genes that our findings indicate are candidate RA target genes for trunk, including *Nr2f1, Nr2f2, Meis1, Meis2,* and *Spry4* that were further examined (Table 1); differential expression of these genes in E8.5 wild-type vs *Aldh1a2*-/- trunk was validated by qRT-PCR (Fig. S2). Overall, our approach of identifying genes with RA-regulated expression that also have nearby RA-regulated peaks for H3K27ac and/or H3K27me3 in wild-type vs *Aldh1a2*-/- embryos is a reliable method for identifying RA target genes for trunk development.

Identification of RARE enhancers and silencers associated with RA-regulated deposition of H3K27ac or H3K27me epigenetic marks

As RA target genes need to be associated with a RARE, the DNA sequences within the RA-regulated H3K27ac/H3K27me3 ChIP-seq peaks we found near our list of 93 RA-regulated genes were searched for RARE sequences using the Homer transcription factor binding site program for the mm10 genome; we searched for three types of RAREs including those with a 6 bp direct repeat separated by either 5 bp (DR5), 2 bp (DR2), or 1 bp (DR1) [5], and the presence or absence of RAREs is summarized (Tables S1 and S2). We found that 46 of these 93 genes contained at least one RARE in their nearby RA-regulated H3K27ac and/or H3K27me3 ChIP-seq peaks, thus narrowing down our list of candidate RA target genes to 49% of the genes originally identified. Our approach identified the three RAREs previously shown to have required functions during trunk development in vivo by knockout studies (RAREs for *Hoxa1*, *Cdx1*, *Fgf8*) plus several RAREs associated with known RA-regulated genes in the E8.5 trunk from *Aldh1a2*-/- studies (*Rarb*, *Crabp2*, *Sox2*, *Dhrs3*, *Cdx2*, *Fst*), thus validating our approach for identifying functional RAREs and hence RA target genes for trunk. The sequences of the RAREs for all the known RA target genes and new candidate RA target genes identified here are summarized; included are 65 RARE enhancers near 34 RA-activated genes (associated with increased H3K27ac and/or decreased H3K27me3) and 20 RARE silencers near 12 RA-repressed genes (associated with increased H3K27me3 and/or decreased H3K27ac) (Table S3).

The results here provide evidence that many of the RA-regulated H3K27ac and H3K27me3

marks are associated with regulation of the nearest genes, however it is possible that some H3K27ac and H3K27me3 RA-regulated peaks may be related to RA-regulated genes located further away in the same topologically associated domain (TAD). In order to address this issue, we assigned each RA-regulated H3K27ac and H3K27me3 peak to a TAD using the 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/view.php). Then the genes in each TAD containing an RA-regulated peak were searched in our RNA-seq database to identify RA-regulated genes, and if at least one gene was found we determined whether a RARE is present in the ChIP-seq peak. This analysis resulted in the identification of 82 additional RARE enhancers near RA-activated genes, and 40 additional RARE silencers near RA-repressed genes, where the gene is not the gene nearest to the RARE in the TAD; in some cases more than one RA-regulated gene was identified in a TAD (Table S3).

Up to now, *Fgf8* represents the only example of a gene that is directly repressed by RA at the transcriptional level as shown by developmental defects upon knockout of the RARE at -4.1 kb, and by the ability of this RARE to stimulate binding of NCOR and PRC2 plus deposition of H3K27me3 in an RA-dependent manner [7, 21]. Here, in addition to *Fgf8*, we found many more candidates for genes repressed by RA in trunk based on identification of nearby RARE silencers (Tables S3).

Analysis of known RA target genes for trunk validates our approach

The RA-regulated H3K27ac and/or H3K27me3 peaks we identified near *Rarb*, *Crabp2*, *Hoxa1*, and *Cdx1* all overlap previously reported RAREs for these genes (Fig. 1). In the case of *Rarb*, the DR5 RARE in the 5'-untranslated region [22] overlaps RA-regulated peaks for both H3K27ac and H3K27me3, suggesting that this RARE in the presence of RA stimulates deposition of H3K27ac and removal of H3K27me3 during activation of *Rarb*; we also identified a DR1 RARE in the 5'-noncoding region of *Rarb* within an RA-regulated H3K27me3 ChIP-seq peak (Fig. 1A). For *Crabp2*, two closely-spaced RAREs previously reported in the 5'-noncoding region [23] associate with RA-regulated peaks for H3K27ac, plus another RARE we identified in the 3'-noncoding region also associates with changes in H3K27ac (Fig. 1B). For *Hoxa1*, the RARE located in the 3'-noncoding region is associated with RA-regulated peaks for both H3K27ac and H3K27me3, plus another RARE we identified in the 3'-untranslated region is associated with RA-regulated peaks for

H3K27me3 (Fig. 1C); importantly, knockout studies on the *Hoxa1* RARE in the 3'-noncoding region demonstrated that it is required *in vivo* for *Hoxa1* expression and normal development [10]. For *Cdx1*, two RAREs have been reported, one in the 5'-noncoding region that was shown by knockout studies to be required for *Cdx1* expression and body axis development [11], plus another RARE in intron 1 [24]. Both of these *Cdx1* RAREs are overlapped by RA-regulated peaks for both H3K27ac and H3K27me3 (Fig. 1D). These findings demonstrate that our approach can identify genes that are already known to be transcriptionally activated by RA via a RARE enhancer.

Identification of RA-regulated epigenetic marks and RAREs near RA-regulated genes known to control neuromesodermal progenitors (NMPs)

Ingenuity Pathway Analysis (IPA) of our list of 93 RA target genes shows enrichment for the pathway "development of body trunk", including *Sox2*, *Cdx2*, and *Fgf8* known to be required for NMP function during trunk development (Fig. S1B). NMPs are bipotential progenitor cells in the caudal region co-expressing *Sox2* and *T/Bra* that undergo balanced differentiation to either spinal cord neuroectoderm or presomitic mesoderm to generate the post-cranial body axis [25-32]. NMPs are first observed in mouse embryos at about E8.0 near the node and caudal lateral epiblast lying on each side of the primitive streak [33-35]. Caudal Wnt and FGF signals are required to establish and maintain NMPs [33, 35-40]. Also, *Cdx2* is required for establishment of NMPs [32]. During development, RA is first produced at E7.5 in presomitic mesoderm expressing *Aldh1a2* to generate an anteroposterior gradient of RA with high activity in the trunk and low activity caudally [5]. Loss of RA does not prevent establishment or maintenance of NMPs, but does result in unbalanced differentiation of NMPs, with decreased caudal *Sox2* expression and decreased appearance of neural progenitors, plus increased caudal *Fgf8* expression and increased appearance of mesodermal progenitors and small somites due to encroachment of caudal *Fgf8* expression into the trunk where it reduces epithelial condensation of presomitic mesoderm needed to form somites [18, 19, 41, 42]. Also, *Cdx2* expression is increased when RA is lost in *Aldh1a2*-/- embryos [43].

Here, when RA is lost we observed RA-regulated H3K27ac and/or H3K27me3 peaks near several genes required for NMP function that show decreased (*Sox2*) or increased (*Fgf8* and *Cdx2*) expression (Fig. 2A-C). Most of these RA-regulated peaks contain RAREs, providing evidence that

Sox2, *Fgf8*, and *Cdx2* are RA target genes (Table S3). For *Sox2*, we observed two RA-regulated H3K27ac ChIP-seq peaks, but only the one in the 3'-noncoding region was found to have a RARE (Fig. 2A). In the case of *Fgf8*, previous studies reporting knockout of the RARE located in the 5'-noncoding region at -4.1 kb resulted in increased caudal *Fgf8* expression and a small somite phenotype (although the defect is not as severe as for *Aldh1a2*-/- embryos), demonstrating that this RARE functions in vivo as a silencer by RA-dependent recruitment of nuclear receptor corepressors [7]; RARE redundancy may explain the milder phenotype as our approach suggests that *Fgf8* has three RARE silencers (Fig. 2B). RARE redundancy may be common as we also observe that *Cdx2* has three RARE silencers (Fig. 2C), and our overall analysis shows that many genes have more than one nearby RARE (Table S3). Overall, these findings indicate that RA controls NMP differentiation at the transcriptional level by activating *Sox2* and repressing *Fgf8* and *Cdx2* as progenitor cells progress from a caudal to a trunk location.

Evidence for genes regulated indirectly by RA at the transcriptional level

Our studies show that many genes that are downregulated or upregulated following loss of RA have nearby RA-regulated peaks for H3K27ac or H3K27me3 that do not contain RAREs (Tables S1-S2). Such genes may be indirectly activated or repressed by RA at the transcriptional level. In the case of *Pax6*, our results indicate that RA stimulates H3K27ac deposition in *Pax6* introns 2 and 6 that do not contain RAREs (Fig. 3A). Previous studies identified an enhancer in *Pax6* intron 6 containing a SOXB1 binding site that is important for activation in the spinal cord [44]. Also, activation of *Pax6* in the spinal cord requires CDX proteins in the posterior-most neural tube, and CDX binding sites have been identified in *Pax6* introns [45]; in addition to expression in the caudal progenitor zone, mouse *Cdx1* is expressed in the posterior neural plate where *Pax6* is activated, and this expression domain requires RA [43]. Activation of *Pax6* also requires that caudal FGF signaling be downregulated [42]. Thus, it is possible that the RA requirement for *Pax6* activation operates through several indirect mechanisms due to the ability of RA to activate *Sox2* and *Cdx1*, and repress *Fgf8* (Figs. 1, 2). Also, other studies have shown that spinal cord *Pax6* is activated by *Neurog2* [46].

We also observed that *Spry4* (shown here to be down-regulated by RA) does not have a RARE

associated with its RA-regulated ChIP-seq peak for H3K27me3 (Fig. 3B). Many of the RA-regulated ChIP-seq peaks observed with our approach that do not contain RAREs may be indirect RA-regulated peaks that contain DNA binding sites for transcription factors other than RARs whose expression or activity is altered by loss of RA, thus resulting in changes for H3K27ac/H3K27me3 marks that are caused by the other transcription factors.

***Nr2f* and *Meis* gene families have nearby RA-regulated epigenetic marks and RARE enhancers**

We identified two gene families (*Nr2f* and *Meis*) where two family members have decreased expression when RA is lost and nearby RA-regulated peaks for H3K27ac or H3K27me3 containing RAREs.

Previous studies suggested that *Nr2f* genes are activated by RA in *Ciona*, zebrafish, and mouse F9 cells [49-51]. Here, *Nf2f1* and *Nr2f2* were both found to have a single RARE in the 5'-noncoding region close to exon 1 that is overlapped by or close to the edge of RA-regulated H3K27ac and H3K27me3 peaks (Fig. 4A-B). Recent studies in zebrafish identified RAREs in similar locations in the *nr2f1a* and *nr2f2* genes [51] and this conservation was detected by our analysis (Table S3).

Meis1 and *Meis2* were previously shown to be upregulated by RA in chick limbs treated with RA [52], *Meis1* is activated by RA in embryonic stem cells [53], and *Meis2* is activated by RA in cell line studies and a RARE located in its 5'-noncoding region was previously reported [9]. Here, *Meis1* was found to have four RAREs in introns 1, 6 and 7 that are overlapped by RA-regulated peaks for H3K27ac and/or H3K27me3, plus we identified the previously reported RARE in the 5-noncoding region that is located at the edge of a small RA-regulated H3K27ac peak (Fig. 4C). *Meis2* was found to have two RAREs that are overlapped by RA-regulated peaks for H3K27ac and/or H3K27me3, one in the 5'-noncoding region close to exon 1 and another in intron 7 (Fig. 4D). Together, these studies identify *Nf2f1*, *Nr2f2*, *Meis1*, and *Meis2* as candidate RA target genes in the developing trunk.

Conservation of RAREs identified with our approach identifies candidate RA target genes

The RARE enhancers and RARE silencers we identified were searched for evolutionary

conservation using the UCSC genome browser. Among the RAREs located very close to RA-regulated genes we found 6 RAREs that are conserved from mouse to zebrafish, 11 conserved to frog (*X. tropicalis*), 18 conserved to reptile (lizard; painted turtle), 20 conserved to bird (chicken; turkey), 39 conserved to human, 65 conserved to rodent (rat), and 20 that are not conserved with rat (Table S3). The large number of RAREs (i.e. 20) conserved beyond mammals to bird, lizard, frog, or fish demonstrate that our approach is able to identify well-conserved RAREs that point to excellent candidates for RA target genes in the developing trunk. Among the additional RAREs we found located further away in the TAD from an RA-regulated gene we identified only 4 more RAREs conserved beyond mammals to bird, lizard, frog, or fish, thus bringing the total to 24 (Table S3). Thus, most of the highly conserved RAREs we identified are located very close to an RA-regulated gene rather than further distant in the TAD. In addition, all these highly conserved RAREs are either identical to the RARE consensus or have only one mismatch. Here we summarize the 24 most highly conserved RAREs and the corresponding 38 best candidates for RA target genes (Table 2).

Our list of best candidate RA target genes (Table 2) includes several for which knockout studies have already demonstrated required functions during trunk development, i.e. in RA signaling (*Rarb, Dhrs3*), body axis formation (*Hoxa1, Hoxa4, Hoxa9, Sox2, Fgf8, Pbx1, Tshz1, Zbtb16*), and foregut formation (*Foxp4*); mouse knockout data summarized by Mouse Genome Informatics (http://www.informatics.jax.org). This list also includes many for which knockout studies have either not been performed or knockouts resulted in no reported early developmental defects; this topic is addressed below.

Nr2f1 **and** ***Nr2f2*** **function redundantly to control body axis formation**

In order to be an RA target gene, the gene must not only be associated with a RARE, but must perform a function downstream of RA during trunk development which can be determined by gene knockout studies. Here, we sought to validate our approach by performing knockout studies on some of the new candidate RA target genes, particularly those that have nearby RAREs that are highly conserved. One can also undertake deletion studies of the RAREs, but as genes are often controlled by redundant enhancers, studies in which predicted enhancers are deleted are often inconclusive [14, 54-56]; this includes knockout studies we performed for two presumed enhancers

for *Tbx5* in forelimb bud, one of which was reported to have a RARE [13]. Below, we describe gene knockout studies on candidate RA target genes with nearby highly conserved RAREs to determine if these genes have a required function in trunk development.

Nr2f1 and *Nr2f2* were selected for gene knockout as they both have nearby RARE enhancers (identified by our H3K27ac/H3K27me3 ChIP-seq analysis) that are conserved from mouse to zebrafish (Table 2). *Nr2f1* (formerly known as COUP-TFI) and *Nr2f2* (formerly known as COUP-TFII) are both expressed at E8.5 in somites and presomitic mesoderm but not spinal cord, suggesting they may function in mesoderm formation during body axis formation [57, 58].

The *Nr2f1* knockout is lethal at birth with brain defects but no somite, spinal cord, or body axis defects are observed [59]. The *Nr2f2* knockout is lethal at E10.5 with defects in heart development but not body axis formation [60]. As redundancy may have masked a body axis defect, we generated *Nr2f1/Nr2f2* double mutants. CRISPR/Cas9 gene editing of fertilized mouse oocytes was employed with sgRNAs designed to generate frameshift knockout deletions in the second exons of both *Nr2f1* and *Nr2f2*. After dissecting embryos at E9.0, we obtained *Nr2f1/Nr2f2* double knockouts that exhibit a body axis growth defect, more similar in size to that of wild-type E8.25 embryos (Fig. 5). Genotyping showed that embryos carrying 1 or 2 knockout alleles are normal in size compared to E9.0 wild-type (Fig. 5A), whereas embryos carrying either 3 or 4 knockout alleles have a defect in body axis extension and are similar in size to E8.25 wild-type; *n*=7 (Fig. 5B-C). Staining for *Uncx* somite expression demonstrated that embryos with 1-2 knockout alleles all have a normal number of somites with normal size (Fig. 5A), whereas embryos with 3-4 knockout alleles all have less somites that are smaller in size; embryos with 3 knockout alleles have a similar defect to those with 4 knockout alleles (Fig. 5B-C). As E9.0 *Nr2f1/Nr2f2* mutants carrying 3-4 knockout alleles are more similar in size to E8.25 wild-type, in order to estimate somite size along the anteroposterior axis we compared them to *Uncx*-stained E8.25 wild-type embryos (Fig. 5D), thus revealing that the E9.0 mutants have somites about 57% the size of somites in E8.25 wild-type embryos, showing they have a specific defect in trunk development rather than a global body growth defect (Fig. 5E). Overall, our findings show that loss of 3 or 4 alleles of *Nr2f1* and *Nr2f2* hinders body axis formation and results in smaller somites. This observation provides evidence that our approach of identifying RA-regulated epigenetic marks in wild-type and RA-deficient embryos can identify new RA target

genes essential for trunk development.

Meis1 and *Meis2* function redundantly to control both body axis and limb formation

Meis1 and *Meis2* were selected for gene knockout as *Meis1* has a nearby RARE enhancer conserved from mouse to frog, and *Meis2* has a nearby RARE enhancer conserved from mouse to bird (Table S3). *Meis1* and *Meis2* are both expressed throughout the trunk and in the proximal regions of limb buds [52].

