Discovery of genes required for body axis and limb formation by global identification of 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

*Correspondence should be addressed to G.D. (duester@SBPdiscovery.org)

Keywords: Body axis formation; retinoic acid response elements; *Aldh1a2* knockout; *Nr2f1*; *Nr2f2*; *Meis1*; *Meis2*

Abstract

Identification of direct target genes for transcription factors is hampered by the large number of genes whose expression changes when the factor is removed and numerous binding sites in the genome. Retinoic acid (RA) receptors bound to RA directly regulate transcription through RA response elements (RAREs) of which there are thousands in the mouse genome. Here, we focused on identification of direct RA target genes in the embryonic trunk during body axis and limb formation. Using ChIP-seq and RNA-seq on trunk tissue from wild-type and *Aldh1a2*-/- embryos lacking RA synthesis, 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 RAREs. Such RAREs were identified in genes known to be required for body axis and limb formation, plus many new direct RA target genes were found. In addition to RARE enhancers, we identified numerous RARE silencers located near genes down-regulated by RA. *Nr2f1*, *Nr2f2*, *Meis1*, and *Meis2* gene family members all have RARE enhancers, and double knockouts of each family demonstrated requirements for body axis and/or limb formation, thus validating our method for identifying RA target genes important for development.

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 direct transcriptional activation or repression of key genes.

Identification of genes that are direct 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 direct targets of RA are dependent upon RAREs, identification of RAREs has been used to identify direct target genes. RAREs have been identified based on finding elements near RA-activated genes that contain a canonical RARE sequence and that can activate transcription in enhancer reporter transgenes either in transfected cell lines or transgenic animals. However, RAR chromatin immunoprecipitation (ChIP) studies on mouse embryoid bodies reported ~14,000 potential RAREs in the mouse genome [8,9]. Thus, it is unclear how many of these RAREs are required to regulate genes during development as only a few RAREs have been reported to have specific activity in vivo in animals carrying enhancer reporter transgenes [5]. Also, even fewer RAREs have been shown to result in 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* at the trunk/caudal border [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 enhancer reporter transgene analysis [12] was found to be unnecessary for *Tbx5* activation and forelimb budding when subjected to deletion analysis [13]. It is now clear that RAREs (and other DNA control elements) that exhibit appropriate tissue-specific expression in enhancer reporter transgene assays (where the element is inserted at a random location in the genome) are often not required in vivo [14]. Thus, additional methods are needed to identify RAREs that are required to activate or repress specific genes.

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 RAREs associated with altered epigenetic marks that are located near genes found to be upregulated or downregulated by RA. This approach was able to identify genes directly regulated by RA at the transcriptional level as opposed to genes whose mRNA levels are changed indirectly or post-transcriptionally when RA is lost. Our findings identified many previously reported

direct RA target genes known to control embryonic trunk development as well as numerous new candidate direct RA target genes that may control trunk development. Knockout studies on several of the genes we identified as candidates for direct RA activation in the developing trunk demonstrated they are required for body axis and/or limb formation.

Results

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

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]; data available at GEO under accession number GSE131624. 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]; data available at GEO under accession number GSE131624.

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 expression in trunk tissue is significantly expressed (FPKM>0.5) and dysregulated (log2 <-0.85 or >0.85) when RA is absent (data available at GEO under accession number GSE131584); a cut-off of log2 <-0.85 or >0.85 was employed to include *Sox2* known to be activated by RA. The much smaller number of RA-regulated peaks found with ChIP-seq compared to the very large number of differentially expressed genes found with RNA-seq suggests that most genes differentially expressed following loss of RA are not directly regulated at the transcriptional level by RA.

In order to identify genes that are good candidates for being transcriptionally activated or repressed by RA, we compared our ChIP-seq and RNA-seq results to identify ChIP-seq peaks where the nearest gene has significant changes in expression in wild-type vs *Aldh1a2*-/- based on RNA-seq. We found 88 RA-regulated peaks for H3K27ac (within or near 75 genes) with significant changes in expression when RA is lost (Table S1), plus 62 RA-regulated peaks for H3K27me3 (within or near 49 genes) with significant changes in expression when RA is lost (Table S2). Thus,

many RA-regulated peaks are found within or near differentially expressed genes, i.e. 28% (88/314) of the peaks for H3K27ac and 24% (62/262) of the peaks for H3K27me3, a total of 150 RA-regulated peaks. We cannot rule out that other H3K27ac and H3K27me3 RA-regulated peaks may be related to RA-regulated genes located further away, however the 88 H3K27ac and 62 H3K27me3 RA-regulated peaks identified here provide evidence that many of these epigenetic marks regulate the nearest gene. As some genes have more than one nearby RA-regulated peak for H3K27ac or H3K27me3, and as 11 genes have nearby RA-regulated peaks for both H3K27ac and H3K27me3 (*Rarb*, *Fgf8*, *Fst*, *Cdx2*, *Meis1*, *Meis2*, *Nr2f2*, *Fgfr2*, *Foxp4*, *Ptprs*, and *Zfhx4*), a total of 113 genes have nearby RA-regulated peaks for H3K27ac and/or H3K27me3 when RA is lost (Tables S1-S2; Fig. S1).

Among the list of the RA-regulated peaks (H3K27ac or H3K27me3) near differentially expressed genes when RA is lost, we found many examples of genes previously reported to be regulated by RA based on studies of *Aldh1a2*-/- embryos [5,20] or RA-treated NMPs [19]; this includes *Rarb*, *Pax6*, *Crabp2*, *Cdx1*, *Sox2*, *Dhrs3*, and *Hoxa1* whose expression is increased by RA, plus *Fgf8*, *Fst*, and *Cdx2* whose expression is decreased by RA (Table 1). H3K27ac peaks near *Rarb*, *Pax6*, *Crabp2*, *Cdx1*, and *Sox2* are reduced in *Aldh1a2*-/- consistent with these being RA-activated genes, whereas H3K27ac peaks near *Fgf8*, *Fst*, and *Cdx2* are increased in *Aldh1a2*-/- consistent with these being genes repressed by RA. H3K27me3 peaks near *Cdx2*, *Fgf8*, and *Fst* are decreased in *Aldh1a2*-/-, whereas H3K27me3 peaks near *Dhrs3*, *Hoxa1*, and *Rarb* are increased in *Aldh1a2*-/-, consistent with the former being genes repressed by RA and the latter being genes activated by RA. In addition to these well established RA target genes, we also further analyzed several additional genes that our findings indicate are RA target genes regulated at the transcriptional level, including *Nr2f1, Nr2f2, Meis1, Meis2, Spry4, and Trp53cor1* (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 differential expression that also have nearby RA-regulated peaks for H3K27ac and H3K27me3 in wild-type vs *Aldh1a2*-/- embryos is a reliable method for identifying known RA target genes, and can thus be used to identify new target genes regulated by RA at the transcriptional level.

Identification of known RAREs associated with RA-regulated deposition of H3K27ac or H3K27me epigenetic marks

As direct RA target genes need to be associated with a RARE, the RA-regulated ChIP-seq peaks we found near our list of 113 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) and the presence or absence of RAREs is summarized (Tables S1 and S2). We found that 60 of these 113 genes contained at least one RARE in their nearby RA-regulated H3K27ac and/or H3K27me3 ChIP-seq peaks, whereas 53 genes did not, indicating that 53% of the genes we identified with our approach are candidate direct RA target genes.

Our analysis here identified the three RAREs previously shown to have required functions in vivo by knockout studies (RAREs for *Hoxa1*, *Cdx1*, *Fgf8*) plus many RAREs previously associated with RA-regulated genes (*Rarb*, *Crabp2*, *Sox2*, *Fst*, *Cdx2*) validating our approach for identifying functional RAREs and direct RA target genes (Table 1). The sequences of the RAREs for these known RA target genes and several candidate new RA target genes that we further examined are summarized (Table S3).

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 [21] 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 to activate *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 [22] 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 [23]. 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.

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 113 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 [24-31]. 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 [32-34]. Caudal Wnt and FGF signals are required to establish and maintain NMPs [32,34-39]. Also, *Cdx2* is required for establishment of NMPs [31]. 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 results 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 [18,19,40,41]. Also, *Cdx2* expression is increased when RA is lost in *Aldh1a2*-/- embryos [42].

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, suggesting that *Sox2*,

Fgf8, and *Cdx2* are direct transcriptional targets of RA (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). Overall, these findings indicate that RA controls NMP differentiation directly at the transcriptional level by activating *Sox2* and repressing *Fgf8* and *Cdx2* as progenitor cells progress from a caudal to a trunk location.

Identification of additional RARE silencers associated with RA-regulated epigenetic marks

Up to now, *Fgf8* represents the only example of a gene that is directly repressed by RA at the transcriptional level as defined by identification of a RARE that RA-dependently decreases expression of a marker gene in a transgenic animal [43] and by knockout of this RARE in mouse embryos [7]. Here, we found many more candidates for genes directly repressed by RA, i.e. genes whose expression is increased when RA is lost that have nearby H3K27me3 marks that are decreased and/or H3K27ac marks that are increased (Table 1; Tables S2-S3). In additional to *Cdx2* (described above), we examined in detail the ChIP-seq findings for *Fst* and *Trp53cor1* to identify potential new RARE silencers (Fig. 3); RAREs are listed in Table S3. For *Fst*, we identified a RARE in intron 1 that is located within a RA-regulated peak for H3K27ac (increased when RA is lost) and is near the edge of a RA-regulated peak for H3K27me3 (decreased when RA is lost) (Fig. 3A). For *Trp53cor1*, a gene not previously reported to be repressed by RA, we found a RARE in the 5'-noncoding region close to exon 1 that is located within a RA-regulated peak for H3K27me3 (decreased when RA is lost) (Fig. 3B). Overall, our findings suggest that RARE silencers may be nearly as common as RARE enhancers.