The *Meis1* knockout is lethal at E11.5 with hematopoietic defects, but no body axis or limb defects are observed [61]. The *Meis2* knockout is lethal at E14.5 with defects in cranial and cardiac neural crest, but no defects in body axis or limb formation were observed [62]. As redundancy may have masked a body axis or limb defect, we generated *Meis1/Meis2* double mutants via CRISPR/Cas9 gene editing of fertilized mouse oocytes employing sgRNAs designed to generate frameshift knockout deletions in the second exons of both *Meis1* and *Meis2*. Embryos were dissected at E10.5 and stained for somite *Uncx* expression. Genotyping showed that E10.5 embryos carrying 1 or 2 knockout alleles for *Meis1/Meis2* are normal in size with normal size somites compared to E10.5 wild-type (Fig. 6A). However, E10.5 embryos carrying 3 or 4 knockout alleles for *Meis1/Meis2* exhibit a body axis extension defect and are either similar in size to *Uncx*-stained E9.5 wild-type embryos (n=3) or smaller (n=4); comparison of somite size along the anteroposterior axis for five of these E10.5 mutants shows that somite sizes range from that seen in E9.5 wild-type to about 40% smaller (Fig. 6B-D). We also observed that E10.5 *Meis1/Meis2* mutants carrying 3-4 knockout alleles that grew similar in size to E9.5 embryos exhibit a lack of forelimb bud outgrowth; n=3 (Fig. 6E). Overall, our findings show that loss of 3 or 4 alleles of *Meis1* and *Meis2* hinders body axis and forelimb formation, thus providing further evidence that the methods we describe here can identify new RA target genes essential for development.

Discussion

Our epigenetic ChIP-seq studies combined with RNA-seq on wild-type vs mutant RA-deficient tissues, along with identification of associated RARE enhancers and silencers, provides a means of identifying new RA target genes by focusing on RA-regulated genes that also have changes in

nearby H3K27ac and/or H3K27me3 epigenetic marks when RA is lost. Our approach can be used to identify target genes for any transcriptional regulator that has an available knockout.

Here, in our studies on *Aldh1a2*-/- trunk tissue, we were able to narrow down 4298 genes identified with RNA-seq that have significant changes in gene expression following loss of RA to less than 40 excellent candidate RA target genes in E8.5 trunk that also have significant changes in H3K27ac and/or H3K27me3 marks (located nearby or further away in the same TAD) associated with highly conserved RAREs. Our method allows one to identify genes that are most likely to be transcriptional targets of the RA signaling pathway as opposed to those whose expression is changed by effects downstream of RARs and RA signaling such as changes in expression or activity of other transcription factors or post-transcriptional changes in mRNA abundance. Our findings allow us to predict that some genes are likely to be indirect transcriptional targets of RA as they have nearby RA-regulated peaks for H3K27ac or H3K27me3 but no RAREs, i.e. *Pax6* that is transcriptionally regulated by factors whose expression is altered by loss of RA including *Sox2* [44], *Cdx* [45], and *Fgf8* [42]. While H3K27ac and H3K27me3 epigenetic marks are quite commonly observed near genes during activation or repression, respectively, it is possible that additional genes regulated transcriptionally by RA may be identified by further ChIP-seq studies examining other epigenetic marks, coactivators, or corepressors.

Our findings provide evidence for additional RARE silencers. Previous methods designed to identify RAREs favored discovery of RARE enhancers as studies were designed to find DNA elements that when fused to a heterologous promoter and marker gene would stimulate expression of the marker gene in the presence of RA. Also, when nuclear receptor coactivators (NCOA) and corepressors (NCOR) that control RA signaling were originally discovered, the model proposed for their function suggested that binding of RA to RAR favored binding of NCOA to activate transcription, with unliganded RAR favoring release of NCOA and binding of NCOR to repress transcription [63]. However, analysis of the *Fgf8* RARE silencer at -4.1 kb demonstrated that RARs bound to RAREs can recruit NCOR in an RA-dependent manner, plus this RARE is required for normal body axis extension [7]. The *Fgf8* RARE silencer was also found to recruit Polycomb Repressive Complex 2 (PRC2) and histone deacetylase 1 (HDAC1) in an RA-dependent manner, providing further evidence that RA can directly control gene silencing [21]. Here, we identified

additional RARE silencers near *Fgf8* and *Cdx2* plus several additional genes. Our studies indicate that RARE silencers are less common than RARE enhancers, and we found that *Fgf8* is the only gene associated with a RARE silencer conserved beyond mammals. These additional RARE silencers can be further examined in comparison to the *Fgf8* RARE silencer to determine the mechanism through which RA directly represses transcription. It will be important to determine how RAREs can function as RA-dependent enhancers for some genes but RA-dependent silencers for other genes.

RA has been shown to be required for balanced NMP differentiation during body axis formation by favoring a neural fate over a mesodermal fate [18, 28, 31]. Our studies provide evidence that RA regulates several genes at the trunk/caudal border needed for NMP differentiation at the transcriptional level; i.e. activation of *Sox2* in the neural plate that favors neural differentiation, repression of *Fgf8* that favors mesodermal differentiation, and repression of *Cdx2* that helps define the location of NMPs. We now provide evidence for a RARE enhancer that activates *Sox2*, three RARE silencers that repress *Cdx2*, and three RARE silencers for *Fgf8*. As the knockout of the original *Fgf8* RARE silencer at -4.1 kb exhibited a body axis phenotype less severe than loss of RA in *Aldh1a2*-/- embryos [7], it is possible that the additional RARE silencers found here provide redundant functions for *Fgf8* repression.

Our observation of highly conserved RARE enhancers near two members of two different gene families (*Nr2f* and *Meis*) was intriguing as it suggested that these gene family members may play redundant roles in body axis formation downstream of RA. Our *Nr2f1/Nr2f2* double knockout studies indeed revealed a defect in body axis formation and small somites that is not observed in each single knockout. Interestingly, zebrafish *nr2f1a/nr2f2* double knockout embryos reported recently exhibit a heart defect more severe than each single knockout, but not a body axis defect [51]. However, this observation is consistent with studies showing that RA is not required for NMP differentiation or body axis formation in zebrafish [64, 65]. Thus, it appears that the ancestral function of *Nr2f* genes in fish was to control heart formation, but that during evolution another function to control body axis formation was added.

The *Meis1/Meis2* double knockouts we describe here revealed an unexpected function for *Meis* genes in body axis extension and forelimb initiation. *Meis1* and *Meis2* are markers of the proximal

limb during forelimb and hindlimb development and were proposed to be activated by RA in the proximal limb as part of the proximodistal limb patterning mechanism in chick embryos [52, 66, 67]. However, knockout of *Rdh10* required to generate RA demonstrated that complete loss of RA in the limb fields prior to and during limb development did not affect hindlimb initiation or patterning, whereas forelimbs were stunted but with *Meis1* and *Meis2* expression still maintained in a proximal position in both stunted forelimbs and hindlimbs [68, 69]; reviewed in [5]. Our epigenetic results here support the previous proposal that RA can up-regulate *Meis1* and *Meis2* (but in the body axis prior to limb formation as opposed to the limb itself) and we provide evidence that *Meis1* and *Meis2* are transcriptional targets of RA in the body axis. Future studies can be directed at understanding the mechanism through which *Meis1* and *Meis2* control body axis and limb formation.

Our studies demonstrate the power of combining gene knockouts, ChIP-seq, and RNA-seq to identify RA target genes for a particular tissue. Such knowledge is essential for determining the mechanisms through which RA controls developmental pathways and adult processes. A similar approach can be used to determine the target genes for any transcriptional regulator for which a knockout is available, thus accelerating the ability to understand gene regulatory networks in general.

Methods

Generation of *Aldh1a2*-/- mouse embryos and isolation of trunk tissue

Aldh1a2-/- mice have been previously described [3]. E8.5 *Aldh1a2*-/- embryos were generated via timed matings of heterozygous parents; genotyping was performed by PCR analysis of yolk sac DNA. E8.5 trunk tissue was released from the rest of the embryo by dissecting across the posterior hindbrain (to remove the head and heart) and just posterior to the most recently formed somite (to remove the caudal progenitor zone) as previously described [21]. All mouse studies conformed to the regulatory standards adopted by the Institutional Animal Care and Use Committee at the SBP Medical Discovery Institute which approved this study under Animal Welfare Assurance Number A3053-01 (approval #18-092).

RNA-seq analysis

Total RNA was extracted from E8.5 trunk tissue (two wild-type trunks and two *Aldh1a2*-/- trunks) and DNA sequencing libraries were prepared using the SMARTer Stranded Total RNA-Seq Kit v2 Pico Input Mammalian (Takara). Sequencing was performed on Illumina NextSeq 500, generating 40 million reads per sample with single read lengths of 75 bp. Sequences were aligned to the mouse mm10 reference genome using TopHat splice-aware aligner; transcript abundance was calculated using Expectation-Maximization approach; fragments per kilobase of transcript per million mapped reads (FPKM) was used for sample normalization; Generalized Linear Model likelihood ratio test in edgeR software was used as a differential test. High throughput DNA sequencing was performed in the Sanford Burnham Prebys Genomics Core.

qRT-PCR analysis

Total RNA was extracted from 20 trunks of either E8.5 wild-type or *Aldh1a2*-/- embryos with the RNeasy Micro Kit (Qiagen #74004). Reverse transcription was performed with the High-Capacity cDNA RT Kit (Thermo Fisher Scientific #4368814). Quantitative PCR (qPCR) was performed using Power SYBR Green PCR Master Mix (Life Tech Supply #4367659). Relative quantitation was performed using the ddCt method with the control being expression of *Rpl10a*. Primers used for PCR (5'-3'):

Rpl10a-F	ACCAGCAGCACTGTGATGAA
Rpl10a-R	cAGGATACGTGGgATCTGCT
Rarb-F	CTCTCAAAGCCTGCCTCAGT
Rarb-R	GTGGTAGCCCGATGACTTGT
Nr2f1-F	TCAGGAACAGGTGGAGAAGC
Nr2f1-R	ACATACTCCTCCAGGGCACA
Nr2f2-F	GACTCCGCCGAGTATAGCTG
Nr2f2-R	GAAGCAAGAGCTTTCCGAAC
Meis1-F	CAGAAAAAGCAGTTGGCACA
Meis1-R	TGCTGACCGTCCATTACAAA
Meis2-F	AACAGTTAGCGCAAGACACG

Meis2-R GGGCTGACCCTCTGGACTAT

Spry4-F CCTGTCTGCTGTGCTACCTG

Spry4-R AAGGCTTGTCAGACCTGCTG

Chromatin immunoprecipitation (ChIP) sample preparation for ChIP-seq

For ChIP-seq we used trunk tissue from E8.5 wild-type or *Aldh1a2*-/- embryos dissected in modified PBS, i.e. phosphate-buffered saline containing 1X complete protease inhibitors (concentration recommended by use of soluble EDTA-free tablets sold by Roche #11873580001) and 10 mM sodium butyrate as a histone deacetylase inhibitor (Sigma # B5887). Samples were processed similar to previous methods [70]. Dissected trunks were briefly centrifuged in 1.5 ml tubes and excess PBS dissection buffer was removed. For cross-linking of chromatin DNA and proteins, 500 μl 1% paraformaldehyde was added, the trunk samples were minced by pipetting up and down with a 200 μl pipette tip and then incubated at room temperature for 15 min. To stop the cross-linking reaction, 55 μl of 1.25 M glycine was added and samples were rocked at room temperature for 5 min. Samples were centrifuged at 5000 rpm for 5 min and the supernatant was carefully removed and discarded. A wash was performed in which 1000 μl of ice-cold modified PBS was added and mixed by vortex followed by centrifugation at 5000 rpm for 5 min and careful removal of supernatant that was discarded. This wash was repeated. Cross-linked trunk samples were stored at -80C until enough were collected to proceed, i.e. 100 wild-type trunks and 100 *Aldh1a2*-/- trunks to perform ChIP-seq with two antibodies in duplicate.

Chromatin was fragmented by sonication. Cross-linked trunk samples were pooled, briefly centrifuged, and excess PBS removed. 490 μl lysis buffer (modified PBS containing 1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.0) was added, mixed by vortexing, then samples were incubated on ice for 10 min. Samples were divided into four sonication microtubes (Covaris AFA Fiber Pre-Slit Snap-Cap 6x16 mm, #520045) with 120 μl per tube. Sonication was performed with a Covaris Sonicator with the following settings - Duty: 5%, Cycle: 200, Intensity: 4, #Cycles: 10, 60 sec each for a total sonication time of 14 min. The contents of the four tubes were re-combined by transfer to a single 1.5 ml microtube which was then centrifuged for 10 min at 13,000 rpm and the supernatants transferred to a fresh 1.5 ml microtube. These conditions were found to shear trunk

DNA to an average size of 300 bp using a 5 μl sample for Bioanalyzer analysis. At this point 20 μl was removed for each sample (wild-type trunks and *Aldh1a2*-/- trunks) and stored at -20C to serve as input DNA for ChIP-seq.

Each sample was divided into four 100 μl aliquots to perform immunoprecipitation with two antibodies in duplicate. Immunoprecipitation was performed using the Pierce Magnetic ChIP Kit (Thermo Scientific, #26157) following the manufacturer's instructions and ChIP-grade antibodies for H3K27ac (Active Motif, Cat#39133) or H3K27me3 (Active Motif, Cat#39155). The immunoprecipitated samples and input samples were subjected to reversal of cross-linking by adding water to 500 μl and 20 μl 5 M NaCl, vortexing and incubation at 65C for 4 hr; then addition of 2.6 μl RNase (10 mg/ml), vortexing and incubation at 37C for 30 min; then addition of 10 μl 0.5 M EDTA, 20 μl 1 M Tris-HCl, pH 8.0, 2 μl proteinase K (10 mg/ml), vortexing and incubation at 45C for 1 hr. DNA was extracted using ChIP DNA Clean & Concentrator (Zymo, # D5201 & D5205), After elution from the column in 50 μl of elution buffer, the DNA concentration was determine using 2 μl samples for Bioanalyzer analysis. The two input samples ranged from 16-20 ng/μl and the eight immunoprecipitated samples ranged from 0.1-0.2ng/μl (5-10 ng per 100 trunks). For ChIP-seq, 2 ng was used per sample to prepare libraries for DNA sequencing.

ChIP-seq genomic sequencing and bioinformatic analysis

Libraries for DNA sequencing were prepared according to the instructions accompanying the NEBNext DNA Ultra II kit (catalog # E7645S; New England Biolabs, Inc). Libraries were sequenced on the NextSeq 500 following the manufacturer's protocols, generating 40 million reads per sample with single read lengths of 75 bp. Adapter remnants of sequencing reads were removed using cutadapt v1.18 [71]. ChIP-Seq sequencing reads were aligned using STAR aligner version 2.7 to Mouse genome version 38 [72]. Homer v4.10 [73] was used to call peaks from ChIP-Seq samples by comparing the ChIP samples with matching input samples. Homer v4.10 was used to annotate peaks to mouse genes, and quantify reads count to peaks. The raw reads count for different peaks were compared using DESeq2 [74]. P values from DESeq2 were corrected using the Benjamini & Hochberg (BH) method for multiple testing errors [75]. Peaks with BH corrected p value <0.05 (BHP<0.05) were selected as significantly differentially marked peaks. Transcription factor binding

sites motif enrichment analyses were performed using Homer v4.10 [73] to analyze the significant RA-regulated ChIP-seq peaks; DR1 RAREs were found by searching for TR4(NR),DR1; DR2 RAREs by Reverb(NR),DR2; and DR5 RAREs by RAR:RXR(NR),DR5. Evolutionary conservation of RAREs was performed via DNA sequence homology searches using the UCSC genome browser software. Ingenuity Pathway Analysis (IPA) was used to identify pathways for our list of target genes; from IPA results, heatmaps were designed with Prism software and associated networks were created using STRING software. High throughput DNA sequencing was performed in the Sanford Burnham Prebys Genomics Core and bioinformatics analysis was performed in the Sanford Burnham Prebys Bioinformatics Core.