Evidence for genes regulated indirectly by RA at the transcriptional level

Our studies show that some genes that are downregulated (*Pax6* and *Dhrs3*) or upregulated (*Spry4*) following loss of RA have nearby RA-regulated peaks for H3K27ac or H3K27me3 that do not contain RAREs (Fig. S3; Table S3). These are likely indirect RA-regulated peaks that contain DNA binding sites for other transcription factors whose expression or activity is altered by loss of RA, with nearby genes being 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. S3A). 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 [42]. Activation of *Pax6* also requires that caudal FGF signaling be downregulated [41]. Thus, it is possible that the RA requirement for *Pax6* activation operates through several indirect mechanisms due to the ability of RA to directly activate *Sox2* and *Cdx1*, and directly repress *Fgf8* (Figs. 1, 2).

***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. *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 [46].

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, and a RARE in the 5-noncoding region near exon 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). *Meis1* and *Meis2* were

previously shown to be upregulated by RA [47]. *Meis1* was previously reported to be activated by RA in embryonic stem cells [48]. *Meis2* has previously been reported to be activated by RA and the RARE located in its 5'-noncdoing region was previously identified in cell line studies [9]. Together, these studies suggest that the *Nr2f* and *Meis* gene families are direct transcriptional targets of RA in the developing trunk.

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

As both *Nr2f1* and *Nr2f2* are activated by RA in the body axis and have nearby RA-regulated epigenetic marks, they may be required for body axis extension. *Nr2f1* (formerly known as COUP-TFI) and *Nr2f2* (formerly known as COUP-TFII) are both expressed at E8.5 in presomitic mesoderm and somites but not spinal cord, suggesting they may function in mesoderm formation during body axis formation [49,50]. The *Nr2f1* knockout is lethal at birth with brain defects but no somite, spinal cord, or body axis defects are observed [51]. The *Nr2f2* knockout is lethal at E10.5 with defects in heart development [52].

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 exhibit a growth defect and are more 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 (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 identification of RA-regulated epigenetic marks in wild-type and RA-deficient embryos can identify new direct RA target genes essential for development.

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

Meis1 and *Meis2* are both expressed throughout the trunk and in the proximal regions of limb buds [47]. As we have shown here that RA contributes to up-regulation of both *Meis1* and *Meis2* in the body axis at E8.5 and as both genes have nearby RA-regulated epigenetic marks, they may be required for body axis or limb formation. The *Meis1* knockout is lethal at E11.5 with hematopoietic defects, but no body axis or limb defects are observed [53]. 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 [54].

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 growth 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 direct RA target genes essential for development.

Discussion

Our epigenetic ChIP-seq studies and identification of associated RARE enhancers and silencers, combined with RNA-seq, provide a means of identifying direct transcriptional targets of RA by focusing on genes that have a change in expression along with nearby changes in H3K27ac and/or H3K27me3 epigenetic marks when RA is lost. Such an approach could be used to identify direct target genes for any transcriptional regulator that has an available knockout. In our case, we were able to narrow down 4298 genes identified with RNA-seq that have significant changes in gene expression following loss of RA to 113 genes that also have significant changes in nearby H3K27ac and/or H3K27me3 marks. By further identifying RA-regulated H3K27ac and H3K27me3 peaks that contain RAREs, our method allows one to identify genes that are most likely to be direct 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. 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* which is transcriptionally regulated by factors whose expression is altered by loss of RA including *Sox2* [44], *Cdx* [45], and *Fgf8* [41]. However, the number of genes regulated directly or indirectly by RA at the transcriptional level appears low compared to the large number of genes detected by RNA-seq, suggesting that most genes with altered expression in the RNA-seq list are due to changes in factors that might affect global mRNA stability and degradation. While H3K27ac and H3K27me3 epigenetic marks are quite commonly observed near genes during activation or repression, 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. Also, we show that many genes whose transcription start sites are nearest the RA-regulated H3K27ac and H3K27me3 marks are in fact bona fide RA target genes; shown most clearly for genes in which a nearby RARE was knocked out, including *Hoxa1* [10], *Cdx1* [11], and *Fgf8* [7]. However, further studies could be performed to determine if some RA-regulated H3K27ac and H3K27me3 marks may relate to RA regulation of genes located further away but still in the same topologically associated domain (TAD).

Our findings provide evidence that RARE silencers may be nearly as common as RARE enhancers. 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 RAR to NCOA to activate transcription, with unliganded RAR favoring release of NCOA and binding of NCOR to repress transcription [55]. 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 [43]. Here, we identified additional RARE silencers near *Fgf8*, *Cdx2*, *Fst* and *Trp53cor1*. These additional RARE silencers along with the original one found for *Fgf8* can be further examined 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 [18,27,30]. Our studies provide evidence that RA regulates several genes at the trunk/caudal border needed for NMP differentiation directly 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. Prior to these studies, RA was known to directly repress caudal *Fgf8* via a RARE silencer [7]. 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 RARE enhancers in 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. Thus, our epigenetic approach for identifying direct RA target genes has led to new targets for gene knockouts that provided new insight into control of body axis formation. 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 [46]. However, this observation is consistent with studies showing that RA is not required for NMP differentiation or body axis formation in zebrafish [56,57]. 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 [47,58,59]. However, in mouse, knockout of *Rdh10* required to generate RA demonstrated that complete loss of RA in the limb fields prior to and during limb initiation and outgrowth did not affect hindlimb initiation or patterning (although interdigital webbing was retained), whereas forelimbs were stunted (*Meis1* and *Meis2* expression was maintained in a proximal position in both forelimbs and hindlimbs [60,61]; 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 direct 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 ChIP-seq and RNA-seq to identify target genes that are controlled directly by RA at the transcriptional level. Such knowledge is essential for determining the mechanisms through which RA controls developmental pathways. A similar approach can be used to determine the direct target genes for other transcriptional regulators, thus accelerating the ability to understand developmental mechanisms 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 [43]. 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.

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
Trp53cor1-F	ACAGGGACCACAAAGTCTGC
Trp53cor1-R	CAAGGCCAGCCTCAGTTTAC

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 [62]. 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 [63]. ChIP-Seq sequencing reads were aligned using STAR aligner version 2.7 to Mouse genome version 38 [64]. Homer v4.10 [65] 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 [66]. P values from DESeq2 were corrected using the Benjamini & Hochberg (BH) method for multiple testing errors [67]. Peaks with BH corrected p value< 0.05 were selected as significantly differentially marked peaks. Transcription factor binding sites motif enrichment analyses were performed using Homer v4.10 [65] 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. 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 [68,69] 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.

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 [70].

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. This work was funded by National Institutes of Health grant R01 AR067731 (G.D.).

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, L. L. *et al.* RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. *Genes Dev.* **21**, 1113-1124 (2007).
2 Niederreither, K., Subbarayan, V., Dollé, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature Genet.* **21**, 444-448 (1999).
3 Mic, F. A., Haselbeck, R. J., Cuenca, A. E. & Duester, G. Novel retinoic acid generating activities in the neural tube and heart identified by conditional rescue of *Raldh2* null mutant mice. *Development* **129**, 2271-2282 (2002).
4 Rhinn, M. & Dolle, P. Retinoic acid signalling during development. *Development* **139**, 843-858 (2012).
5 Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nature Rev. Mol. Cell Biol.* **16**, 110-123 (2015).
6 Mark, M., Ghyselinck, N. B. & Chambon, P. Function of retinoid nuclear receptors: lessons from genetic and pharmacological dissections of the retinoic acid signaling pathway during mouse embryogenesis. *Annual Review of Pharmacology & Toxicology* **46**, 451-480 (2006).
7 Kumar, S., Cunningham, T. J. & Duester, G. Nuclear receptor corepressors *Ncor1* and *Ncor2* (*Smrt*) are required for retinoic acid-dependent repression of *Fgf8* during somitogenesis. *Dev. Biol.* **418**, 204-215 (2016).
8 Moutier, E. *et al.* Retinoic Acid Receptors Recognize the Mouse Genome through Binding Elements with Diverse Spacing and Topology. *J. Biol. Chem.* **287**, 26328-26341 (2012).
9 Lalevee, S. *et al.* Genome-wide in Silico Identification of New Conserved and Functional Retinoic Acid Receptor Response Elements (Direct Repeats Separated by 5 bp). *Journal of Biological Chemistry* **286**, 33322-33334 (2011).
10 Dupé, V. *et al.* In vivo functional analysis of the *Hoxa-1* 3' retinoic acid response element (3'RARE). *Development* **124**, 399-410 (1997).
11 Houle, M., Sylvestre, J. R. & Lohnes, D. Retinoic acid regulates a subset of Cdx1 function in vivo. *Development* **130**, 6555-6567 (2003).
12 Nishimoto, S., Wilde, S. M., Wood, S. & Logan, M. P. RA Acts in a Coherent Feed-Forward Mechanism with Tbx5 to Control Limb Bud Induction and Initiation. *Cell Reports* **12**, 879-891 (2015).
13 Cunningham, T. J., Lancman, J. J., Berenguer, M., Dong, P. D. S. & Duester, G. Genomic knockout of two presumed forelimb Tbx5 enhancers reveals they are nonessential for limb development. *Cell Reports* **23**, 3146-3151 (2018).
14 Duester, G. Knocking out enhancers to enhance epigenetic research. *Trends in Genetics* **35**, 89 (2019).
15 Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283 (2011).
16 Laugesen, A. & Helin, K. Chromatin repressive complexes in stem cells, development, and cancer. *Cell Stem Cell* **14**, 735-751 (2014).
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* **136**, 665-676 (2009).
18 Cunningham, T. J. *et al.* Retinoic acid activity in undifferentiated neural progenitors Is sufficient to fulfill Its role in restricting Fgf8 expression for somitogenesis. *PLOS ONE* **10**, e0137894 (2015).
19 Cunningham, T. J., Colas, A. & Duester, G. Early molecular events during retinoic acid induced differentiation of neuromesodermal progenitors. *Biology Open* **5**, 1821-1833 (2016).
20 Niederreither, K. & Dolle, P. Retinoic acid in development: towards an integrated view. *Nature Rev. Genet.* **9**, 541-553 (2008).

21 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* **113**, 723-734 (1991).
22 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* **71**, 73-85 (1992).
23 Gaunt, S. J. & Paul, Y. L. Origins of Cdx1 regulatory elements suggest roles in vertebrate evolution. *Int. J. Dev. Biol.* **55**, 93-98 (2011).
24 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* **146**, 24 (2019).
25 Wilson, V., Olivera-Martinez, I. & Storey, K. G. Stem cells, signals and vertebrate body axis extension. *Development* **136**, 1591-1604 (2009).
26 Kondoh, H. & Takemoto, T. Axial stem cells deriving both posterior neural and mesodermal tissues during gastrulation. *Curr. Opin. Genet. Dev.* **22**, 374-380 (2012).
27 Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864-2875 (2015).
28 Kimelman, D. Tales of tails (and trunks): forming the posterior body in vertebrate embryos. *Curr. Top. Dev. Biol.* **116**, 517-536 (2016).
29 Koch, F. *et al.* Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors. *Dev. Cell* **42**, 514-526 (2017).
30 Gouti, M. *et al.* A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev. Cell* **41**, 243-261 (2017).
31 Amin, S. *et al.* Cdx and T brachyury co-activate growth signaling in the embryonic axial progenitor niche. *Cell Reports* **17**, 3165-3177 (2016).
32 Olivera-Martinez, I., Harada, H., Halley, P. A. & Storey, K. G. Loss of FGF-Dependent Mesoderm Identity and Rise of Endogenous Retinoid Signalling Determine Cessation of Body Axis Elongation. *PLOS Biology* **10**, e1001415 (2012).
33 Tsakiridis, A. *et al.* Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development* **141**, 1209-1221 (2014).
34 Garriock, R. J. *et al.* Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development* **142**, 1628-1638 (2015).
35 Martin, B. L. & Kimelman, D. Canonical Wnt Signaling Dynamically Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation. *Dev. Cell* **22**, 223-232 (2012).
36 Naiche, L. A., Holder, N. & Lewandoski, M. FGF4 and FGF8 comprise the wavefront activity that controls somitogenesis. *Proc. Natl. Acad. Sci. USA* **108**, 4018-4023 (2011).
37 Jurberg, A. D., Aires, R., Novoa, A., Rowland, J. E. & Mallo, M. Compartment-dependent activities of Wnt3a/beta-catenin signaling during vertebrate axial extension. *Dev. Biol.* **394**, 253-263 (2014).
38 Takemoto, T. *et al.* Tbx6-dependent *Sox2* regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394-398 (2011).
39 Wymeersch, F. J. *et al.* Position-dependent plasticity of distinct progenitor types in the primitive streak. *eLife* **5**, e10042 (2016).
40 Diez del Corral, R. *et al.* Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65-79 (2003).
41 Patel, N. S. *et al.* FGF Signalling Regulates Chromatin Organisation during Neural Differentiation via Mechanisms that Can Be Uncoupled from Transcription. *PLOS Genetics* **9**, e1003614 (2013).

42 Zhao, X. & Duester, G. Effect of retinoic acid signaling on Wnt/beta-catenin and FGF signaling during body axis extension. *Gene Expr. Patterns* **9**, 430-435 (2009).
43 Kumar, S. & Duester, G. Retinoic acid controls body axis extension by directly repressing *Fgf8* transcription. *Development* **141**, 2972-2977 (2014).
44 Oosterveen, T. *et al.* SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. *Proc. Natl. Acad. Sci. USA* **110**, 7330-7335 (2013).
45 Joshi, P., Darr, A. J. & Skromne, I. CDX4 regulates the progression of neural maturation in the spinal cord. *Dev. Biol.* **449**, 132-142 (2019).
46 Dohn, T. E. *et al.* Nr2f-dependent allocation of ventricular cardiomyocyte and pharyngeal muscle progenitors. *PLoS Genetics* **15**, e1007962 (2019).
47 Mercader, N. *et al.* Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961-3970 (2000).
48 Kashyap, V. *et al.* RARg is essential for retinoic acid induced chromatin remodeling and transcriptional activation in embryonic stem cells. *J. Cell Sci.* **126**, 999-1008 (2013).
49 Béland, M. & Lohnes, D. Chicken ovalbumin upstream promoter-transcription factor members repress retinoic acid-induced *Cdx1* expression. *J. Biol. Chem.* **280**, 13858-13862 (2005).
50 Vilhais-Neto, G. C. *et al.* Rere controls retinoic acid signalling and somite bilateral symmetry. *Nature* **463**, 953-957 (2010).
51 Qiu, Y. *et al.* Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. *Genes Dev.* **11**, 1925-1937 (1997).
52 Pereira, F. A., Qiu, Y. H., Zhou, G., Tsai, M. J. & Tsai, S. Y. The orphan nuclear receptor COUP-TFII is required for angiogenesis and heart development. *Genes Dev.* **13**, 1037-1049 (1999).
53 Hisa, T. *et al.* Hematopoietic, angiogenic and eye defects in *Meis1* mutant animals. *EMBO Journal* **23**, 450-459 (2004).
54 Machon, O., Masek, J., Machonova, O., Krauss, S. & Kozmik, Z. Meis2 is essential for cranial and cardiac neural crest development. *BMC Developmental Biology* **15**, 40 (2015).
55 Perissi, V., Aggarwal, A., Glass, C. K., Rose, D. W. & Rosenfeld, M. G. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell* **116**, 511-526 (2004).
56 Begemann, G., Schilling, T. F., Rauch, G. J., Geisler, R. & Ingham, P. W. The zebrafish *neckless* mutation reveals a requirement for *raldh2* in mesodermal signals that pattern the hindbrain. *Development* **128**, 3081-3094 (2001).
57 Berenguer, M., Lancman, J. J., Cunningham, T. J., Dong, P. D. S. & Duester, G. Mouse but not zebrafish requires retinoic acid for control of neuromesodermal progenitors and body axis extension. *Dev. Biol.* **441**, 127-131 (2018).
58 Cooper, K. L. *et al.* Initiation of proximal-distal patterning in the vertebrate limb by signals and growth. *Science* **332**, 1083-1086 (2011).
59 Rosello-Diez, A., Ros, M. A. & Torres, M. Diffusible signals, not autonomous mechanisms, determine the main proximodistal limb subdivision. *Science* **332**, 1086-1088 (2011).
60 Cunningham, T. J., Chatzi, C., Sandell, L. L., Trainor, P. A. & Duester, G. *Rdh10* mutants deficient in limb field retinoic acid signaling exhibit normal limb patterning but display interdigital webbing. *Dev. Dyn.* **240**, 1142-1150 (2011).
61 Cunningham, T. J. *et al.* Antagonism between retinoic acid and fibroblast growth factor signaling during limb development. *Cell Reports* **3**, 1503-1511 (2013).
62 Voss, A. K., Dixon, M. P., McLennan, T., Kueh, A. J. & Thomas, T. in *Transcriptional Regulation: Methods and Protocols, Methods in Molecular Biology* Vol. 809 (ed A. Vancura) 335-352 (Humana Press, 2012).

63 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17** (2011).
64 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
65 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589 (2010).
66 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67 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* **57**, 289-300 (1995).
68 Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**, 910-918 (2013).
69 Tan, E. P., 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: the Journal of Genetics & Development* **53**, 225-236 (2015).
70 Sirbu, I. O. & Duester, G. Retinoic acid signaling in node ectoderm and posterior neural plate directs left-right patterning of somitic mesoderm. *Nature Cell Biol.* **8**, 271-277 (2006).

Table 1. Comparison of *Aldh1a2* KO and WT E8.5 trunk tissue for ChIP-seq and RNA-seq.

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**	**nearest gene with altered expression in *Aldh1a2* KO**	**log2 fold change for nearest gene: RNA-seq for *Aldh1a2* KO vs WT**
chr13:78197222-78204291	-1.23	DR1	Nr2f1	-2.02
chr14:16571405-16576397	-0.63	DR5	Rarb	-1.64
chr11:18962656-18965461	-0.61	DR5, DR1	Meis1	-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	-1.10
chr7:70348715-70369942	-0.57	DR1	Nr2f2	-2.32
chr11:18956989-18958835	-0.57	DR5	Meis1	-2.64
chr11:19012000-19025444	-0.54	DR1	Meis1	-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**	**nearest gene with altered expression in *Aldh1a2* KO**	**log2 fold change for nearest gene: RNA-seq for *Aldh1a2* KO vs WT**
chr18:38598986-38601292	-1.20	-	Spry4	3.43
chr5:147297983-147318733	-0.63	DR2	Cdx2	1.98
chr17:29080591-29082455	-0.59	DR2	Trp53cor1	1.91
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	-1.10
chr7:70356085-70361002	0.63	DR1	Nr2f2	-2.32
chr6:52156115-52158253	0.73	DR5, DR2	Hoxa1	-5.43
chr11:19015536-19017169	0.78	DR1	Meis1	-2.64
chr11:19007512-19012358	0.87	DR2	Meis1	-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 and Figs. S1 and S2.