Generation of mutant embryos by CRISPR/Cas9 mutagenesis

CRISPR/Cas9 gene editing was performed using methods similar to those previously described by others [76, 77] and by our laboratory [7]. Single-guide RNAs (sgRNAs) were generated that target exons to generate frameshift null mutations, with two sgRNAs used together for each gene. sgRNAs were designed with maximum specificity using the tool at crispr.mit.edu to ensure that each sgRNA had no more than 17 out of 20 matches with any other site in the mouse genome and that those sites are not located within exons of other genes. DNA templates for sgRNAs were generated by PCR amplification (Phusion DNA Polymerase; New England Biolabs) of ssDNA oligonucleotides (purchased from Integrated DNA Technologies) containing on the 5' end a minimal T7 promoter, then a 20 nucleotide sgRNA target sequence (underlined below), and finally the tracrRNA sequence utilized by Cas9 on the 3' end, shown as follows:

5'-GCGTAATACGACTCACTATAGG<u>NNNNNNNNNNNNNNNNNNNN</u>GTTTTAGAGCTAGAAATA

GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT-3'

The 20 nucleotide target sequences used were as follows:

Nf2f1 exon 2 (#1) TTTTTATCAGCGGTTCAGCG

Nf2f1 exon 2 (#2) GGTCCATGAAGGCCACGACG

Nf2f2 exon 2 (#1) GGTACGAGTGGCAGTTGAGG

Nf2f2 exon 2 (#2) CGCCGAGTATAGCTGCCTCA

Meis1 exon 2 (#1) CGACGACCTACCCCATTATG

Meis1 exon 2 (#2) TGACCGAGGAACCCATGCTG

Meis2 exon 2 (#1) GATGAGCTGCCCCATTACGG

Meis2 exon 2 (#2) CGACGCCTTGAAAAGAGACA

sgRNAs were then transcribed from templates using HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) and purified using Megaclear Kit (Life Technologies). sgRNAs were tested in vitro for their cleavage ability in combination with Cas9 nuclease (New England Biolabs); briefly, genomic regions flanking the target sites were PCR amplified, then 100 ng was incubated with 30 nM Cas9 nuclease and 30 ng sgRNA in 30 µl for 1 hour at 37˚C, followed by analysis for cleavage by gel electrophoresis.

For injection into mouse embryos, a solution containing 50 ng/µl Cas9 mRNA (Life Technologies) and 20 ng/µl for each sgRNA used was prepared in nuclease free water. Fertilized oocytes were collected from 3-4 week-old superovulated C57Bl6 females prepared by injecting 5 IU each of pregnant mare serum gonadotrophin (PMSG) (Sigma Aldrich) and human chorionic gonadotropin (hCG) (Sigma Aldrich). Fertilized oocytes were then transferred into M2 medium (Millipore) and injected with the Cas9 mRNA/sgRNA solution into the cytoplasm. Injected embryos were cultured in KSOMaa medium (Zenith) in a humidified atmosphere with 5% CO_2 at 37˚C overnight to maximize the time for CRISPR/Cas9 gene editing to occur at the 1-cell stage, then re-implanted at the 2-cell stage into recipient pseudo-pregnant ICR female mice. Implanted females were sacrificed to obtain F0 E9.0 embryos (*Nr2f1/Nr2f2*) or F0 E10.5 embryos (*Meis1/Meis2*). As fertilized mouse oocytes spend a long time at the 1-cell and 2-cell stages, this facilitates CRISPR/Cas9 gene editing at early stages and allows many F0 embryos to be examined for mutant phenotypes [7]. For genotyping, yolk sac DNA was collected and PCR products were generated using primers flanking the sgRNA target sites; PCR products were subjected to DNA sequence analysis from both directions using either upstream or downstream primers. For each gene analyzed, embryos were classified as heterozygous (het) if the DNA sequence contained both a wild-type allele and a frame-shift allele; embryos were classified as homozygous (hom) if only frame-shift alleles were detected but no wild-type sequence.

In situ gene expression analysis

Embryos were fixed in paraformaldehyde at 4˚C overnight, dehydrated into methanol, and stored at -20˚C. Detection of mRNA was performed by whole mount in situ hybridization as previously described [78].

Acknowledgments

We thank the Genomics Core Facility and Bioinformatics Core Facility at SBP Medical Discovery Institute for help with ChIP-seq and RNA-seq analysis. We thank the Animal Resources Core Facility at SBP Medical Discovery Institute for conducting timed-matings to generate mouse embryos. Research reported in this publication was supported by NIAMS of the National Institutes of Health under award number R01AR067731 (G.D.). 100%/$2,145,500 of the total project costs were financed with Federal funding. 0%/$0 of the total costs were financed with non-Federal funding. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

M.B, K.F.M., and G.D. designed the study and performed the experiments. M.B., J.Y., and G.D. analyzed the data and wrote the paper.

Competing financial interests:

The authors declare no competing financial interests.

Data availability

RNA-seq data have been deposited in GEO under accession number GSE131584. ChIP-seq data have been deposited in GEO under accession number GSE131624.

References

1. Sandell LL, Sanderson BW, Moiseyev G, Johnson T, Mushegian A, Young K, et al. RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ

development. Genes Dev. 2007;21(9):1113-24. PubMed PMID: 17473173; PubMed Central PMCID: PMCPMCID: PMC1855236
2. Niederreither K, Subbarayan V, Dollé P, Chambon P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nature Genet. 1999;21:444-8.
3. Mic FA, Haselbeck RJ, Cuenca AE, Duester G. Novel retinoic acid generating activities in the neural tube and heart identified by conditional rescue of *Raldh2* null mutant mice. Development. 2002;129:2271-82. PubMed Central PMCID: PMCPMCID: PMC2833017.
4. Rhinn M, Dolle P. Retinoic acid signalling during development. Development. 2012;139(5):843-58. PubMed PMID: 22318625.
5. Cunningham TJ, Duester G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. Nature Rev Mol Cell Biol. 2015;16:110-23. PubMed Central PMCID: PMCPMCID: PMC4636111.
6. Mark M, Ghyselinck NB, Chambon P. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. Ann Rev PharmacolToxicol. 2006;46:451-80. PubMed PMID: 16402912.
7. Kumar S, Cunningham TJ, Duester G. Nuclear receptor corepressors *Ncor1* and *Ncor2* (*Smrt*) are required for retinoic acid-dependent repression of *Fgf8* during somitogenesis. Dev Biol. 2016;418:204-15. PubMed Central PMCID: PMCPMCID: PMC5031541.
8. Moutier E, Ye T, Choukrallah MA, Urban S, Osz J, Chatagnon A, et al. Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology. J Biol Chem. 2012;287(31):26328-41. PubMed PMID: 22661711; PubMed Central PMCID: PMCPMCID: PMC3406717.
9. Lalevee S, Anno YN, Chatagnon A, Samarut E, Poch O, Laudet V, et al. Genome-wide in Silico Identification of New Conserved and Functional Retinoic Acid Receptor Response Elements (Direct Repeats Separated by 5 bp). J Biol Chem. 2011;286(38):33322-34. PubMed PMID: 21803772; PubMed Central PMCID: PMCPMCID: PMC3190930.
10. Dupé V, Davenne M, Brocard J, Dollé P, Mark M, Dierich A, et al. In vivo functional analysis of the *Hoxa-1* 3' retinoic acid response element (3'RARE). Development. 1997;124:399-410.
11. Houle M, Sylvestre JR, Lohnes D. Retinoic acid regulates a subset of Cdx1 function in vivo. Development. 2003;130:6555-67.
12. Nishimoto S, Wilde SM, Wood S, Logan MP. RA Acts in a Coherent Feed-Forward Mechanism with Tbx5 to Control Limb Bud Induction and Initiation. Cell Reports. 2015;12(5):879-91. PubMed PMID: 26212321; PubMed Central PMCID: PMCPMCID: PMC4553633.
13. Cunningham TJ, Lancman JJ, Berenguer M, Dong PDS, Duester G. Genomic knockout of two presumed forelimb Tbx5 enhancers reveals they are nonessential for limb development. Cell Reports. 2018;23:3146-51. PubMed Central PMCID: PMCPMCID: PMC6034701.
14. Duester G. Knocking out enhancers to enhance epigenetic research. Trends Genet. 2019;35(2):89. PubMed PMID: 30466730; PubMed Central PMCID: PMCPMCID: PMC6338484.
15. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011;470(7333):279-83. PubMed PMID: 21160473; PubMed Central PMCID: PMCPMCID: PMC4445674.
16. Laugesen A, Helin K. Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell. 2014;14(6):735-51. PubMed PMID: 24905164.
17. Ribes V, Le Roux I, Rhinn M, Schuhbaur B, Dolle P. Early mouse caudal development relies on crosstalk between retinoic acid, Shh and Fgf signalling pathways. Development. 2009;136(4):665-76. PubMed PMID: 19168680.
18. Cunningham TJ, Brade T, Sandell LL, Lewandoski M, Trainor PA, Colas A, et al. Retinoic acid activity in undifferentiated neural progenitors Is sufficient to fulfill Its role in restricting Fgf8 expression for somitogenesis. PLOS ONE. 2015;10(9):e0137894. PubMed PMID: 26368825; PubMed Central PMCID: PMCPMCID: PMC4569375.

19. Cunningham TJ, Colas A, Duester G. Early molecular events during retinoic acid induced differentiation of neuromesodermal progenitors. Biology Open. 2016;5:1821-33. PubMed Central PMCID: PMCPMCID: PMC5200905.
20. Niederreither K, Dolle P. Retinoic acid in development: towards an integrated view. Nature Rev Genet. 2008;9(7):541-53. PubMed PMID: 18542081.
21. Kumar S, Duester G. Retinoic acid controls body axis extension by directly repressing *Fgf8* transcription. Development. 2014;141:2972-7. PubMed Central PMCID: PMCPMCID: PMC4197666.
22. Mendelsohn C, Ruberte E, LeMeur M, Morriss-Kay G, Chambon P. Developmental analysis of the retinoic acid-inducible RAR-b2 promoter in transgenic animals. Development. 1991;113:723-34.
23. Durand B, Saunders M, Leroy P, Leid M, Chambon P. All-trans and 9-cis retinoic acid induction of CRABPII transcription is mediated by RAR-RXR heterodimers bound to DR1 and DR2 repeated motifs. Cell. 1992;71:73-85.
24. Gaunt SJ, Paul YL. Origins of Cdx1 regulatory elements suggest roles in vertebrate evolution. Int J Dev Biol. 2011;55(1):93-8. PubMed PMID: 21425084.
25. Edri S, Hayward P, Jawaid W, Martinez Arias A. Neuro-mesodermal progenitors (NMPs): a comparative study between pluripotent stem cells and embryo-derived populations. Development. 2019;146(12):24. PubMed PMID: 31152001.
26. Wilson V, Olivera-Martinez I, Storey KG. Stem cells, signals and vertebrate body axis extension. Development. 2009;136(10):1591-604. PubMed PMID: 19395637.
27. Kondoh H, Takemoto T. Axial stem cells deriving both posterior neural and mesodermal tissues during gastrulation. Curr Opin Genet Dev. 2012;22(4):374-80. PubMed PMID: 22575683.
28. Henrique D, Abranches E, Verrier L, Storey KG. Neuromesodermal progenitors and the making of the spinal cord. Development. 2015;142(17):2864-75. PubMed PMID: 26329597; PubMed Central PMCID: PMCPMCID: PMC4958456.
29. Kimelman D. Tales of tails (and trunks): forming the posterior body in vertebrate embryos. Curr Top Dev Biol. 2016;116:517-36. PubMed PMID: 26970638; PubMed Central PMCID: PMCPMCID: PMC4883064.
30. Koch F, Scholze M, Wittler L, Schifferl D, Sudheer S, Grote P, et al. Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. Dev Cell. 2017;42(5):514-26. PubMed PMID: 28826820.
31. Gouti M, Delile J, Stamataki D, Wymeersch FJ, Huang Y, Kleinjung J, et al. A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. Dev Cell. 2017;41(3):243-61. PubMed PMID: 28457792; PubMed Central PMCID: PMCPMCID: PMC5425255.
32. Amin S, Neijts R, Simmini S, van Rooijen C, Tan SC, Kester L, et al. Cdx and T brachyury co-activate growth signaling in the embryonic axial progenitor niche. Cell Reports. 2016;17(12):3165-77. PubMed PMID: 28009287.
33. Olivera-Martinez I, Harada H, Halley PA, Storey KG. Loss of FGF-Dependent Mesoderm Identity and Rise of Endogenous Retinoid Signalling Determine Cessation of Body Axis Elongation. PLOS Biology. 2012;10(10):e1001415. PubMed PMID: 23118616; PubMed Central PMCID: PMCPMCID: PMC3484059.
34. Tsakiridis A, Huang Y, Blin G, Skylaki S, Wymeersch F, Osorno R, et al. Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. Development. 2014;141(6):1209-21. PubMed PMID: 24595287; PubMed Central PMCID: PMCPMCID: PMC3943179.
35. Garriock RJ, Chalamalasetty RB, Kennedy MW, Canizales LC, Lewandoski M, Yamaguchi TP. Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. Development. 2015;142(9):1628-38. PubMed PMID: 25922526; PubMed Central PMCID: PMCPMCID: PMC4419273.

36. Martin BL, Kimelman D. Canonical Wnt Signaling Dynamically Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation. Dev Cell. 2012;22(1):223-32. PubMed PMID: 22264734; PubMed Central PMCID: PMCPMCID: PMC3465166.
37. Naiche LA, Holder N, Lewandoski M. FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. Proc Natl Acad Sci USA. 2011;108:4018-23. PubMed Central PMCID: PMCPMCID: PMC3054031.
38. Jurberg AD, Aires R, Novoa A, Rowland JE, Mallo M. Compartment-dependent activities of Wnt3a/beta-catenin signaling during vertebrate axial extension. Dev Biol. 2014;394(2):253-63. PubMed PMID: 25152336.
39. Takemoto T, Uchikawa M, Yoshida M, Bell DM, Lovell-Badge R, Papaioannou VE, et al. Tbx6-dependent *Sox2* regulation determines neural or mesodermal fate in axial stem cells. Nature. 2011;470:394-8. PubMed Central PMCID: PMCPMCID: PMC3042233.
40. Wymeersch FJ, Huang Y, Blin G, Cambray N, Wilkie R, Wong FC, et al. Position-dependent plasticity of distinct progenitor types in the primitive streak. eLife. 2016;5:e10042. PubMed PMID: 26780186; PubMed Central PMCID: PMCPMCID: PMC4798969.
41. Diez del Corral R, Olivera-Martinez I, Goriely A, Gale E, Maden M, Storey K. Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. Neuron. 2003;40:65-79.
42. Patel NS, Rhinn M, Semprich CI, Halley PA, Dolle P, Bickmore WA, et al. FGF Signalling Regulates Chromatin Organisation during Neural Differentiation via Mechanisms that Can Be Uncoupled from Transcription. PLOS Genetics. 2013;9(7):e1003614. PubMed PMID: 23874217.
43. Zhao X, Duester G. Effect of retinoic acid signaling on Wnt/beta-catenin and FGF signaling during body axis extension. Gene Expr Patterns. 2009;9(6):430-5. PubMed PMID: 19539783; PubMed Central PMCID: PMCPMCID: PMC2791796.
44. Oosterveen T, Kurdija S, Enstero M, Uhde CW, Bergsland M, Sandberg M, et al. SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. Proc Natl Acad Sci USA. 2013;110(18):7330-5. PubMed PMID: 23589857.
45. Joshi P, Darr AJ, Skromne I. CDX4 regulates the progression of neural maturation in the spinal cord. Dev Biol. 2019;449(2):132-42. PubMed PMID: 30825428.
46. Scardigli R, Schuurmans C, Gradwohl G, Guillemot F. Crossregulation between Neurogenin2 and pathways specifying neuronal identity in the spinal cord. Neuron. 2001;31(2):203-17. PubMed PMID: 11502253.
47. Feng L, Hernandez RE, Waxman JS, Yelon D, Moens CB. Dhrs3a regulates retinoic acid biosynthesis through a feedback inhibition mechanism. Dev Biol. 2010;338(1):1-14. PubMed PMID: 19874812.
48. Billings SE, Pierzchalski K, Butler Tjaden NE, Pang XY, Trainor PA, Kane MA, et al. The retinaldehyde reductase DHRS3 is essential for preventing the formation of excess retinoic acid during embryonic development. FASEB Journal. 2013;27(12):4877-89. PubMed PMID: 24005908.
49. Ishibashi T, Usami T, Fujie M, Azumi K, Satoh N, Fujiwara S. Oligonucleotide-based microarray analysis of retinoic acid target genes in the protochordate, *Ciona intestinalis*. DevDyn. 2005;233:1571-8.
50. Laursen KB, Mongan NP, Zhuang Y, Ng MM, Benoit YD, Gudas LJ. Polycomb recruitment attenuates retinoic acid-induced transcription of the bivalent NR2F1 gene. Nucleic Acids Res. 2013;41(13):6430-43. PubMed PMID: 23666625.
51. Dohn TE, Ravisankar P, Tirera FT, Martin KE, Gafranek JT, Duong TB, et al. Nr2f-dependent allocation of ventricular cardiomyocyte and pharyngeal muscle progenitors. PLoS Genetics. 2019;15(2):e1007962. PubMed PMID: 30721228.
52. Mercader N, Leonardo E, Piedra ME, Martínez-A C, Ros MA, Torres M. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. Development. 2000;127:3961-70.