Fig. 1

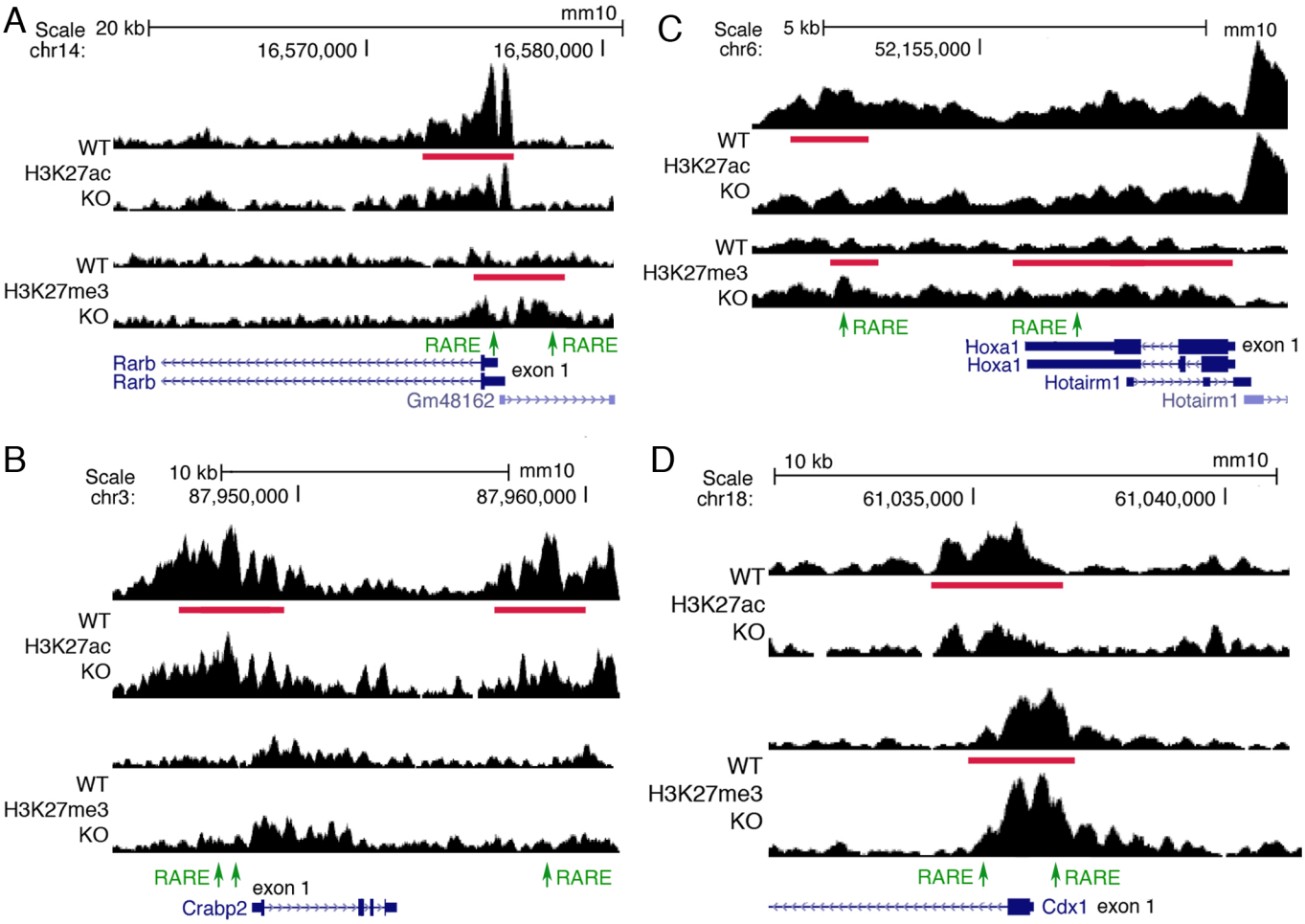


Fig. 1. ChIP-seq findings for *Rarb*, *Crabp2*, *Hoxa1*, and *Cdx1* showing that RA-regulated peaks for H3K27ac and H3K7me3 are located near known RARE enhancers. (A) Shown for *Rarb* are RA-regulated ChIP-seq peaks for H3K27ac and H3K27me3 (red bars) when RA is lost in E8.5 trunk comparing wild-type (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

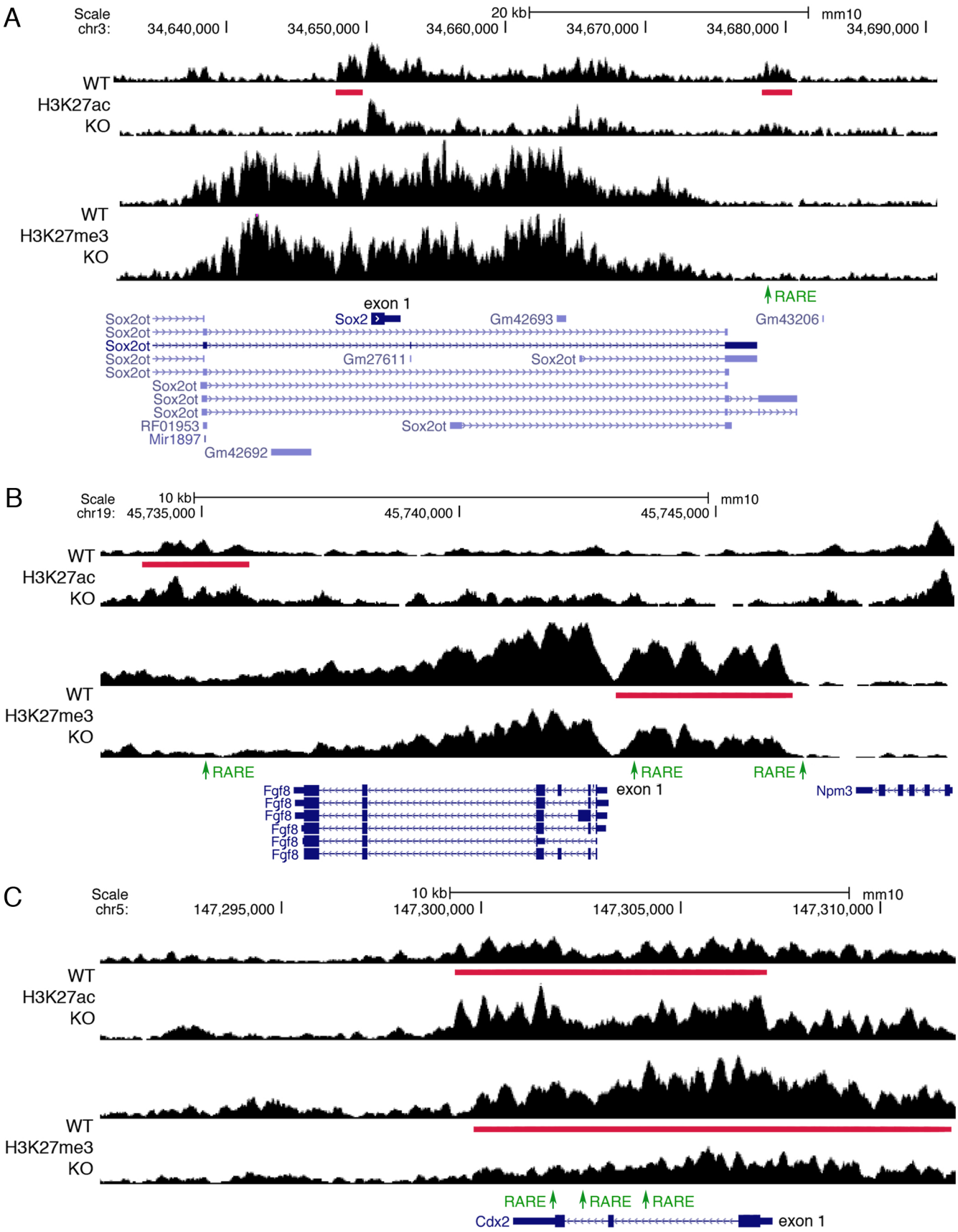


Fig. 2. ChIP-seq findings identify RAREs near *Sox2*, *Fgf8*, and *Cdx2* that regulate NMPs. See below.

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

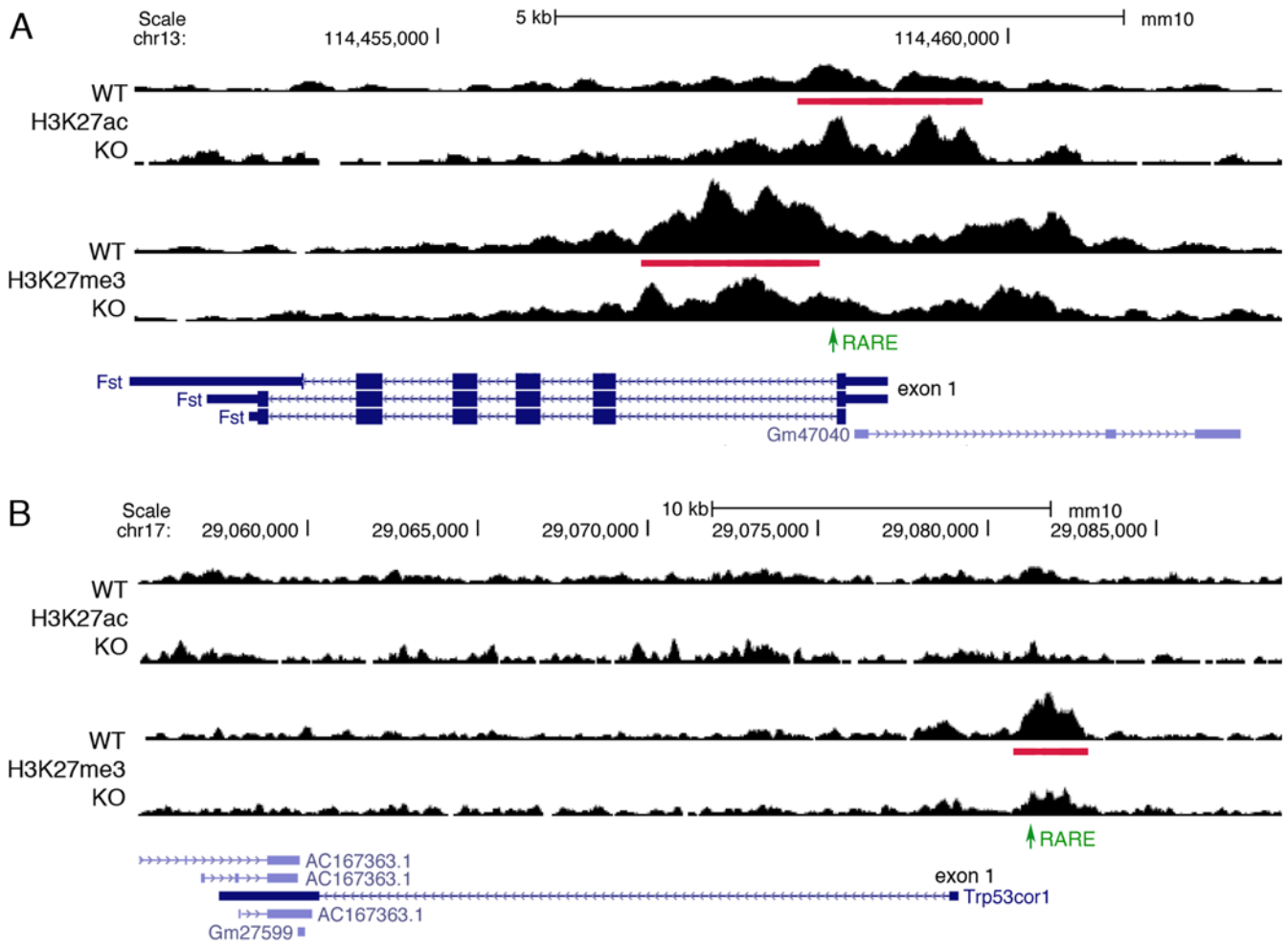


Fig. 3. ChIP-seq findings for *Fst* and *Trp53cor1* identify new RARE silencers. (A) *Fst* has nearby differential peaks (red bars) for both H3K27ac (increased) and H3K27me3 (decreased) when RA is lost in E8.5 trunk from *Aldh1a2-/-* (KO) compared to wild-type (WT). A RARE (green) contained within (or near) each peak is a candidate for a RARE silencer as H3K27ac is increased and H3K27me3 is decreased when RA is lost. (B) *Trp53cor1* has an H3K27me3 peak that is reduced when RA is lost, plus an associated RARE silencer candidate.

Fig. 4

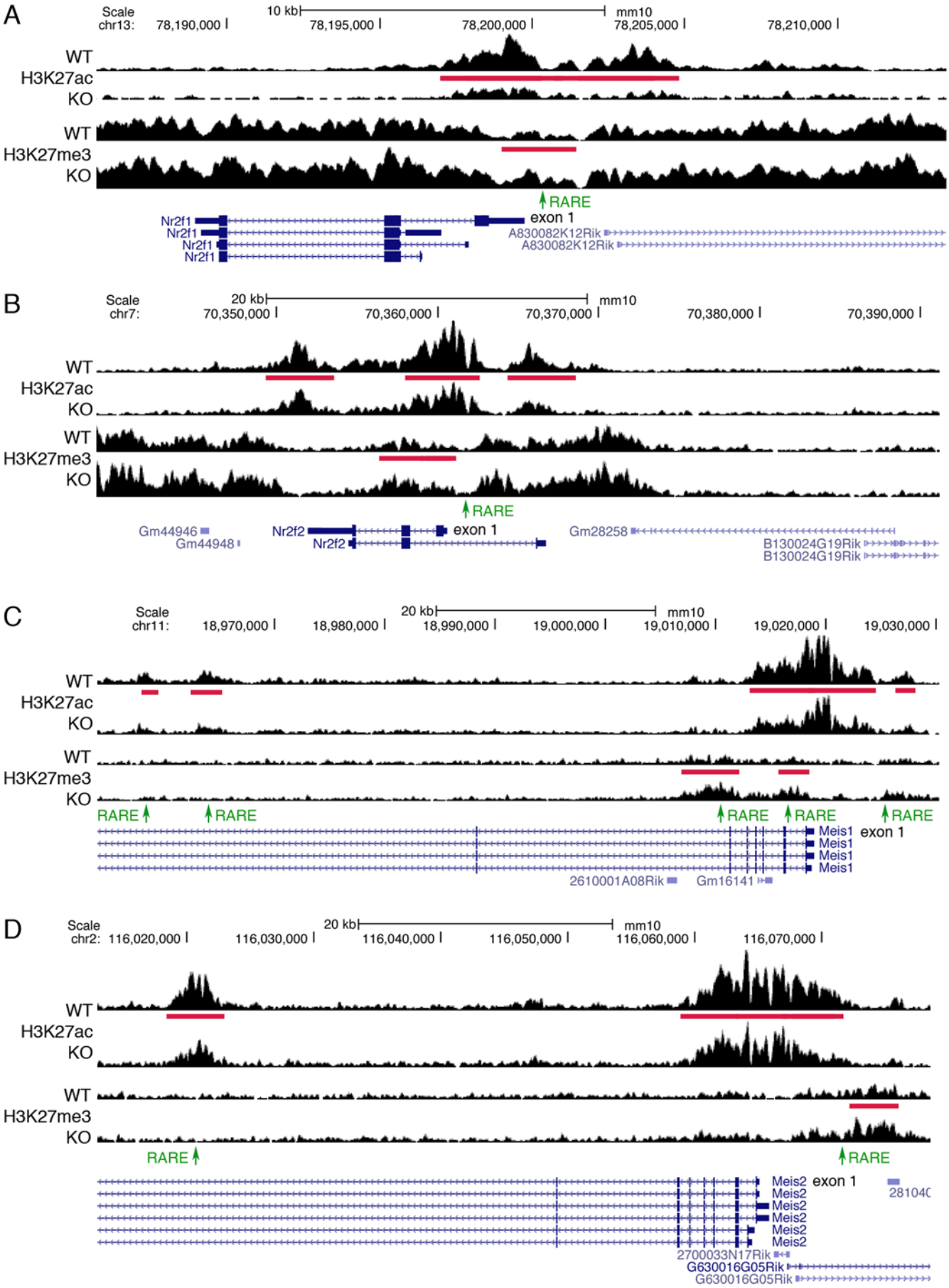


Fig. 4. ChIP-seq findings for *Nr2f1*, *Nr2f2*, *Meis1*, and *Meis2* identify RARE enhancers in gene families.
(continued on next page)

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

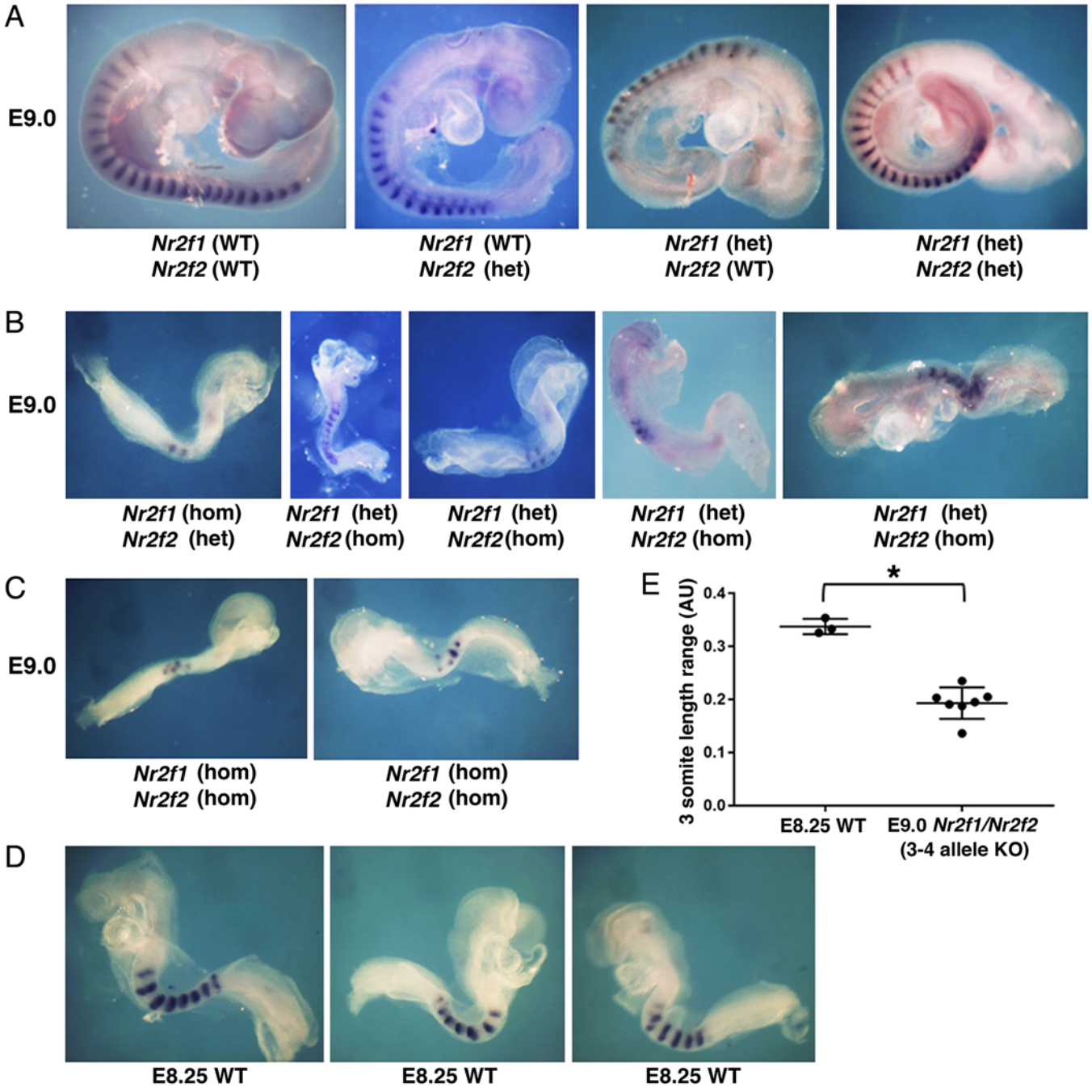


Fig. 5. *Nr2f1/Nr2f2* double mutants exhibit defects in body axis formation. (A) Embryos dissected at E9.0 carrying 0-2 knockout alleles for *Nr2f1* or *Nr2f2* have normal somites and body axis formation based on expression of the somite marker *Uncx*. (B-C) Embryos dissected at E9.0 and stained for *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