53. Kashyap V, Laursen KB, Brenet F, Viale AJ, Scandura JM, Gudas LJ. RARg is essential for retinoic acid induced chromatin remodeling and transcriptional activation in embryonic stem cells. J Cell Sci. 2013;126(Pt 4):999-1008. PubMed PMID: 23264745.
54. Will AJ, Cova G, Osterwalder M, Chan WL, Wittler L, Brieske N, et al. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nature Genet. 2017;49(10):1539-45. PubMed PMID: 28846100; PubMed Central PMCID: PMCPMCID: PMC5617800.
55. Osterwalder M, Barozzi I, Tissieres V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. Nature. 2018;554(7691):239-43. PubMed PMID: 29420474; PubMed Central PMCID: PMCPMCID: PMC5808607.
56. Dickel DE, Ypsilanti AR, Pla R, Zhu Y, Barozzi I, Mannion BJ, et al. Ultraconserved enhancers are required for normal development. Cell. 2018;172(3):491-9. PubMed PMID: 29358049; PubMed Central PMCID: PMCPMCID: PMC5786478.
57. Béland M, Lohnes D. Chicken ovalbumin upstream promoter-transcription factor members repress retinoic acid-induced *Cdx1* expression. J Biol Chem. 2005;280:13858-62.
58. Vilhais-Neto GC, Maruhashi M, Smith KT, Vasseur-Cognet M, Peterson AS, Workman JL, et al. Rere controls retinoic acid signalling and somite bilateral symmetry. Nature. 2010;463(7283):953-7. PubMed PMID: 20164929.
59. Qiu Y, Pereira FA, DeMayo FJ, Lydon JP, Tsai SY, Tsai MJ. Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. Genes Dev. 1997;11:1925-37.
60. Pereira FA, Qiu YH, Zhou G, Tsai MJ, Tsai SY. The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. Genes Dev. 1999;13:1037-49.
61. Hisa T, Spence SE, Rachel RA, Fujita M, Nakamura T, Ward JM, et al. Hematopoietic, angiogenic and eye defects in *Meis1* mutant animals. EMBO Journal. 2004;23:450-9.
62. Machon O, Masek J, Machonova O, Krauss S, Kozmik Z. Meis2 is essential for cranial and cardiac neural crest development. BMC Developmental Biology. 2015;15(1):40. PubMed PMID: 26545946; PubMed Central PMCID: PMCPMCID: PMC4636814.
63. Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. Cell. 2004;116:511-26.
64. Begemann G, Schilling TF, Rauch GJ, Geisler R, Ingham PW. The zebrafish *neckless* mutation reveals a requirement for *raldh2* in mesodermal signals that pattern the hindbrain. Development. 2001;128:3081-94.
65. Berenguer M, Lancman JJ, Cunningham TJ, Dong PDS, Duester G. Mouse but not zebrafish requires retinoic acid for control of neuromesodermal progenitors and body axis extension. Dev Biol. 2018;441:127-31. PubMed Central PMCID: PMCPMCID: PMC6064660.
66. Cooper KL, Hu JK, ten Berge D, Fernandez-Teran M, Ros MA, Tabin CJ. Initiation of proximal-distal patterning in the vertebrate limb by signals and growth. Science. 2011;332(6033):1083-6. PubMed PMID: 21617075.
67. Rosello-Diez A, Ros MA, Torres M. Diffusible signals, not autonomous mechanisms, determine the main proximodistal limb subdivision. Science. 2011;332(6033):1086-8. PubMed PMID: 21617076.
68. Cunningham TJ, Chatzi C, Sandell LL, Trainor PA, Duester G. *Rdh10* mutants deficient in limb field retinoic acid signaling exhibit normal limb patterning but display interdigital webbing. Dev Dyn. 2011;240:1142-50. PubMed Central PMCID: PMCPMCID: PMC3081420.
69. Cunningham TJ, Zhao X, Sandell LL, Evans SM, Trainor PA, Duester G. Antagonism between retinoic acid and fibroblast growth factor signaling during limb development. Cell Reports. 2013;3:1503-11. PubMed Central PMCID: PMCPMCID: PMC3745640.

70. Voss AK, Dixon MP, McLennan T, Kueh AJ, Thomas T. Chromatin immunoprecipitation of mouse embryos. In: Vancura A, editor. Transcriptional Regulation: Methods and Protocols, Methods in Molecular Biology. 809. New York: Humana Press; 2012. p. 335-52.
71. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17.
72. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.
73. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576-89.
74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
75. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 1995;57:289-300.
76. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell. 2013;153(4):910-8. PubMed PMID: 23643243; PubMed Central PMCID: PMCPMCID: PMC3969854.
77. Tan EP, Li Y, Del Castillo Velasco-Herrera M, Yusa K, Bradley A. Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells. Genesis. 2015;53(2):225-36. PubMed PMID: 25378133.
78. Sirbu IO, Duester G. Retinoic acid signaling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. Nature Cell Biol. 2006;8:271-7. PubMed Central PMCID: PMCPMCID: PMC2805411.

Table 1. Comparison of ChIP-seq and RNA-seq for *Aldh1a2*-/- vs wild-type E8.5 trunk tissue showing identification of previously known RA target genes plus a few of the new candidate RA target genes for early trunk development that were further studied here.

A. H3K27ac ChIP-seq vs. RNA-seq

H3K27ac ChIP-seq RA-regulated peak for *Aldh1a2* KO vs WT (mm10)	**log2 fold change: H3K27ac ChIP-seq for *Aldh1a2* KO vs WT**	**RARE: based on Homer TFBS analysis**	**nearby gene with altered expression in *Aldh1a2* KO**	**log2 fold change for nearby gene: RNA-seq for *Aldh1a2* KO vs WT**
chr13:78197222-78204291	-1.23	DR1	Nr2f1 - new	-2.02
chr4:145033496-145035860	-0.65	DR5	Dhrs3	-1.11
chr14:16571405-16576397	-0.63	DR5	Rarb	-1.64
chr11:18962656-18965461	-0.61	DR5, DR1	Meis1 - new	-2.64
chr2:105689278-105690982	-0.58	-	Pax6	-3.02
chr3:87956774-87961235	-0.58	DR2, DR1	Crabp2	-2.82
chr2:116019003-116024272	-0.58	DR2	Meis2 - new	-1.10
chr7:70348715-70369942	-0.57	DR1	Nr2f2 - new	-2.32
chr11:18956989-18958835	-0.57	DR5	Meis1 - new	-2.64
chr11:19012000-19025444	-0.54	DR1	Meis1 - new	-2.64
chr3:34678267-34680699	-0.54	DR2	Sox2	-0.86
chr18:61033064-61036494	-0.52	DR2, DR1	Cdx1	-2.00
chr3:34647848-34655776	-0.51	-	Sox2	-0.86
chr19:45733505-45735997	0.53	DR1	Fgf8	5.24
chr13:114456392-114460659	0.72	DR2	Fst	1.15
chr5:147298587-147311126	0.73	DR2	Cdx2	1.98

B. H3K27me3 ChIP-seq vs. RNA-seq

H3K27me3 ChIP-seq RA-regulated peak for *Aldh1a2* KO vs WT (mm10)	**log2 fold change: H3K27me3 ChIP-seq for *Aldh1a2* KO vs WT**	**RARE: based on Homer TFBS analysis**	**nearby gene with altered expression in *Aldh1a2* KO**	**log2 fold change for nearby gene: RNA-seq for *Aldh1a2* KO vs WT**
chr18:38598986-38601292	-1.20	-	Spry4 - new	3.43
chr5:147297983-147318733	-0.63	DR2	Cdx2	1.98
chr19:45735049-45746658	-0.49	DR2	Fgf8	5.24
chr13:114456076-114460873	-0.47	DR2	Fst	1.15
chr4:144893360-144895562	0.59	-	Dhrs3	-1.11
chr2:116072251-116077455	0.61	DR5	Meis2 - new	-1.10
chr7:70356085-70361002	0.63	DR1	Nr2f2 - new	-2.32
chr6:52156115-52158253	0.73	DR5, DR2	Hoxa1	-5.43
chr11:19015536-19017169	0.78	DR1	Meis1 - new	-2.64
chr11:19007512-19012358	0.87	DR2	Meis1 - new	-2.64
chr14:16574377-16578138	1.02	DR5, DR1	Rarb	-1.64

ChIP-seq values for RA-regulated peaks between *Aldh1a2*-/- (KO) and wild-type (WT) for H3K7ac (log2 <-0.51 or >0.51) and H3K27me3 (log2 <-0.47 or >0.47) with BHP <0.05; a cut-off of log2 <-0.51 or >0.51 for H3K27ac was employed to include a RA-regulated peak near *Sox2* known to be activated by RA; a cut-off of log2 <-0.47 or >0.47 was employed for H3K27me3 to include a RA-regulated peak near *Fst* known to be repressed by RA. RNA-seq values are log2 <-0.85 or >0.85 for differentially expressed genes with FPKM values (KO and WT) >0.5; a cut-off of log2 <-0.85 or >0.85 was employed to include the known RA target gene *Sox2*. RARE, retinoic acid response element; DR1 or DR2 or DR5, direct repeat with 1 or 2 or 5 bp between each repeat; TFBS, transcription factor binding site. Also see related data in Tables S1-S3 (describing all known and new candidate RA target genes) and Figs. S1 and S2.

Table 2. DNA sequences of highly conserved RAREs located in RA-regulated ChIP-seq peaks for H3K27ac or H3K27me3 near all RA-regulated genes in same TAD. RAREs shown here are conserved from mouse to bird, reptile, frog, or fish. RAREs contain no more than one mismatch to Homer consensus DR5, DR2, or DR1 RARE motifs shown here; DR, direct repeat.

RARE MOTIFS: (Homer)	DR5 = RAR:RXR(NR),DR5		DR2 = Reverb(NR),DR2							DR1 = TR4(NR),DR1
Nearest gene with decreased or increased expression in *Aldh1a2* KO	**Other genes in same TAD with decreased or increased expression in *Aldh1a2* KO**	**RARE DNA sequence 5'-3' overall consensus:** AGGTCA N5,N2 AGGTCA G T or N1 G T	**Type of RARE**	**Conserved**						**Genomic coordinates (mm10)**
				rodent	**human**	**bird**	**reptile**	**frog**	**fish**	
RARE ENHANCERS (RA stimulates gain of H3K27ac and/or loss of H3K27me3 near RARE and activates gene in same TAD)										
C1d	none	GGGTCA G GGGTTA	DR1	x	x	x	x	x	x	chr11:18748180-18748192
Clstn1	Lzic, Nmnat1 Kif1b	GGGTCA GA AGGTCA	DR2	x	x	x	x	x		chr4:149907094-149907107
Dach1	none	AGTTCA CACAA AGTTCA	DR5	x	x	x	x	x	x	chr14:98035388-98035404
Dhrs3	none	GGGTCA TTCCA AGTTCA	DR5	x	x	x	x	x		chr4:145034810-145034826
		GGTTCA TCGGG AGGGCA	DR5	x	x	x	x	x		chr4:145034847-145034863
Foxp4	none	GGGTGA C AGGTCA	DR1	x	x	x	x			chr17:47898625-47898637
Hoxa1	Hoxa4, Hoxa9 Skap2	GGTTCA CCGAA AGTTCA	DR5	x	x	x	x	x		chr6:52153426-52153442
		GGTTCA AGAAG AGTTCA	DR5	x	x	x	x	x	x	chr6:52175533-52175549
Meis1	none	AGGCCA CTGAG AGGTCA	DR5	x	x	x	x	x		chr11:18963875-18963891
Meis2	Dph6	AGGTCA AAAAC AGTTCA	DR5	x	x	x	x			chr2:116071242-116071258
Nr2f1	none	GTGTCA A AGTTCA	DR1	x	x	x	x	x	x	chr13:78200425-78200437
Nr2f2	none	GTGTCA A AGTTCA	DR1	x	x	x	x	x	x	chr7:70361772-70361784
Pbx1	Lmx1a	GGGTCG CT GGGTCA	DR2	x	x	x	x			chr1:169238844-169238857
Rarb	none	GGTTCA CCGAA AGTTCA	DR5	x	x	x	x			chr14:16575513-16575529
Sox2	none	GGGTCA GG AGGTCA	DR2	x	x	x	x	x	x	chr3:34679067-34679080
Tshz1	none	GGGTCA TTCAT AGTTCA	DR5	x	x	x	x			chr18:84073476-84073492
		AGGTCA GG AGGTGA	DR2	x	x	x	x			chr18:83839858-83839871
		GGGTGA ACTCA GGTTCA	DR5	x	x	x	x			chr18:83839869-83839885
Zbtb16	none	GGGTCA CA GGGTCA	DR2	x	x	x	x	-	x	chr9:48694721-48694734
		GGGTCA G GGGTTA	DR1	x	x	x	x			chr9:48695827-48695839
Zfhx4	Pex2	GGGTCA GCCTG AGGTCA	DR5	x	x	x	x	x	x	chr3:5388103-5388119
Zfp386	none	GAGTCA A AGGTCA	DR1	x	-	x	x			chr12:117352086-117352098
Zfp638	none	GGTTCA GCCAA AGGTGA	DR5	x	x	x	x	x		chr6:84976840-84976856
RARE SILENCERS (RA stimulates gain of H3K27me3 and/or loss of H3K27ac near RARE and represses gene in same TAD)										
Fgf8	Poll, Btrc Mrpl43, Chuk Sema4g Dnmbp, Erlin1 Entpd7, Got1 Slc25a28	GGGTCA GC AGTTCA	DR2	x	x	x				chr19:45747043-45747056

Fig. 1

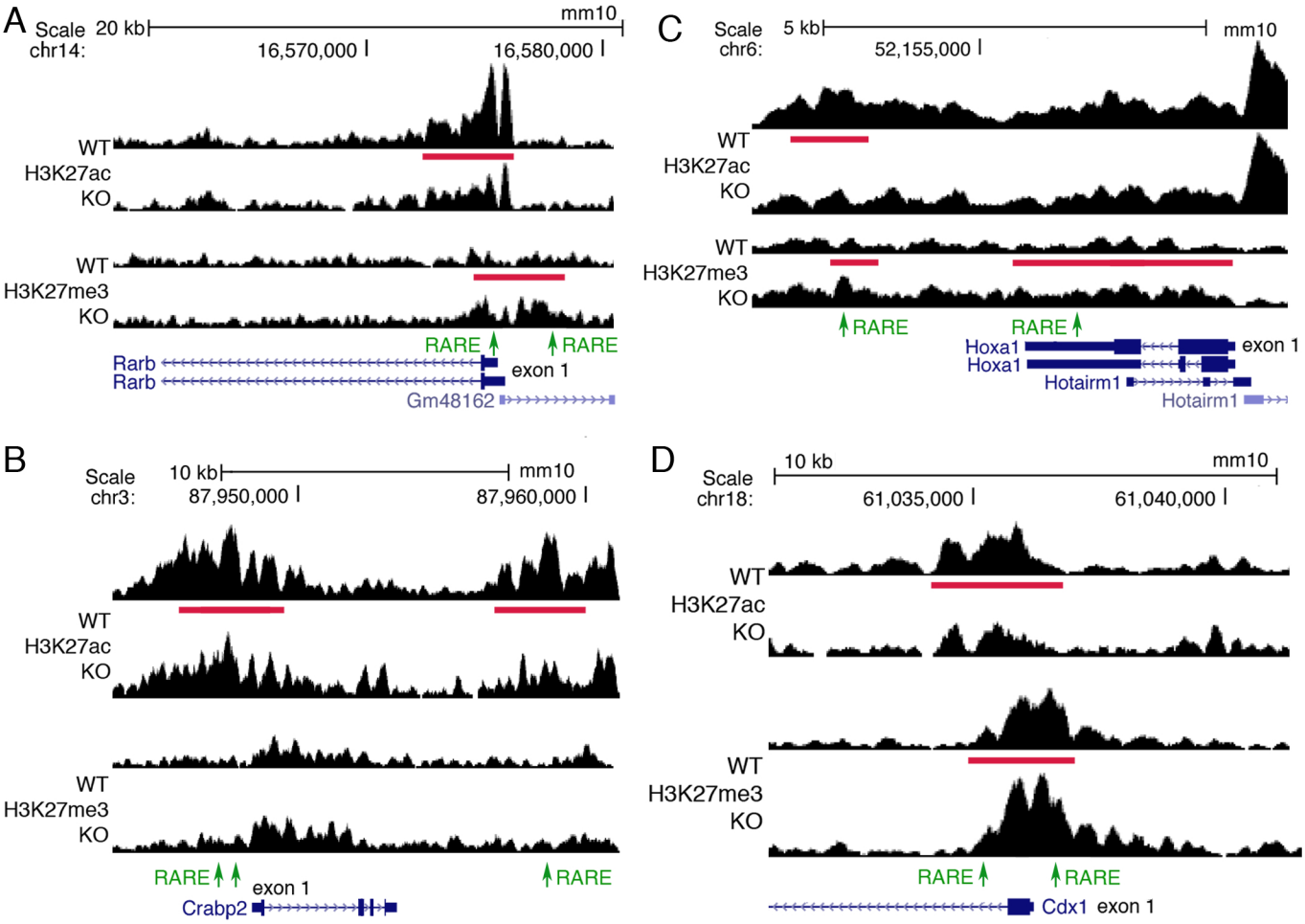


Fig. 1. ChIP-seq findings for *Rarb*, *Crabp2*, *Hoxa1*, and *Cdx1* showing that RA-regulated peaks for H3K27ac and H3K7me3 are located near known RARE enhancers. (A) Shown for *Rarb* are RA-regulated ChIP-seq peaks for H3K27ac and H3K27me3 (red bars) when RA is lost in E8.5 trunk comparing wild-type (WT) vs *Aldh1a2-/-* (KO) as well as RAREs (green). A RARE in the 5'-untranslated region is known to function as an RA-dependent enhancer in mouse transgene studies (ref. 21); here, H3K27ac is decreased and H3K27me3 increased near the native RARE when RA is lost in trunk tissue, supporting its function as a RARE enhancer in vivo. We also found a RARE in the 5'-noncoding region of *Rarb* within an H3K27me3 ChIP-seq peak that is increased when RA is lost. (B) RA-regulated peaks for H3K27ac and RAREs are shown for *Crabp2*. The two RAREs in the 5'-noncoding region were previously shown to function as RA-dependent enhancers in cell line studies (ref. 22). Our epigenetic studies also identified another RARE enhancer in the 3'-noncoding region. (C) RA-regulated peaks for H3K27ac and/or H3K27me3 and RAREs are shown for *Hoxa1*. Knockout studies in mouse embryos have shown that the RARE in the 3'-noncoding region is essential for hindbrain *Hoxa1* expression and development (ref. 10). (D) RA-regulated peaks for H3K27ac and H3K27me3 and RAREs are shown for *Cdx1*. Knockout studies in mouse embryos have shown that the RARE in the 5'-noncoding region is essential for *Cdx1* expression and body axis development (ref. 11). RA-regulated peaks in the genome browser view shown here and elsewhere are for one replicate, with the other replicate showing a similar result.

Fig. 2

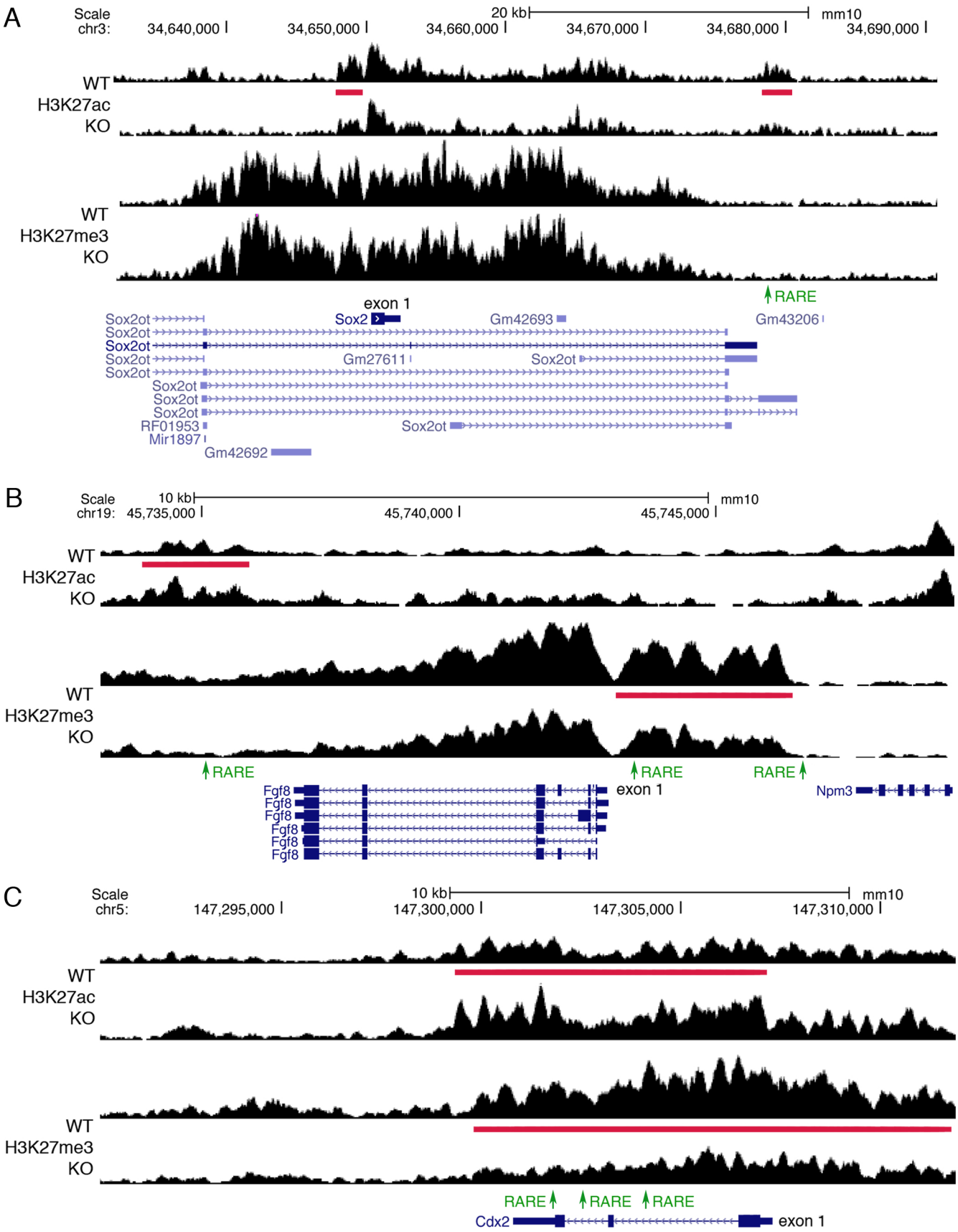


Fig. 2. ChIP-seq findings identify RAREs near *Sox2*, *Fgf8*, and *Cdx2* that regulate NMPs. See below.

Fig. 2. ChIP-seq findings identify RAREs near *Sox2*, *Fgf8*, and *Cdx2* that regulate NMPs.
(A) Two RA-regulated ChIP-seq peaks for H3K27ac (red bars) near *Sox2* are shown for trunk tissue from E8.5 wild-type (WT) vs *Aldh1a2*-/- (KO). A RARE (green) was found in the 3'-noncoding peak (but not the 5'-noncoding peak) suggesting it may function as a RARE enhancer as the H3K27ac peak is decreased when RA is lost. (B) Shown are RA-regulated ChIP-seq peaks for H3K27me3 and H3K27ac near *Fgf8*. In the 5'-noncoding region of *Fgf8* we found two RAREs on either end of the peak for H3K27me3 (repressive mark) that is decreased in KO, indicating they are candidate RARE silencers; the RARE furthest upstream in the 5'-noncoding region at -4.1 kb was shown by knockout studies to function as an RA-dependent RARE silencer required for caudal *Fgf8* repression and somitogenesis (ref. 7). We also found another RARE in the 3'-noncoding region of *Fgf8* that is another candidate for a RARE silencer as it is contained within a RA-regulated peak for H3K27ac (activating mark) that is increased when RA is lost. (C) *Cdx2* has a peak for H3K27ac that is increased and an overlapping peak for H3K27me3 that is decreased, along with three RAREs included within both peaks, indicating that all these RAREs are candidates for RARE silencers.

Fig. 3

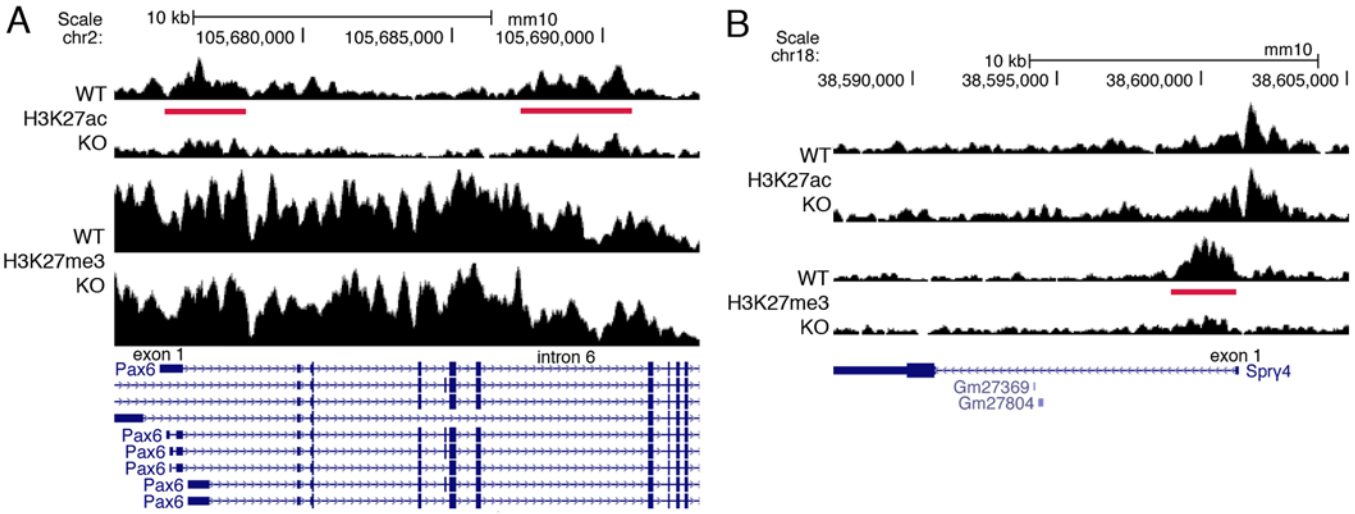


Fig. 3. **ChIP-seq findings for *Pax6* and *Spry4* that lack RARE enhancers or silencers.** These genes are good candidates for being indirect transcriptional targets of RA as their RA-regulated ChIP-seq peaks do not contain RAREs. (A) *Pax6* has two RA-regulated peaks (red bars) for H3K27ac (decreased) when RA is lost in E8.5 trunk tissue from *Aldh1a2*-/- (KO) compared to wild-type (WT); these RA-regulated peaks do not contain RAREs suggesting that transcription of *Pax6* is indirectly activated by RA. (B) *Spry4* has an RA-regulated peak for H3K27me3 (decreased) when RA is lost with no associated RARE suggesting that transcription of *Spry4* is indirectly repressed by RA.

Fig. 4

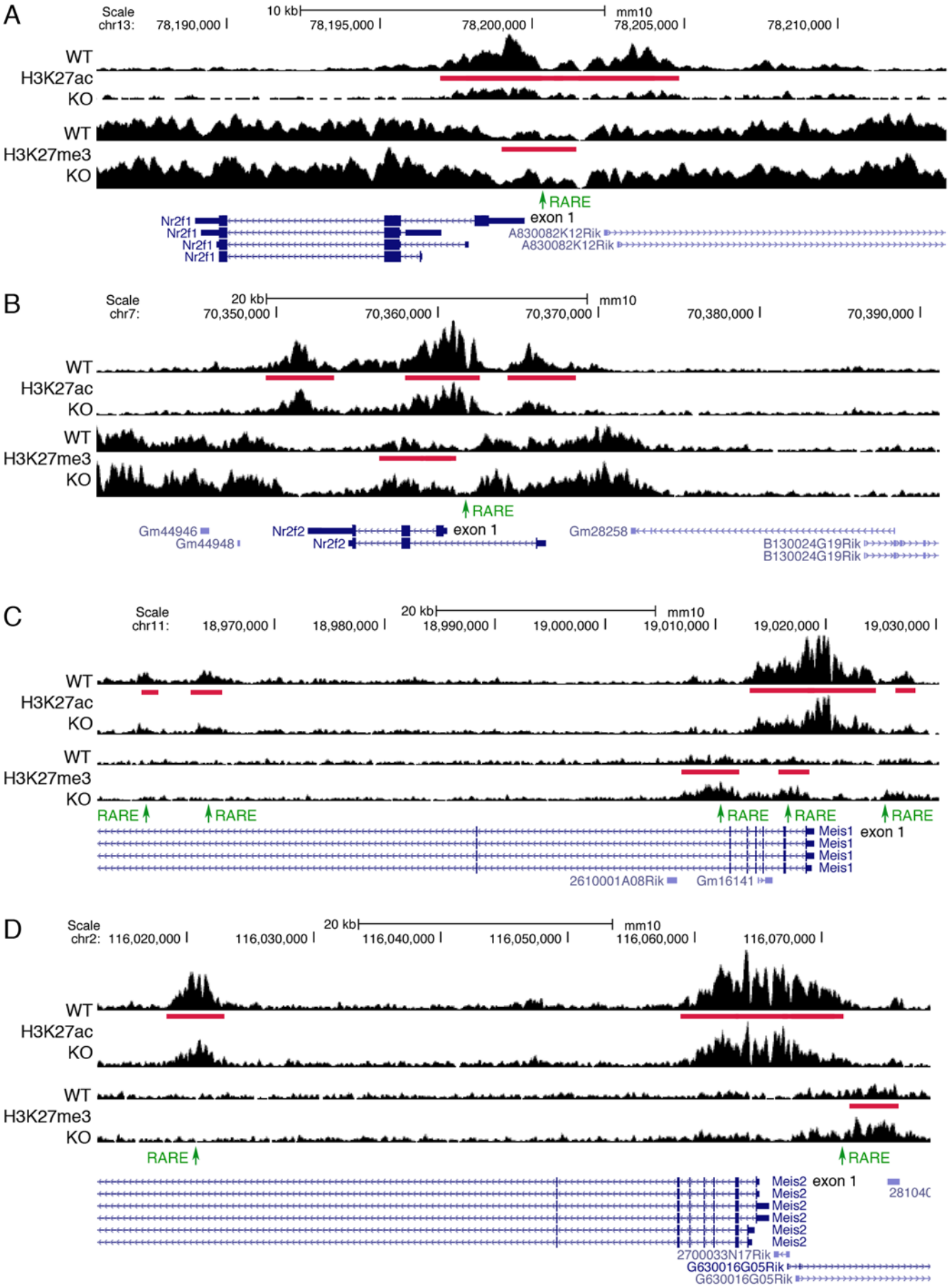


Fig. 4. ChIP-seq findings for *Nr2f1*, *Nr2f2*, *Meis1*, and *Meis2* identify RARE enhancers in gene families.
(continued on next page)

Fig. 4. ChIP-seq findings for *Nr2f1*, *Nr2f2*, *Meis1*, and *Meis2* identify RARE enhancers in gene families. (A-B) *Nr2f1* and *Nr2f2* have differential peaks (red bars) for both H3K27ac (decreased) and H3K27me3 (increased) when RA is lost in E8.5 trunk from *Aldh1a2*-/- (KO) compared to wild-type (WT). Each family member has one RARE (green) contained within these differential peaks that are candidates for RARE enhancers. (C-D) *Meis1* and *Meis2* have differential peaks for both H3K27ac (all decreased) and H3K27me3 (all increased) when RA is lost, along with associated RAREs for each peak that are candidates for RARE enhancers.