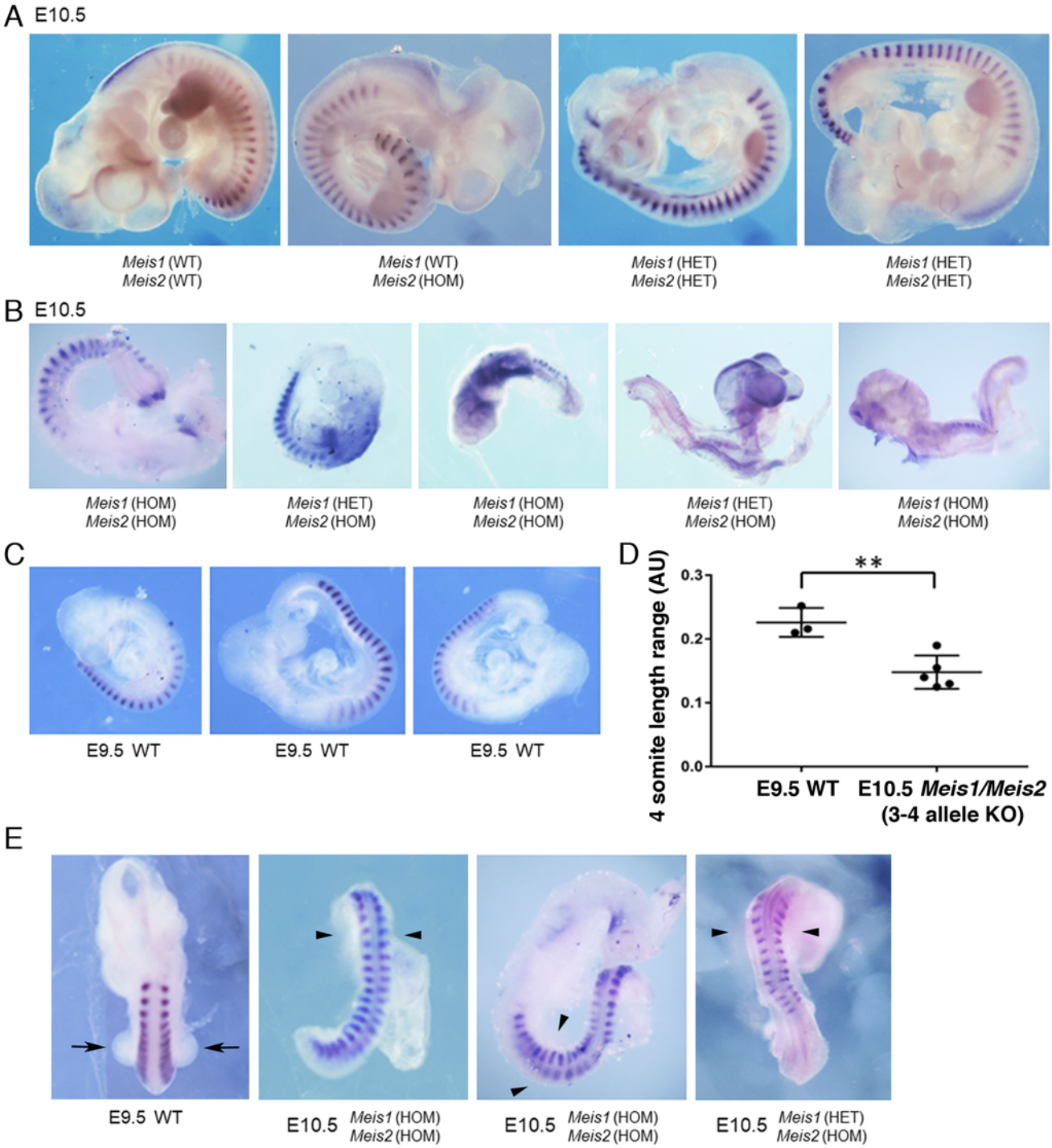


Fig. 6. *Meis1/Meis2* double mutants exhibit defects in body axis and forelimb formation. (A) Embryos dissected at E10.5 carrying 0-2 knockout alleles for *Meis1* or *Meis2* have normal somites and body axis formation based on expression of the somite marker *Uncx*; also, limb formation is normal. (B) Embryos dissected at E10.5 and stained for *Uncx* that carry 3 or 4 knockout alleles for *Meis1* or *Meis2* exhibit small somites and reduced body axis growth resembling the size of embryos at E9.5. (C) Wild-type (WT) E9.5 embryos stained for *Uncx* expression. (D) Comparison of somite size along the anteroposterior axis between E9.5 WT and E10.5 *Meis1/Meis2* 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; *Meis1/Meis2* 3-4 allele double knockout, $n = 5$ biological replicates. (E) Forelimb buds (arrows) normally observed in an E9.5 WT embryo are absent in E10.5 *Meis1/Meis2* knockout embryos with 3-4 knockout alleles (arrowheads).

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**	**nearest gene with altered expression in *Aldh1a2* KO vs WT**	**log2 fold change for nearest gene: RNA-seq for *Aldh1a2* KO vs WT**
chr13:78197222-78204291	-1.23	DR1	Nr2f1	-2.02
chr11:35545982-35547217	-1.22	-	Slit3	1.45
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	Exoc6b	3.14
chr4:149906238-149908597	-0.97	DR2	Spsb1	-1.19
chr8:14991303-14993068	-0.96	DR1	Arhgef10	-1.56
chr5:24321289-24322853	-0.94	DR1	Kcnh2	1.19
chr19:56029205-56030546	-0.93	-	Tcf7l2	1.22
chr11:68088685-68091257	-0.90	-	Stx8	-2.87
chr17:56468893-56471601	-0.90	DR5, 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
chr5:74239903-74242919	-0.83	-	Scfd2	1.11
chr6:5028587-5031108	-0.83	DR5, DR2	Ppp1r9a	-1.47
chr9:114798509-114801621	-0.81	-	Cmtm8	-2.12
chr4:149254383-149256741	-0.80	DR5	Kif1b	-1.72
chr5:36373121-36374553	-0.79	DR5, DR2	Sorcs2	-5.20
chr3:5387155-5389128	-0.79	DR5	Zfhx4 *	-2.26
chr17:30549676-30551173	-0.78	-	Btbd9	-1.18
chr1:74102745-74105259	-0.78	DR5, DR2	Tns1	0.96
chr3:5235978-5239655	-0.77	-	Zfhx4 *	-2.26
chr11:5128664-5130395	-0.76	-	Emid1	2.40
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
chr12:8912431-8914944	-0.64	-	Laptm4a	-1.97
chr18:83927674-83929824	-0.64	-	Tshz1	-1.33
chr14:14069903-14072065	-0.64	-	Atxn7	-1.50
chr7:130367064-130369885	-0.64	-	Fgfr2 *	-3.06
chr5:111242680-111244777	-0.63	DR5, DR1	Ttc28	-1.87
chr17:66410064-66414806	-0.63	DR5, DR1	Mtcl1	-1.32
chr4:148106909-148109328	-0.63	DR5	Draxin	-3.43
chr14:16571405-16576397	-0.63	DR5	Rarb *	-1.64
chr16:74395975-74399535	-0.63	DR1	Robo2	-1.22
chr1:38435452-38439196	-0.62	DR5, DR2	Aff3	-3.86
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	DR5, 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