Fig. 5

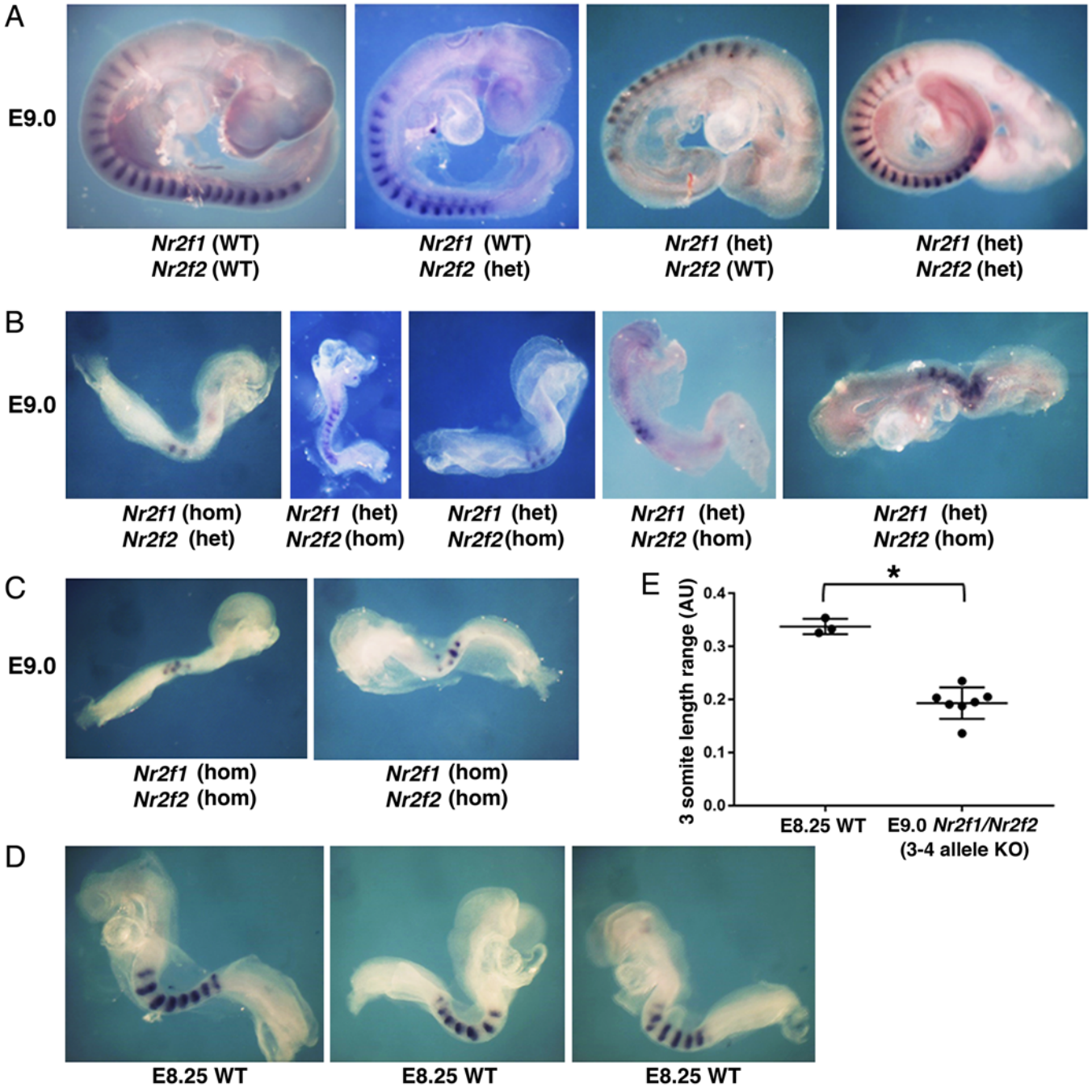
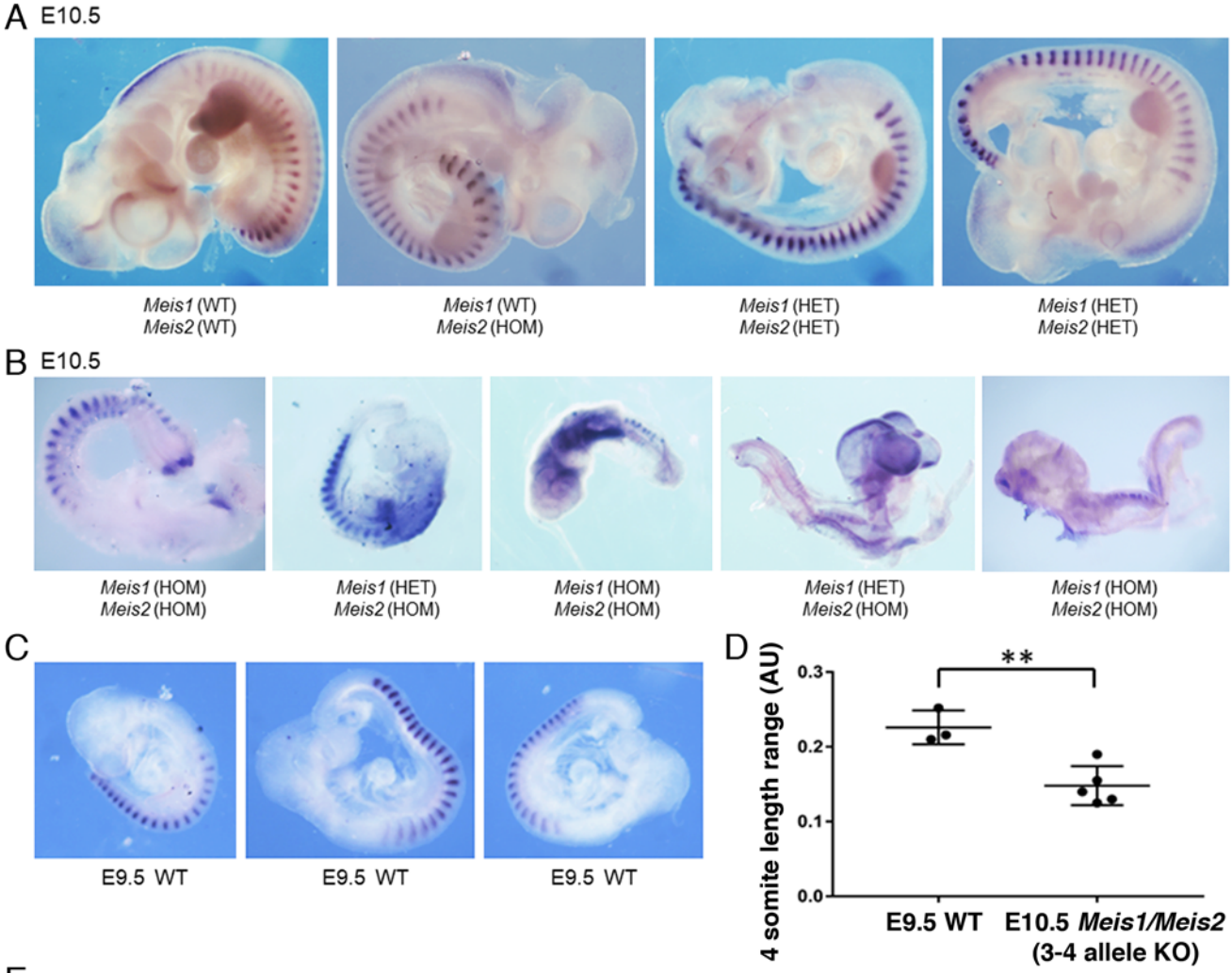
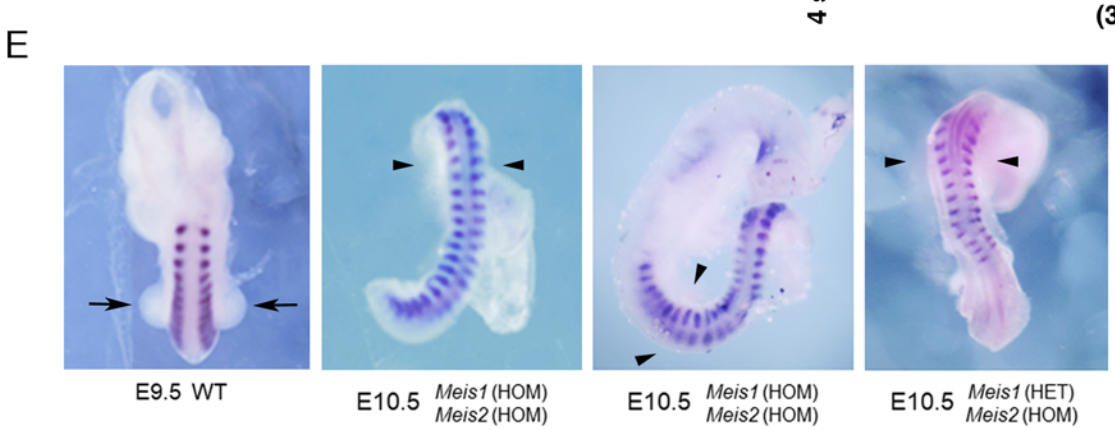


Fig. 5. *Nr2f1/Nr2f2* double mutants exhibit defects in body axis formation. (A) Embryos dissected at E9.0 carrying 0-2 knockout alleles for *Nr2f1* or *Nr2f2* have normal somites and body axis formation based on expression of the somite marker *Uncx*. (B-C) Embryos dissected at E9.0 and stained for *Uncx* that carry 3 or 4 knockout alleles for *Nr2f1* or *Nr2f2* exhibit small somites and reduced body axis growth resembling the size of embryos at E8.25. (D) Wild-type (WT) E8.25 embryos stained for *Uncx* expression. (E) Comparison of somite size along the anteroposterior axis between E8.25 WT and E9.0 *Nr2f1/Nr2f2* knockout embryos (3-4 knockout alleles); *, $p < 0.05$, data expressed as mean ± SD, one-way ANOVA (non-parametric test); WT, $n = 3$ biological replicates; *Nr2f1/Nr2f2* 3-4 allele double knockout, $n = 7$ biological replicates.

Fig. 6


A E10.5
Meis1 (WT)
Meis2 (WT)
Meis1 (WT)
Meis2 (HOM)
Meis1 (HET)
Meis2 (HET)
Meis1 (HET)
Meis2 (HET)
B E10.5
Meis1 (HOM)
Meis2 (HOM)
Meis1 (HET)
Meis2 (HOM)
Meis1 (HOM)
Meis2 (HOM)
Meis1 (HET)
Meis2 (HOM)
Meis1 (HOM)
Meis2 (HOM)
C
E9.5 WT
E9.5 WT
E9.5 WT
D
4 somite length range (AU)
0.3
0.2
0.1
0.0
**
E9.5 WT
E10.5 Meis1/Meis2
(3-4 allele KO)


E
E9.5 WT
E10.5 Meis1 (HOM) Meis2 (HOM)
E10.5 Meis1 (HOM) Meis2 (HOM)
E10.5 Meis1 (HET) Meis2 (HOM)

Supplemental Table S1. Comparison of *Aldh1a2-/-* and wild-type E8.5 trunk tissue for H3K27ac ChIP-seq and RNA-seq results to identify RA-regulated H3K27ac ChIP-seq peaks near genes with RA-regulated expression.

H3K27ac ChIP-seq differential peak for *Aldh1a2* KO vs WT (mm10)	**log2 fold change: H3K27ac ChIP-seq for *Aldh1a2* KO vs WT**	**RARE: based on Homer TFBS analysis**	**nearby gene with altered expression in *Aldh1a2* KO vs WT**	**log2 fold change for nearby gene: RNA-seq for *Aldh1a2* KO vs WT**
chr13:78197222-78204291	-1.23	DR1	Nr2f1	-2.02
chr13:34133342-34134366	-1.13	-	Tubb2b	-1.03
chr18:83838984-83841358	-1.10	DR5, DR2	Tshz1	-1.33
chr11:68049983-68051745	-1.07	DR2, DR1	Stx8	-2.87
chr6:84975748-84978033	-1.03	DR5	Zfp638	-1.33
chr8:14991303-14993068	-0.96	-	Arhgef10	-1.56
chr11:68088685-68091257	-0.90	-	Stx8	-2.87
chr17:56468893-56471601	-0.90	DR2, DR1	Ptprs *	-3.31
chr9:63715651-63717278	-0.89	DR5	Smad3	-2.09
chr19:21164673-21168435	-0.86	DR2	Zfand5	-2.37
chr9:16152232-16154557	-0.86	-	Fat3	-3.02
chr9:118653618-118656175	-0.84	DR5, DR1	Itga9	-1.63
chr6:5028587-5031108	-0.83	DR5, DR2	Ppp1r9a	-1.47
chr9:114798509-114801621	-0.81	-	Cmtm8	-2.12
chr4:149254383-149256741	-0.80	-	Kif1b	-1.72
chr5:36373121-36374553	-0.79	DR5	Sorcs2	-5.20
chr3:5387155-5389128	-0.79	DR5	Zfhx4 *	-2.26
chr17:30467016-30471968	-0.78	DR1	Btbd9	-1.18
chr3:5235978-5239655	-0.77	-	Zfhx4 *	-2.26
chr9:35442832-35445303	-0.74	-	Cdon	-1.60
chr14:98034746-98040239	-0.71	DR5, DR1	Dach1	-2.98
chr6:14897860-14903160	-0.67	DR1	Foxp2	-1.26
chr16:44529549-44532004	-0.67	-	Boc	-0.87
chr14:52330133-52333035	-0.67	-	Sall2	-1.17
chr4:145033496-145035860	-0.65	DR5	Dhrs3 *	-1.11
chr12:8912431-8914944	-0.64	-	Laptm4a	-1.97
chr18:83927674-83929824	-0.64	-	Tshz1	-1.33
chr14:14069903-14072065	-0.64	-	Atxn7	-1.50
chr5:111242680-111244777	-0.63	DR1	Ttc28	-1.87
chr17:66410064-66414806	-0.63	DR5, DR1	Mtcl1	-1.32
chr4:148106909-148109328	-0.63	-	Draxin	-3.43
chr14:16571405-16576397	-0.63	DR5	Rarb *	-1.64
chr16:74395975-74399535	-0.63	DR1	Robo2	-1.22
chr18:84069541-84075594	-0.62	DR5	Tshz1	-1.33
chr11:18962656-18965461	-0.61	DR5, DR1	Meis1 *	-2.64
chr9:96989027-96991630	-0.60	DR5	Spsb4	-2.23
chr3:108409804-108412280	-0.60	-	Celsr2	-7.29
chr17:47897351-47899993	-0.59	DR1	Foxp4 *	-1.02
chr14:78848672-78851835	-0.58	DR5, DR2, DR1	Vwa8	-1.08
chr2:105689278-105690982	-0.58	-	Pax6	-3.02
chr3:87956774-87961235	-0.58	DR2, DR1	Crabp2	-2.82
chr2:116019003-116024272	-0.58	DR2	Meis2 *	-1.10
chr4:107834342-107836683	-0.58	DR2	Lrp8	-1.54
chr7:70348715-70369942	-0.57	DR1	Nr2f2 *	-2.32
chr11:18956989-18958835	-0.57	DR5	Meis1 *	-2.64
chr13:34129519-34132640	-0.55	DR2	Tubb2b	-1.03
chr11:19012000-19025444	-0.54	DR1	Meis1 *	-2.64
chr9:96956410-96959728	-0.54	DR2	Spsb4	-2.23
chr9:48692264-48699040	-0.54	DR2, DR1	Zbtb16	1.36
chr3:34678267-34680699	-0.54	DR2	Sox2	-0.86
chr6:144250107-144252835	-0.52	-	Sox5	-2.33

chr18:61033064-61036494	-0.52	DR2, DR1	Cdx1	-2.00
chr3:34647848-34655776	-0.51	-	Sox2	-0.86
chr17:56475307-56476820	-0.51	-	Ptprs *	-3.30
chr7:133035031-133040386	-0.51	DR2	Ctbp2	-3.06
chr18:53463017-53465407	0.53	-	Prdm6	0.93
chr19:45733505-45735997	0.53	DR1	Fgf8 *	5.24
chr11:54891361-54894784	0.57	DR1	Gpx3	2.58
chr4:86669294-86671201	0.59	-	Plin2	2.29
chr3:127457093-127465482	0.62	-	Ank2	2.26
chr18:60492434-60494743	0.62	-	Smim3	1.25
chr10:17704835-17706602	0.65	DR1	Cited2	2.13
chr11:57831171-57833707	0.66	-	Hand1	1.46
chr6:52310576-52314619	0.71	DR1	Evx1	1.54
chr13:114456392-114460659	0.72	DR2	Fst *	1.15
chr5:147298587-147311126	0.73	DR2	Cdx2 *	1.98
chr5:53106977-53110254	0.74	DR1	Sel1l3	3.67
chr10:59957002-59959223	0.75	-	Ddit4	3.33
chr5:107216257-107218023	0.77	-	Tgfbr3	1.37
chr5:104021158-104022631	1.000	-	Hsd17b11	3.89
chr1:118647742-118649473	1.11	DR5	Tfcp2l1	2.77

ChIP-seq values represent differentially marked H3K27ac peaks comparing *Aldh1a2*-/- (KO) and wild-type (WT) with BHP <0.05; a cut-off of log2 <-0.51 or >0.51 was employed to include a differential peak near *Sox2* known to be activated by RA. RNA-seq values represent differentially expressed genes comparing KO and WT in which FPKM >0.5; a cut-off of log2 <-0.85 or >0.85 was employed to include *Sox2* known to be activated by RA. Genes that have differential peaks for both H3K27ac and H3K27me3 (Table S2) are marked with an asterisk. RARE, retinoic acid response element; DR1 or DR2 or DR5, direct repeat with 1 or 2 or 5 bp between each repeat; TFBS, transcription factor binding site.

Supplemental Table S2. Comparison of *Aldh1a2-/-* and wild-type E8.5 trunk tissue for H3K27me3 ChIP-seq and RNA-seq results to identify RA-regulated H3K27me3 ChIP-seq peaks near genes with RA-regulated expression.