chr7:70348715-70369942	-0.57	DR1	Nr2f2 *	-2.32
chr11:18956989-18958835	-0.57	DR5	Meis1 *	-2.64
chr7:130181974-130183270	-0.56	-	Fgfr2 *	-3.06
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	DR5, DR2	Ctbp2	-0.86
chr18:53463017-53465407	0.53	-	Prdm6	0.93
chr19:45733505-45735997	0.53	DR1	Fgf8 *	5.24
chr16:44378768-44382520	0.56	DR5, DR2	Spice1	-3.66
chr11:54891361-54894784	0.57	DR1	Gpx3	2.58
chr10:81166840-81168973	0.58	-	Pias4	-3.38
chr4:86669294-86671201	0.59	DR2	Plin2	2.29
chr14:61054604-61059202	0.60	-	Tnfrsf19	-1.18
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
chr10:29532842-29536192	0.65	-	Rspo3	-1.03
chr11:57831171-57833707	0.66	-	Hand1	1.46
chr6:52310576-52314619	0.71	DR5, DR1	Evx1	1.54
chr13:114456392-114460659	0.72	DR2	Fst *	1.15
chr5:147298587-147311126	0.73	DR2	Cdx2 *	1.98
chr11:31369048-31371820	0.73	-	Stc2	-2.17
chr5:53106977-53110254	0.74	DR2, DR1	Sel1l3	3.67
chr10:59957002-59959223	0.75	-	Ddit4	3.33
chr5:107216257-107218023	0.77	-	Tgfbr3	1.37
chr8:111875922-111876960	0.78	-	Tmem170	-2.26
chr13:116293452-116295196	0.85	-	Isl1	-1.25
chr1:84331820-84334081	0.99	-	Pid1	-3.29
chr5:104021158-104022631	1.000	-	Hsd17b11	3.89
chr4:44251608-44253515	1.01	-	Rnf38	-1.05
chr4:33269655-33271170	1.02	-	Pnrc1	-1.37
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 more than one differential H3K27ac peak nearby are highlighted in yellow. Genes that have differential peaks for both H3K27ac and H3K27me3 (Table S3) are marked with an asterisk. Some genes respond opposite to what is expected (discordant): red values refer to genes downregulated when H3K27ac is increased; green values refer to genes upregulated when H3K7ac is decreased. 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**	**nearest gene with altered expression in *Aldh1a2* KO vs WT**	**log2 fold change for nearest gene: RNA-seq for *Aldh1a2* KO vs WT**
chr18:38598986-38601292	-1.20	-	Spry4	3.43
chr17:15533901-15538178	-0.95	DR5	Pdcd2	2.22
chr17:15533901-15538178	-0.95	DR5	Tbp	2.94
chrX:104569467-104572914	-0.90	DR5	Zdhhc15	1.5
chrX:104546302-104549188	-0.89	-	Zdhhc15	1.5
chr2:118901989-118904591	-0.85	-	Bahd1	1.12
chr4:129226495-129228276	-0.85	DR5	C77080	0.98
chrX:104281534-104284135	-0.78	-	Abcb7	-1.35
chr4:129221927-129223300	-0.77	-	C77080	0.98
chrX:169610313-169612750	-0.76	-	Mid1	-5.21
chr5:15980910-15984533	-0.75	-	Cacna2d1	1.01
chr3:89278454-89282169	-0.72	-	Efna1	1.23
chrX:169601660-169604571	-0.69	-	Mid1	-5.21
chrX:169914954-169917261	-0.67	-	Mid1	-5.21
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	DR5, DR2	Tnfrsf1a	1.39
chrX:169904958-169910424	-0.63	-	Mid1	-5.21
chr12:54201904-54203715	-0.63	DR5	Egln3	3.46
chr5:147297983-147318733	-0.63	DR2	Cdx2 *	1.98
chr7:4150034-4153470	-0.63	-	Cdc42ep5	-6.48
chrX:104536138-104539780	-0.62	DR2	Zdhhc15	1.50
chr7:27354380-27357877	-0.62	DR1	Shkbp1	-1.96
chr4:98726175-98729089	-0.61	-	L1td1	1.50
chr17:29080591-29082455	-0.59	DR2	Trp53cor1	1.91
chr5:136580307-136583065	-0.54	DR5	Cux1	-1.53
chr10:60828600-60835032	-0.49	DR5	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	DR5	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
chr6:65669463-65675294	0.50	-	Ndnf	1.02
chr19:4709301-4715701	0.51	DR5, DR2	Sptbn2	-1.59
chr3:5219339-5224436	0.52	-	Zfhx4 *	-2.26
chr10:86838475-86844777	0.53	DR5, DR1	Stab2	9.70
chr11:90004468-90012858	0.54	DR5, DR2	Pctp	0.88
chr7:130260543-130263682	0.56	-	Fgfr2 *	-3.06
chr17:56471489-56479605	0.57	DR5, DR1	Ptprs *	-3.31
chr8:95698227-95701503	0.58	-	Ndrg4	1.17
chr4:144893360-144895562	0.59	-	Dhrs3	-1.11

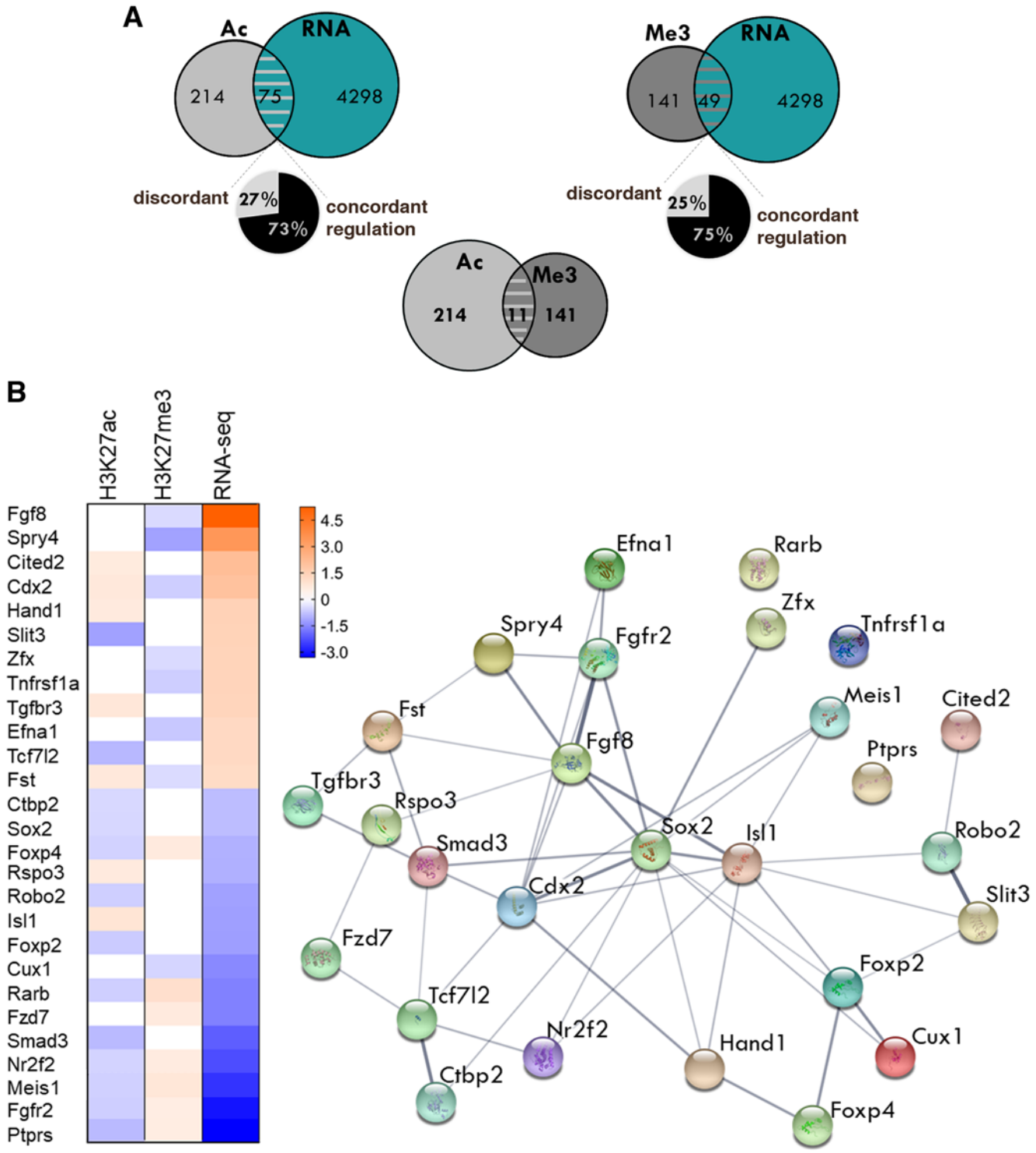
chr10:86846149-86852177	0.59	-	Stab2	9.70
chr11:95009493-95013361	0.60	DR2, DR1	Samd14	3.64
chr2:116072251-116077455	0.61	DR5	Meis2 *	-1.10
chr7:70356085-70361002	0.63	DR1	Nr2f2 *	-2.32
chr15:98621217-98623590	0.65	-	Cacnb3	-2.73
chr1:59473538-59476300	0.66	DR5, DR1	Fzd7	-1.64
chr17:47915056-47917877	0.67	-	Foxp4 *	-1.02
chr11:96351406-96353558	0.69	DR5	Hoxb2	0.94
chr10:8548515-8549917	0.72	DR5	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
chr19:3763564-3766537	0.76	DR5, DR2	Suv420h1	0.94
chr16:21204917-21206806	0.77	DR2	Ephb3	-5.25
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 more than one differential H3K27me3 peak nearby are highlighted in yellow. Genes that have differential peaks for both H3K27ac (Table S2) and H3K27me3 are marked with an asterisk. Some genes respond opposite to what is expected (discordant): red values refer to genes downregulated when H3K27me3 is decreased; green values refer to genes upregulated when H3K7me3 is increased. 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 differential peaks for H3K27ac and H3K27me3. The RAREs shown here contain no more than two mismatches to Homer consensus DR5, DR2, or DR1 RARE motifs shown at bottom.