H3K27me3 ChIP-seq differential peak for *Aldh1a2* KO vs WT (mm10)	**log2 fold change: H3K27me3 ChIP-seq for *Aldh1a2* KO vs WT**	**RARE: based on Homer TFBS analysis**	**nearby gene with altered expression in *Aldh1a2* KO vs WT**	**log2 fold change for nearby gene: RNA-seq for *Aldh1a2* KO vs WT**
chr18:38598986-38601292	-1.20	-	Spry4	3.43
chr17:15533901-15538178	-0.95	-	Pdcd2	2.22
chr17:15533901-15538178	-0.95	-	Tbp	2.94
chrX:104569467-104572914	-0.90	-	Zdhhc15	1.5
chrX:104546302-104549188	-0.89	-	Zdhhc15	1.5
chr2:118901989-118904591	-0.85	-	Bahd1	1.12
chr4:129226495-129228276	-0.85	-	C77080	0.98
chr4:129221927-129223300	-0.77	-	C77080	0.98
chr5:15980910-15984533	-0.75	-	Cacna2d1	1.01
chr3:89278454-89282169	-0.72	-	Efna1	1.23
chr11:103110581-103113995	-0.67	-	Acbd4	6.49
chr4:129246599-129252553	-0.64	DR5	C77080	0.98
chr10:21991375-21993681	-0.64	-	Sgk1	1.38
chr6:125360514-125365732	-0.63	DR2	Tnfrsf1a	1.39
chr12:54201904-54203715	-0.63	-	Egln3	3.46
chr5:147297983-147318733	-0.63	DR2	Cdx2 *	1.98
chrX:104536138-104539780	-0.62	DR2	Zdhhc15	1.50
chr4:98726175-98729089	-0.61	-	L1td1	1.50
chr17:29080591-29082455	-0.59	-	Trp53cor1	1.91
chr11:117780323-117784425	-0.58	DR2	Tmc8	1.33
chr6:122800166-122804076	-0.54	-	Slc2a3	1.35
chr10:60828600-60835032	-0.49	-	Unc5b	3.58
chr19:45735049-45746658	-0.49	DR2	Fgf8 *	5.24
chrX:94129800-94133754	-0.48	-	Zfx	1.40
chr2:119235078-119238967	-0.47	-	Spint1	2.74
chr13:114456076-114460873	-0.47	DR2	Fst *	1.15
chr19:11816637-11820013	-0.47	-	Stx3	0.90
chr6:72232803-72239355	0.50	-	Atoh8	-2.59
chr19:4709301-4715701	0.51	-	Sptbn2	-1.59
chr3:5219339-5224436	0.52	-	Zfhx4 *	-2.26
chr7:130260543-130263682	0.56	-	Fgfr2	-3.06
chr17:56471489-56479605	0.57	DR1	Ptprs *	-3.31
chr4:144893360-144895562	0.59	-	Dhrs3	-1.11
chr2:116072251-116077455	0.61	DR5	Meis2 *	-1.10
chr7:70356085-70361002	0.63	DR1	Nr2f2 *	-2.32
chr15:98621217-98623590	0.65	-	Cacnb3	-2.73
chr1:59473538-59476300	0.66	DR1	Fzd7	-1.64
chr17:47915056-47917877	0.67	-	Foxp4 *	-1.02
chr10:8548515-8549917	0.72	-	Ust	-1.40
chr6:52156115-52158253	0.73	DR5, DR2	Hoxa1	-5.43
chr7:96211108-96212622	0.75	-	Tenm4	-2.65
chrX:162887815-162889313	0.76	-	Syap1	-1.55

chr11:19015536-19017169	0.78	DR1	Meis1 *	-2.64
chr18:58208120-58210286	0.81	-	Fbn2	-1.65
chr14:21983733-21987831	0.85	-	Zfp503	-2.68
chr11:19007512-19012358	0.87	DR2	Meis1 *	-2.64
chr14:16574377-16578138	1.02	DR5, DR1	Rarb *	-1.64

ChIP-seq values represent differentially marked H3K27me3 peaks comparing *Aldh1a2*-/- (KO) and wild-type (WT) with BHP <0.05; a cut-off of log2 <-0.47 or >0.47 was employed to include a differential peak near *Fst* known to be repressed by RA. RNA-seq values represent differentially expressed genes comparing KO and WT in which FPKM >0.5; a cut-off of log2 <-0.85 or >0.85 was employed to include the known RA target gene *Sox2*. Genes that have differential peaks for both H3K27me3 and H3K27ac (Table S1) are marked with an asterisk. RARE, retinoic acid response element; DR1 or DR2 or DR5, direct repeat with 1 or 2 or 5 bp between each repeat; TFBS, transcription factor binding site.

Supplemental Table S3. DNA sequences of RAREs located in RA-regulated ChIP-seq peaks for H3K27ac or H3K27me3 near all RA-regulated genes in same TAD. RAREs contain no more than two mismatches to Homer consensus DR5, DR2, or DR1 RARE motifs shown here; DR, direct repeat.

RARE MOTIFS: (Homer)	DR5 = RAR:RXR(NR),DR5		DR2 = Reverb(NR),DR2							DR1 = TR4(NR),DR1
Nearest gene with decreased or increased expression in *Aldh1a2* KO	**Other genes in same TAD with decreased or increased expression in *Aldh1a2* KO**	**RARE DNA sequence 5'-3' overall consensus:** AGGTCA N5,N2 AGGTCA G T or N1 G T	**Type of RARE**	**Conserved**						**Genomic coordinates (mm10)**
				rodent	**human**	**bird**	**reptile**	**frog**	**fish**	
RARE ENHANCERS (RA stimulates gain of H3K27ac and/or loss of H3K27me3 near RARE and activates gene in same TAD)										
gene nearest to RARE is RA-activated	gene in same TAD that is RA-activated									
Btbd9	none	AGTTAA C AGGTCA	DR1	x						chr17:30471156-30471168
Cdx1	Arsi	GGGTTA G GGGTCA	DR1	x	x					chr18:61036704-61036716
	Dctn4	GGGTCA AG AGTTCA	DR2	x	x					chr18:61035133-61035146
Crabp2	Prcc	AGTTCA CC AGGTCA	DR2	x						chr3:87947525-87947538
	Mef2d	AGGGCA G AGGTCA	DR1	x	x					chr3:87948035-87948047
		AGGTCA GG AGGGCA	DR2	x	x					chr3:87958618-87958631
Ctbp2	Zranb1	AGGTCT CT GTGTCA	DR2	-						chr7:133035195-133035208
	Fam53b	GGGTCA AT GGGTCT	DR2	x						chr7:133037547-133037560
	Edrf1									
Dach1	none	AGTTCA CACAA AGTTCA	DR5	x	x	x	x	x	x	chr14:98035388-98035404
		GGGACA A AGGTCA	DR1	x	x					chr14:98037394-98037406
Dhrs3	none	GGGTCA TTCCA AGTTCA	DR5	x	x	x	x	x		chr4:145034810-145034826
		GGTTCA TCGGG AGGGCA	DR5	x	x	x	x	x		chr4:145034847-145034863
Foxp2	none	AGGTGA A AGTTCA	DR1	x	x					chr6:14898480-14898492
Foxp4	none	GGGTGA C AGGTCA	DR1	x	x	x	x			chr17:47898625-47898637
Fzd7	Tmem237	AGGTCA G GGTTCA	DR1	x	x					chr1:59475768-59475780
	Stradb									
Hoxa1	Hoxa4	GGTTCA CCGAA AGTTCA	DR5	x	x	x	x	x		chr6:52153426-52153442
	Hoxa9	AGGTCA CT AAATCA	DR2	-						chr6:52156158-52156171
	Skap2	GGTTCA AGAAG AGTTCA	DR5	x	x	x	x	x	x	chr6:52175533-52175549
Itga9	none	AGGTCA GCCGG AGGGCA	DR5	x	x					chr9:118655221-118655237
		AGGCCA A AAGTCA	DR1	x						chr9:118656142-118656154
Lrp8	none	AGGTCA CT GGGGGA	DR2	-						chr4:107836195-107836208
Meis1	none	ATGTCA G AGGTCA	DR1	x	x					chr11:19025437-19025449
		GGGTCA G AGGCCA	DR1	x	x					chr11:19016387-19016399
		AGGGCA GG GGGCCA	DR2	x						chr11:19010468-19010481
		AGGCCA CTGAG AGGTCA	DR5	x	x	x	x	x		chr11:18963875-18963891
		ATGTCA A AGGACA	DR1	x						chr11:18958299-18958311
Meis2	Dph6	AGGTCA AAAAC AGTTCA	DR5	x	x	x	x			chr2:116071242-116071258
		CTCTCA AA GGGTCA	DR2	x	x					chr2:116020707-116020720
Mtcl1	none	GGGTCA GGAGG AGTTGA	DR5	-						chr17:66412903-66412919
		GGGTCA C AGGTCA	DR1	x	x					chr17:66413174-66413186

Nr2f1	none	GTGTCA A AGTTCA	DR1	x	x	x	x	x	x	chr13:78200425-78200437
Nr2f2	none	GTGTCA A AGTTCA	DR1	x	x	x	x	x	x	chr7:70361772-70361784
Ppp1r9a	Casd1 Gng11	GGGTCA AGGGC ATATCA	DR5	-						chr6:5030269-5030285
		AGGGCA CT GGCTCA	DR2	-						chr6:5031044-5031057
Ptprs	Zfp119a Ranbp3 Ndufa11	GGGTCA CG AGGTCA	DR2	x						chr17:56470658-56470671
		AGGTCA CC AGGTCA	DR2	x						chr17:56470666-56470679
		AGGTCA C AGGGCA	DR1	-						chr17:56470674-56470686
		GGATCA GG AGTTCA	DR2	x						chr17:56471720-56471733
Rarb	none	GGTTCA CCGAA AGTTCA	DR5	x	x	x	x			chr14:16575513-16575529
		AGGACA G AGGTCA	DR1	-						chr14:16578037-16578049
Robo2	none	GTGGCA A AGGTCA	DR1	-						chr16:74398869-74398881
Smad3	none	GGGTCA TGTGA AGTTCA	DR5	x	x					chr9:63716397-63716413
Sox2	none	GGGTCA GG AGGTCA	DR2	x	x	x	x	x	x	chr3:34679067-34679080
Spsb4	none	GGGTCA CGCAC GGGTCA	DR5	x						chr9:96989808-96989824
		AGCTCA CT GGGGCA	DR2	-						chr9:96958298-96958311
Stx8	Ntn1	AGTTCA G AGTTCA	DR1	x						chr11:68051350-68051362
		AGTTCA CT GTGGCA	DR2	x	x					chr11:68051567-68051580
Tshz1	none	GGGTCA TTCAT AGTTCA	DR5	x	x	x	x			chr18:84073476-84073492
		AGGTCA CCCAG AGTTCA	DR5	x						chr18:84075146-84075162
		AGGTCA GG AGGTGA	DR2	x	x	x	x			chr18:83839858-83839871
		GGGTGA ACTCA GGTTCA	DR5	x	x	x	x			chr18:83839869-83839885
Ttc28	Chek2	AGGTCA G AGGTTA	DR1	x						chr5:111244585-111244597
Tubb2b	Psmg4	GTGTCA GT GGGTCT	DR2	x						chr13:34130267-34130280
Vwa8	none	GAGTCA A AGGTCA	DR1	x						chr14:78849683-78849695
		AGGTCA TACAC AGGCCA	DR5	-						chr14:78850773-78850789
		GGCTTA CT GGGTCA	DR2	-						chr14:78851047-78851060
		GGGTCA A AGTTCA	DR1	x						chr14:78851055-78851067
Zbtb16	none	GGGTCA CA GGGTCA	DR2	x	x	x	x	-	x	chr9:48694721-48694734
		GGGTCA G GGGTTA	DR1	x	x	x	x			chr9:48695827-48695839
		GGGTCA G AGGCCA	DR1	x	x					chr9:48696900-48696912
Zfand5	none	GGGTCA TT GGGTAA	DR2	x						chr19:21165208-21165221
Zfhx4	Pex2	GGGTCA GCCTG AGGTCA	DR5	x	x	x	x	x	x	chr3:5388103-5388119
Zfp638	none	GGTTCA GCCAA AGGTGA	DR5	x	x	x	x	x		chr6:84976840-84976856
1700017B05Rik	Commd4 Man2c1	AGGTAA A AGGTCA	DR1	x	x					chr9:57265274-57265286
		GGGTCT CT GGGTCT	DR2	-						chr9:57266886-57266899
		65 RAREs - nearest gene is RA-activated								
(gene nearest to RARE is not RA-activated)	gene in same TAD that is RA-activated									
(Gm15428)	Ralgps2 Fam20b	GGGTCA G AGATCA	DR1	x						chr1:156789360-156789372
		GGGTCA GTGAG GGGTCA	DR5	x						chr1:156790306-156790322
(Gm37839)	Lmx1a Pbx1	GGGTCA AACGC AGGGCA	DR5	x	x					chr1:169238665-169238681
		GGGTCG CT GGGTCA	DR2	x	x	x	x			chr1:169238844-169238857
(Gm6075)	Crispld1	ATGTCA GT AGGACA	DR2	-						chr1:17450456-17450469

(Gm37068)	Ccdc115	AGGTCA TTCAA AGGTCA	DR5	x							chr1:35694541-35694557
(Col4a4)	Irs1	AGGTCA A AGGTCA	DR1	-							chr1:82475916-82475928
(Gm28884)	B3gnt7	GGGTCA GACAC AGGGGA	DR5	-							chr1:85933157-85933173
(Acoxl)	Bcl2l11	GGGTCA G AGGCCA	DR1	x	x						chr2:127907834-127907846
(1700001O22Rik) (Ptges)	Prrx2 Ntmt1	AGTTCA A GGTTAT	DR1	-							chr2:30779109-30779121
		AGGCCA GGCAG AGGTCA	DR5	x	x						chr2:30779417-30779433
		GGGTCA CAGAG AGGTCA	DR5	x	x						chr2:30779575-30779591
		GGGTCA G AGGCGA	DR1	x							chr2:30893114-30893126
		AGTTCA A AGTTGA	DR1	x							chr2:30893953-30893965
		AGTTCA AGGTC AGTGCT	DR5	x							chr2:30894156-30894172
(Gm25869)	Olfml3	AGGTCA GGGAG AAGTCA	DR5	x							chr3:103419827-103419843
		AGGTCA AGGAG GATTCT	DR5	x	x						chr3:103420712-103420728
(Mcoln2)	Prkacb	AGGTCA C AGGTCA	DR1	x							chr3:146204096-146204108
		GGGTCA CACAG GGGTCA	DR5	x							chr3:146204537-146204553
(Gm37359)	Tiparp	AGGTCA CA GGGTCA	DR2	x							chr3:65859155-65859168
(Trim62)	Zscan20	GGGTCA CA GGGTCA	DR2	x							chr4:128897746-128897759
		AGGTCT GG GGGGCA	DR2	-							chr4:128898695-128898708
(Gm12992)	Ythdf2	AGGTCA CACAG AGGCCA	DR5	x	x						chr4:131918909-131918925
		GGGCCA G AGTTCA	DR1	x							chr4:131919741-131919753
(Grrp1)	Pdik1l Dhdds	AGGTGG G AGGTCA	DR1	x							chr4:134256564-134256576
		AGTTGA G AGGTGA	DR1	-							chr4:134257704-134257716
(Gm13200)	Mad2l2	GGGGCA AGCAG GGGTCA	DR5	x							chr4:148383603-148383619
		AAGTCA CC GGGTCA	DR2	-							chr4:148383993-148384006
(Spsb1)	Clstn1 Lzic Nmnat1 Kif1b	GGGTCA GA AGGTCA	DR2	x	x	x	x	x			chr4:149907094-149907107
		AGGTCA G AGGGCA	DR1	x							chr4:150095743-150095755
(Whrn)	Akna	GGGTCA CG GGGTCG	DR2	-							chr4:63466825-63466838
(Stx2)	Fzd10 Mmp17 Snora15 Sumf2 Chchd2 Gusb	AGGTCA G AGGGAA	DR1	-							chr5:128975872-128975884
		AGGTCA TCCTG AGGGCA	DR5	x							chr5:128976529-128976545
(Clip2)	Abhd11 Bcl7b	GGGTCA CCGAG AGGTCA	DR5	x	x						chr5:134541736-134541752
		AGGTTA T AGGTCA	DR1	-							chr5:134542981-134542993
(Wasf3)	Rnf6	GGGTGG G AGGTCA	DR1	-							chr5:146386603-146386615
(Uspl1)	Ubl3	AGGTCA A AGGTCA	DR1	-							chr5:149196756-149196768
		GGGTCA AACTC AGGTCA	DR5	x							chr5:149196825-149196841
(Dgki)	Ptn	GGGTCA GGGTG AGCACA	DR5	x							chr6:36880813-36880829
(8030453O22Rik)	Ndufb2	AGTTCA GT GGCTCA	DR2	-							chr6:39545227-39545240
(Gm5876)	Prdm5	AGGTCA GCAGC AGGTCA	DR5	x	x						chr6:66077462-66077478
(Efcc1)	Isy1	AGGTCA G AGGTCA	DR1	x							chr6:87739484-87739496
(Gm23847)	Fgfr2 Ate1	GGTTCA G AGTTCA	DR1	x							chr7:130009273-130009285
(Nav2)	Prmt3	AGGTCA TAAAC AAGTCA	DR5	x							chr7:49333913-49333929
(Gse1)	Tldc1 Hsdl1 Zdhhc7 Gins2	GGGTCA A AGGGGA	DR1	-							chr8:120408671-120408683
		AGGGCA G AGGGCA	DR1	-							chr8:120409479-120409491
(Igdcc3)	Dpp8	GAGTCA A AGGGCA	DR1	x	x						chr9:65163164-65163176