Nearest gene with altered expression in *Aldh1a2* KO	**RARE DNA sequence 5'-3'**	**Type**	**Location**	**Genomic location (mm10)**
RARE ENHANCERS				
Cdx1	GGGTTA G GGGTCA	DR1	5'-noncoding	chr18:61036704-61036716
Cdx1	GGGTCA AG AGTTCA	DR2	intron 1	chr18:61035133-61035146
Crabp2	AGTTCA CC AGGTCA	DR2	5'-noncoding	chr3:87947525-87947538
Crabp2	AGGGCA G AGGTCA	DR1	5'-noncoding	chr3:87948035-87948047
Crabp2	AGGTCA GG AGGGCA	DR2	3'-noncoding	chr3:87958618-87958631
Hoxa1	GGTTCA CCGAA AGTTCA	DR5	3'-noncoding	chr6:52153426-52153442
Hoxa1	AGGTCA CT AAATCA	DR2	3'-untranslated	chr6:52156158-52156171
Meis1	ATGTCA G AGGTCA	DR1	5'-noncoding	chr11:19025437-19025449
Meis1	GGGTCA G AGGCCA	DR1	intron 1	chr11:19016387-19016399
Meis1	AGGGCA GG GGGCCA	DR2	intron 6	chr11:19010468-19010481
Meis1	AGGCCA CTGAG AGGTCA	DR5	intron 7	chr11:18963875-18963892
Meis1	ATGTCA A AGGACA	DR1	intron 7	chr11:18958299-18958311
Meis2	AGGTCA AAAAC AGTTCA	DR5	5'-noncoding	chr2:116071242-116071258
Meis2	CTCTCA AA GGGTCA	DR2	intron 7	chr2:116020707-116020720
Nr2f1	GTGTCA A AGTTCA	DR1	5'-noncoding	chr13:78200425-78200437
Nr2f2	GTGTCA A AGTTCA	DR1	5'-noncoding	chr7:70361772-70361785
Rarb	GGTTCA CCGAA AGTTCA	DR5	5'-untranslated	chr14:16575513-16575529
Rarb	AGGACA G AGGTCA	DR1	5'-noncoding	chr14:16578037-16578049
Sox2	GGGTCA GG AGGTCA	DR2	3'-noncoding	chr3:34679067-34679080
RARE SILENCERS				
Cdx2	GGGTCA CT GGGTGA	DR2	3'-untranslated	chr5:147301839-147301852
Cdx2	GGCTCA CA GTGTCA	DR2	intron 2	chr5:147302661-147302674
Cdx2	AGGTCA CT TGGTCA	DR2	intron 1	chr5:147303936-147303949
Fgf8	GGGTCA GC AGTTCA	DR2	5'-noncoding	chr19:45747045-45747058
Fgf8	AGGTCT CT GGGTCG	DR2	5'-noncoding	chr19:45743342-45743355
Fgf8	AGGGCA G AGGCCA	DR1	3' noncoding	chr19:45735030-45735042
Fst	GGGGCA GG GGTTCT	DR2	intron 1	chr13:114458455-114458468
Trp53cor1	AAATGA GG GGTTCA	DR2	5'-noncoding	chr17:29081305-29081318
DIFFERENTIAL PEAKS WITH NO RAREs (DNA elements that are indirect targets of RA signaling):				
Dhrs3	up-regulated by RA	H3K27me3 differential peak with no RAREs: chr4:144893360-144895562 (from intron 1 to intron 2)		
Pax6	up-regulated by RA	H3K27ac differential peaks with no RAREs: chr2:105674357-105677203 (within intron 1) chr2:105689278-105690982 (within intron 6)		
Spry4	down-regulated by RA	H3K27me3 differential peak with no RAREs: chr18:38598986-38601292 (within intron 1)		
RARE MOTIFS:	DR5	DR2	DR1	

Fig. S1

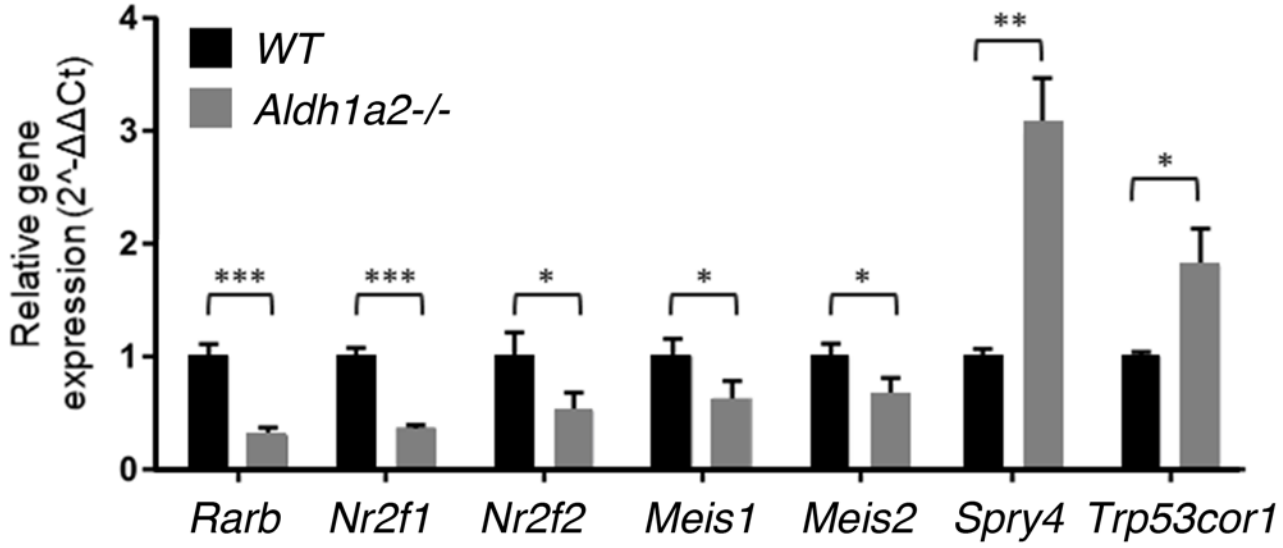


Supplemental Figure S1. Bioinformatic analysis of genes identified as RA target genes regulated at the transcriptional level. Related to Table 1.

(A) Venn diagram showing the number of genes that have both differential expression (RNA-seq) during RA deficiency and differential deposition of nearby H3K27ac and H3K27me3 marks determined by ChIP-seq.

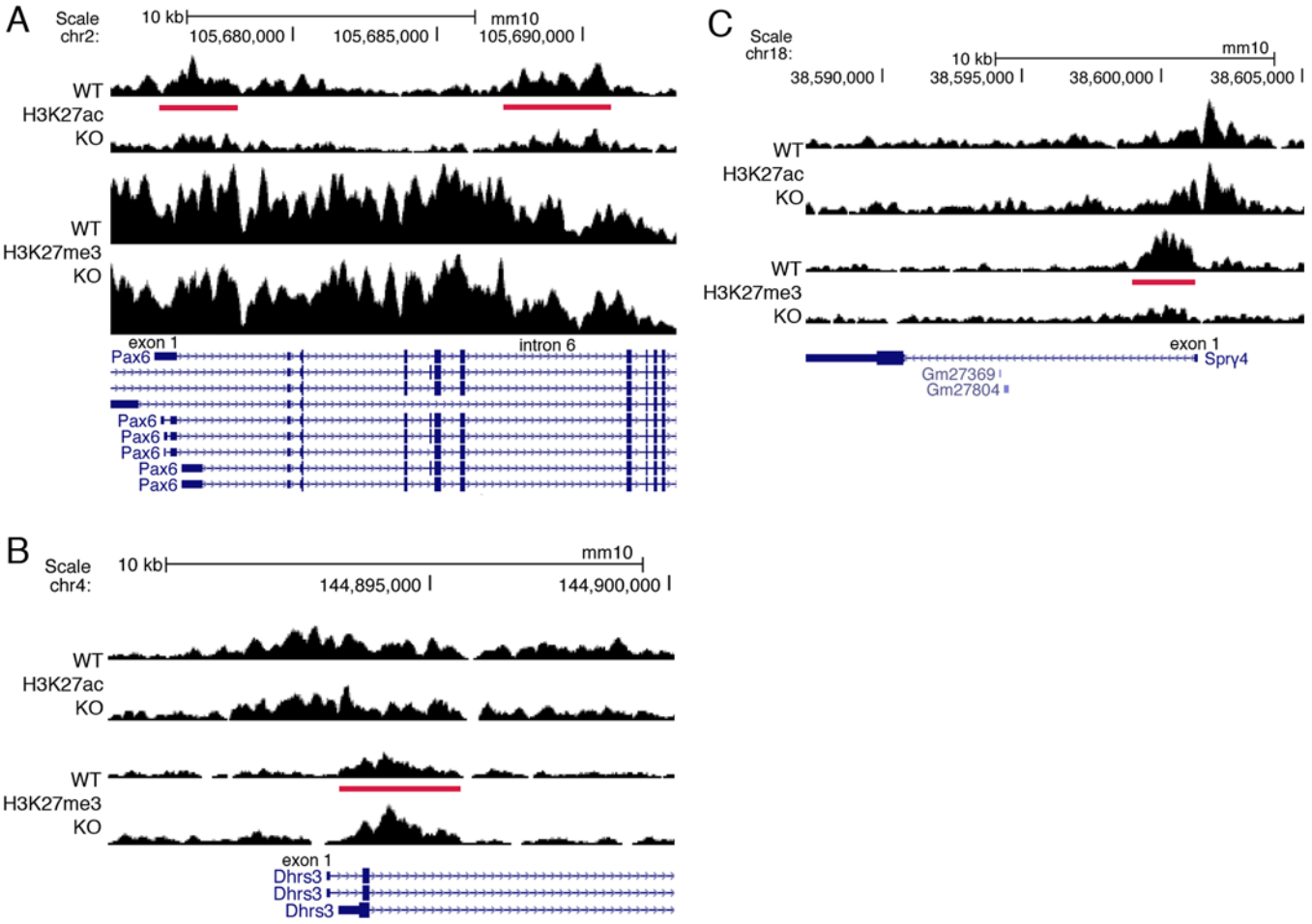
(B) IPA analysis showing that the RA target genes identified by our analysis are enriched for the pathway "Development of body trunk."

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.

Fig. S3



Supplemental Figure S3. ChIP-seq findings for *Pax6*, *Dhrs3*, and *Spry4* that lack RARE enhancers or silencers. These genes are good candidates for being indirect transcriptional targets of RA as none of the differential peaks contain RAREs.

(A-B) *Pax6* and *Dhrs3* have differential peaks (red bars) for H3K27ac (decreased) or H3K27me3 (increased), respectively, when RA is lost in E8.5 trunk from *Aldh1a2*-/- (KO) compared to wild-type (WT). The absence of RAREs associated with these differential peaks suggests that transcription of these genes is indirectly activated by RA.

(C) *Spry4* has a differential peak for H3K27me3 (decreased) when RA is lost, with no associated RAREs suggesting that transcription of *Spry4* is indirectly repressed by RA.