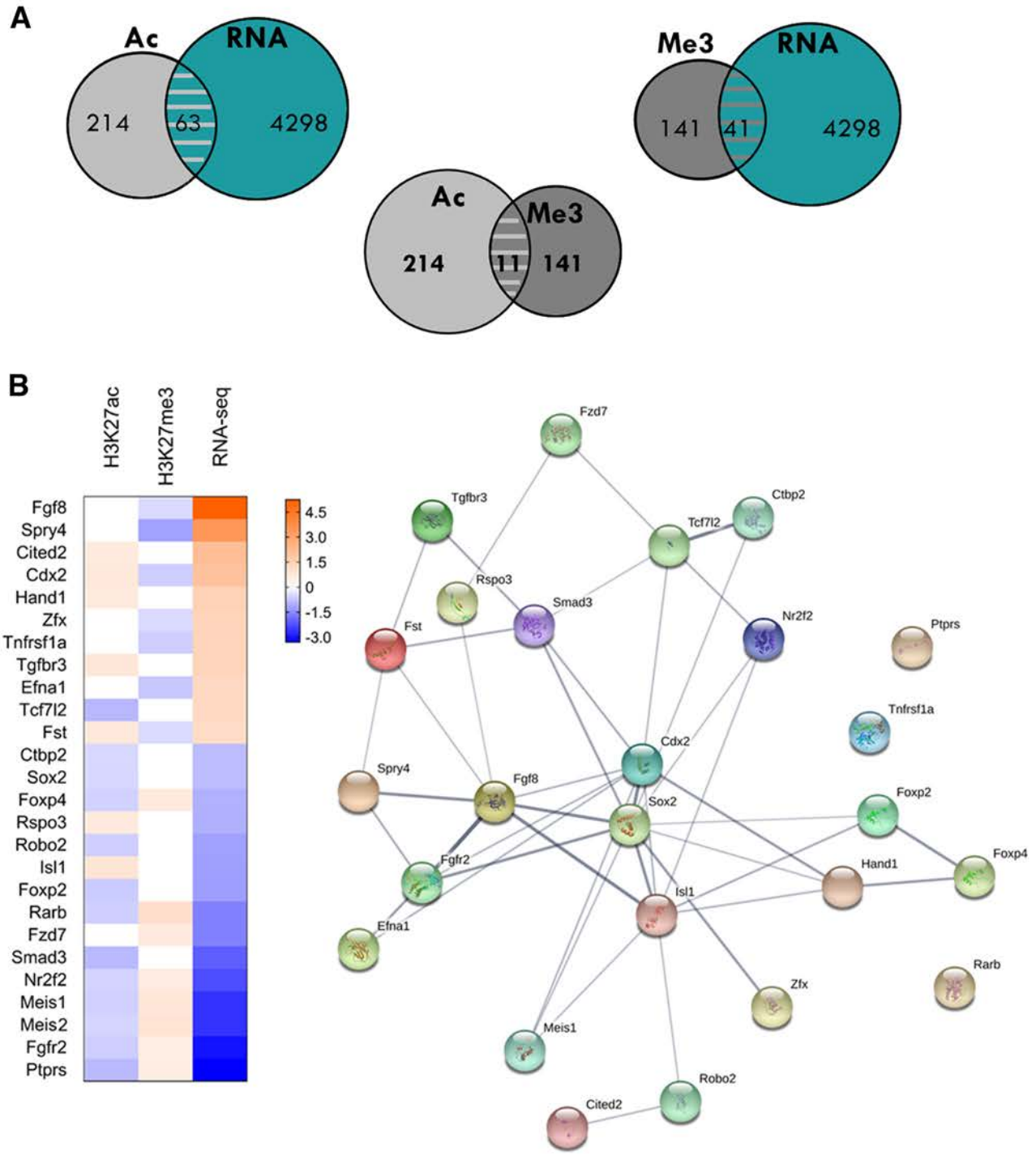
(E130307A14Rik)	Rpf2	GGGTCA A AGGTCA	DR1	x	x						chr10:39660884-39660896
(Polr3b)	Ric8b	GTTTCA A AGGTCA	DR1	-							chr10:84710728-84710740
(Evpl)	Unk	GGATCA GA AGTTCA	DR2	x							chr11:116241241-116241254
	Mrpl38										
	Srp68										
(Gm28401)	C1d	GGGTCA G GGGTTA	DR1	x	x	x	x	x	x		chr11:18748180-18748192
		GGATCA GT GTGTCA	DR2	x	x						chr11:18749733-18749746
(Cdk5rap3) (Gm11523)	Scrn2	GGGTCA G AGGGCA	DR1	-							chr11:96896439-96896451
	Hoxb3	AGGTCA G AGGTGA	DR1	-							chr11:96897098-96897110
	Hoxb5os	AGGACA G AGGTCA	DR1	-	x						chr11:96899490-96899502
		GGGTCA G GGGGAGA	DR1	-							chr11:96899889-96899902
		GACTCA AG AGTTCA	DR2	-							chr11:96877903-96877916
(Fam181a)	Otub2	GGTTAC TG AGGTCA	DR2	x	x						chr12:103314689-103314702
(Gm5441)	Zfp386	GAGTCA A AGGTCA	DR1	x	-	x	x				chr12:117352086-117352098
(Gm25538) (Ddx1)	Fam49a	AGGTCA A AGGTGA	DR1	x	x						chr12:12723995-12724007
	Mycn	GGGTGA A AGGTCA	DR1	x							chr12:13207084-13207096
		AGTTCA A GGTCCT	DR1	x							chr12:13209724-13209736
(Agr3)	Ispd	GGGGCA ATGTG AGGTCA	DR5	x							chr12:35894512-35894528
		AGTTCA G GGGTCA	DR1	x	x						chr12:35894569-35894581
(Unc5a)	Rab24	GGGTTA G AGTTCA	DR1	x	x						chr13:54993593-54993605
		GGGTCA CT GGCTTA	DR2	x	x						chr13:54994828-54994841
(B230219D22Rik)	Grk6	AGGTCA TA GGGTCA	DR2	x							chr13:55688082-55688095
	Ddx41										
(Kif13b)	Ints9	GGGTCA AACTC AGGTCA	DR5	x							chr14:64640716-64640732
	Fzd3										
(Snora31)	Afap1	GGGTCA GTCAG GGGTCA	DR5	x							chr14:75863890-75863906
	Trmt44										
(Celsr1)	Cerk	AGGTCA GA GGCTCA	DR2	x							chr15:85971844-85971857
	Trmu										
(Gm15742)	Ppp1r2	AGTTGA A AGGCCA	DR1	x							chr16:30939753-30939765
(Spice1)	Naa50	AGGTCA CA GTGTGA	DR2	x	x						chr16:44382205-44382218
		ATTTCT G AGTTCA	DR1	-							chr16:44382403-44382415
(Gm26694)	Capn15	AGGTCA CCACC AGGCCA	DR5	x	x						chr17:25795888-25795904
	Wdr24										
	Fbxl16										
	Wdr90										
	Ube2i										
	Ift140										
	Nme3										
(Gm26682)	Zeb1	AGGTCA ACCAG AGGGCA	DR5	x	x						chr18:5186998-5187014
(Gm14505)	Cask	GGTTCA CAGAA AGTTCA	DR5	x							chrX:14628834-14628850
(Actrt3)	Skil	GAGGCA GG AGGTCA	DR2	-							chr3:30596235-30596248
		GAGTCA CT GGGTCA	DR2	-							chr3:30596543-30596556
(Slfn14)	Nle1	AGGGCA CAGCA AGGTCA	DR5	x							chr11:83277949-83277965
	Zfp830										
(Sacs)	Tnfrsf19	AGGTCA GA GGGAGA	DR2	-							chr14:61167047-61167060
	Mipep										
(Sp7)	Aaas	AGGTGA GCTTG AGGCCA	DR5	x							chr15:102365172-102365188
	Pfdn5	GGGTCA G AGGGCA	DR1	x	x						chr15:102366013-102366025

		82 RAREs - nearest gene not RA-activated but other gene(s) in TAD are								
RARE SILENCERS (RA stimulates gain of H3K27me3 and/or loss of H3K27ac near RARE and represses gene in same TAD)										
gene nearest to RARE is RA-repressed	gene in same TAD that is RA-repressed									
C77080	Zbtb8os	GGGTGA TCCAA AGGCCA	DR5	-						chr4:129248733-129248749
Cdx2	Lnx2 Usp12	GGGTCA CT GGGTGA	DR2	x	x					chr5:147301839-147301852
		GGCTCA CA GTGTCA	DR2	-						chr5:147302661-147302674
		AGGTCA CT TGGTCA	DR2	x	x					chr5:147303936-147303949
Cited2	Heca	AGATGA G AGGTCA	DR1	-						chr10:17705156-17705168
Evx1	none	GGGGCA G AGGTGA	DR1	-						chr6:52312241-52312253
Fgf8	Poll Btrc Mrpl43 Sema4g Chuk Erlin1 Dnmbp Entpd7 Got1 Slc25a28	GGGTCA GC AGTTCA	DR2	x	x	x				chr19:45747043-45747056
		AGGTCT CT GGGTCG	DR2	-						chr19:45743342-45743355
		AGGGCA G AGGCCA	DR1	x						chr19:45735030-45735042
Fst	Ndufs4 Mocs2	GGGGCA GG GGTTCT	DR2	x						chr13:114458455-114458468
Gpx3	none	GGATCA A AGTTCA	DR1	x						chr11:54892620-54892632
		GGGTCA G AGGTCG	DR1	-						chr11:54892779-54892791
Sel1l3	Slc34a2	AGGTCA G AGGTCA	DR1	x						chr5:53109896-53109908
		GAGTCA A AGTTCA	DR1	x	x					chr7:27356622-27356634
Tfcp2l1	none	AGGTCA TTATC AGGTGA	DR5	x						chr1:118648735-118648751
Tmc8	Syngr2 Tk1	TGGTCA GT GGGTCT	DR2	x						chr11:117782965-117782978
		GGGTCA TG GGGACA	DR2	x						chr11:117784333-117784346
Tnfrsf1a	Cd9 Nop2 Spsb2 Ptpn6	AGGTCA TG GAGTCA	DR2	x						chr6:125360617-125360630
Zdhhc15	none	AGGTCT GT GGGCCA	DR2	-						chrX:104539271-104539284
		GGGTCC CTGTG AGTTCA	DR5	-						chrX:104571199-104571215
		20 RAREs - nearest gene is RA-repressed								
nearest gene not RA-repressed	gene in same TAD that is RA-repressed									
(Gm13686)	Tfpi	GGGTCA A AGGTGA	DR1	x						chr2:83709948-83709960
(Accsl)	Cd82 Tspan18	AGGGCA A AGGTCA	DR1	x						chr2:93873962-93873974

(Wnt5b)	Erc1	AGGTCA AG GGCTCA	DR2	-						chr6:119448467-119448480
	Adipor2									
(Lrrc27)	Pwwp2b	AGTTCA A AGTCCA	DR1	x						chr7:139234821-139234833
(Ldha)	Hps5	GAATCA G AGTTCA	DR1	x						chr7:46837621-46837633
		AGCTCA CT AGGCCA	DR2	x	x					chr7:46839662-46839675
		AGGGCA A AGGTGA	DR1	x	x					chr7:46839730-46839742
(Arid5b)	Rhobtb1	AGGTCA GAGAA AGGTCA	DR5	x						chr10:68229857-68229873
(C030005K06Rik)	Dtwd2	AGGGCA A AGGGCA	DR1	-						chr18:50052943-50052955
(Epha2)	Plekhm2	GAGGCA G AGGTCA	DR1	x						chr4:141300510-141300522
	Dnajc16									
(Hpse)	Plac8	GGATCA GC AGTTCA	DR2	x						chr5:100717951-100717964
		AGATCA AA AGTTCA	DR2	x						chr5:100718336-100718349
		AGGTCA GT GGGACA	DR2	-						chr5:100718666-100718679
(Sh2b3)	Trafd1	AGGTCA G GGTCAA	DR1	-						chr5:121839252-121839264
(Cux1)	Znhit1	AGGCTA TG AGTTCA	DR2	-						chr5:136582974-136582987
(4933427G23Rik)	Rint1	ACGTCA GT GGGACA	DR2	-						chr5:23830755-23830768
(Shkbp1)	Numbl	GAGTCA A AGTTCA	DR1	x	x					chr7:27356622-27356634
	Sertad1	GGATTA TATTG AGTTCA	DR5	x						chr7:28377231-28377247
	Akt2 Zfp60 Rps16 Dll3 Paf1 Fbxo17									
(Usp10)	Cotl1	AGGGCA GG AGGACA	DR2	x						chr8:119898220-119898233
	Taf1c Necab2									
(Gm4895)	Sgk1	AGGCCA G AAGTCA	DR1	-						chr10:22156107-22156119
		AGGGAA A AGGTCA	DR1	-						chr10:22160643-22160655
(Rnf213)	Sgsh	GGGGCA G AGGCCA	DR1	-						chr11:119392214-119392226
	Gaa	AGTCCA GGCCA AGGACA	DR5	-						chr11:119392289-119392305
	Ccdc40	AGGTCA TAGGT AGTCCA	DR5	x						chr11:119392300-119392316
	Chmp6									
(Camk2b)	Ykt6	GGGTCA T AGGCCA	DR1	-						chr11:6007158-6007170
	Ddx56 Zmiz2									
(Gm16505)	Asb13	AGTCCA A AGGTCA	DR1	-						chr13:3362568-3362580
	Klf6									
(Il17rd)	Arhgef3	TGGTCA A AGGGCA	DR1	-						chr14:27039866-27039878
	Fam208a									
(Gm7030) (2410017I17Rik)	Ddr1	GTGTCA G AGGTCA	DR1	x						chr17:36128857-36128869
	Flot1	AGGCCA G AGGTCA	DR1	-						chr17:36155723-36155735
	Prr3	GGGTCG G AGGTCA	DR1	x						chr17:36167901-36167913
(4930519F16Rik)	Cdx4	ATTTCT G AGTTCA	DR1	-						chrX:103254783-103254795
	Chic1	AGTTCA G AGGTTA	DR1	x						chrX:103256648-103256660
		AGGCCA G AGGGCA	DR1	x						chrX:103257644-103257656
		ATTTCT G AGTTCA	DR1	-						chrX:103285450-103285462
		ATTTCT G AGTTCA	DR1	-						chrX:103337616-103337628
		ATTTCT G AGTTCA	DR1	-						chrX:103339320-103339332
		ATTTCT G AGTTCA	DR1	-						chrX:103342427-103342439
		ATTTCT G AGTTCA	DR1	-						chrX:103344477-103344489

(Gm15232)	Rab9	AAGTCA G AGGTCA	DR1	x						chrX:167125774-167125786
		AGGCCA G AGGTCA	DR1	x						chrX:167125965-167125977
		40 RARES - nearest gene not RA-repressed but other gene(s) in TAD are								

Fig. S1

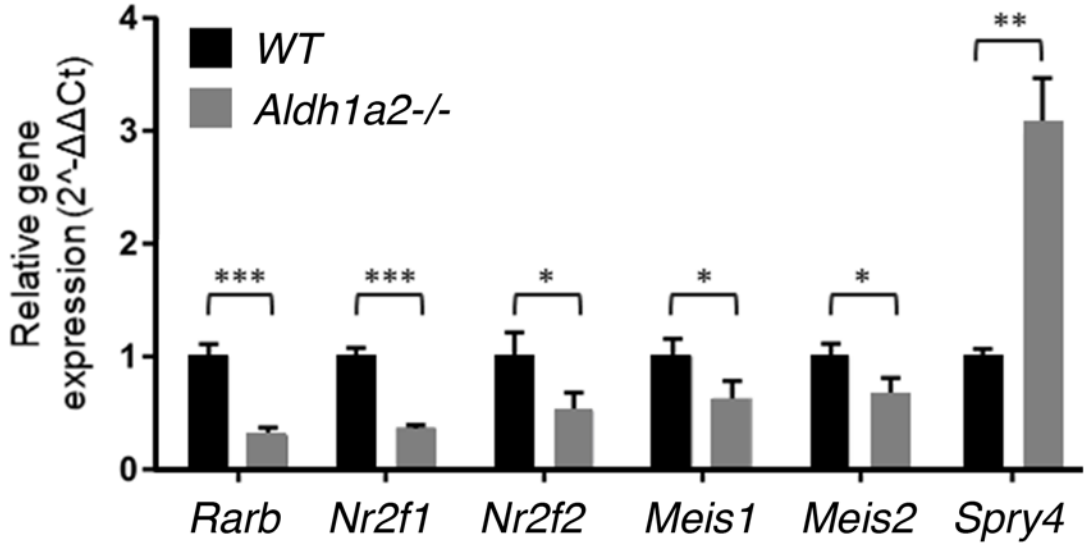


Supplemental Figure S1. Bioinformatic analysis of genes identified as RA target genes. Related to Table 1.

(A) Venn diagram showing the number of genes that have both RA-regulated expression and RA-regulated deposition of nearby H3K27a or H3K27me3 marks following loss of RA.

(B) The heatmap was designed with Prism software (left panel) from the list of genes involved in "Development of Body Trunk" obtained by IPA analysis of RA target genes identified by loss of RA, and the associated network was created using STRING software (right panel).

Fig. S2



Supplemental Figure S2. Analysis of differential gene expression of new RA target genes by qRT-PCR analysis of E8.5 wild-type vs *Aldh1a2*-/- trunk tissue. Related to Table 1. *p<0.05, **p<0.01, ***p<0.005; data expressed as mean ± SD (Student's t-test); WT and *Aldh1a2*-/-, n = 3 technical replicates.