Characterising open chromatin in chick embryos identifies cis-regulatory elements important for paraxial mesoderm formation and axis extension

Gi Fay Mok¹, Leighton Folkes¹, Shannon A. Weldon¹, Eirini Maniou¹,³, Victor Martinez-Heredia¹, Alice M. Godden¹, Ruth M. Williams², Tatjana Sauka-Spengler², Grant N. Wheeler¹, Simon Moxon¹ & Andrea E. Münsterberg¹

Somites arising from paraxial mesoderm are a hallmark of the segmented vertebrate body plan. They form sequentially during axis extension and generate musculoskeletal cell lineages. How paraxial mesoderm becomes regionalised along the axis and how this correlates with dynamic changes of chromatin accessibility and the transcriptome remains unknown. Here, we report a spatiotemporal series of ATAC-seq and RNA-seq along the chick embryonic axis. Footprint analysis shows differential coverage of binding sites for several key transcription factors, including CDX2, LEF1 and members of HOX clusters. Associating accessible chromatin with nearby expressed genes identifies cis-regulatory elements (CRE) for TCF15 and MEOX1. We determine their spatiotemporal activity and evolutionary conservation in Xenopus and human. Epigenome silencing of endogenous CREs disrupts TCF15 and MEOX1 gene expression and recapitulates phenotypic abnormalities of anterior–posterior axis extension. Our integrated approach allows dissection of paraxial mesoderm regulatory circuits in vivo and has implications for investigating gene regulatory networks.
The partitioning of paraxial mesoderm into repetitive segments, termed somites, is a key feature of vertebrate embryos. During amniote gastrulation, mesoderm cells emerge from the primitive streak and migrate in characteristic trajectories to generate axial, paraxial and lateral plate mesoderm (LPM)\(^1\). Paraxial mesoderm is located on either side of the midline tissues, neural tube and notochord. As the body axis extends, it consecutively generates pairs of somites\(^3\) epithelial spheres comprised of multipotent progenitor cells. In response to extrinsic signals, epithelial somites (ES) undergo dramatic morphogenetic changes and reorganise\(^4\). On the ventral side cells undergo an epithelial to mesenchymal transition (EMT) to form the sclerotome, while on the dorsal side the cells in the dermomyotome remain epithelial. From the dermomyotome edges cells transition to form the myotome, in-between the sclerotome and dermomyotome\(^5\). Concomitantly with somite morphogenesis, the differentiation potential of somite cells becomes more restricted, with cells eventually becoming specified towards the lineages of the musculoskeletal system, including chondrocytes and skeletal muscle cells\(^6\). Overall the process of somitogenesis generates a spatiotemporal gradient of differentiation within the paraxial mesoderm along the embryonic body axis\(^3\).

In addition, somite derivatives exhibit regional differences depending on their anterior–posterior axial position. Regional identity is already established at gastrula stages and is controlled by the stepwise transcriptional activation of HOX gene expression\(^9\)-\(^11\). For example, members of the HOXB cluster are first activated in a temporal colinear fashion in prospective paraxial mesoderm, prior to ingress through the primitive streak\(^12\). The colinear activation of HOX genes culminates in nested expression domains within the paraxial mesoderm, thereby conferring regional identity along the axis\(^13\). To determine the structural features associated with colinear expression, the 3D organisation of HOX clusters has been investigated\(^10\). It has also been shown that posterior Wnt signalling and CDX transcription factors (TFs) are important regulators of the “trunk” HOX genes in the centre of HOX clusters\(^15\). In particular, CDX2 is essential for axial elongation with mutations leading to posterior truncations associated with changes in HOX expression domains\(^16\). CDX activity is associated with histone acetylation and mediates chromatin accessibility of regulatory elements\(^17\).

Superimposed onto regional differences is the control of cell identity and differentiation, and several well-characterised TFs serve as markers for musculoskeletal lineages. Chondrogenic cells express PAX1, PAX9 and SOX9 and dermomyotomal myogenic progenitors are characterised by PAX3 and PAX7. Committed myoblasts express MYF5 and MYOD, while MYOG and KHLH31 are markers for differentiated myocytes\(^18\)-\(^20\). Other transcriptional regulators that are important in paraxial mesoderm include TCF15 (Paraxis), a bHLH TF required for somite epithelialization\(^21\), CDX (Caudal), which is necessary for axis elongation\(^22\), and MEOX1, which is involved in somite morphogenesis, patterning and differentiation, particularly of sclerotome-derived structures\(^23\),\(^24\). In human, mutations of MEOX1 are found in patients with Klippel-Feil Syndrome, which is associated with fusion and numerical defects in the cervical spine as well as scoliosis\(^25\),\(^26\). Whilst the sequence of marker gene expression in paraxial mesoderm is well defined\(^19\),\(^20\), the epigenetic and genomic mechanisms that control these transcriptional programmes remain largely unknown. The identification of enhancers has improved through high-throughput sequencing assays and comparative genomic analysis, however, experimental validation of enhancer activity remains challenging. In this study, we assay spatiotemporal changes in both gene expression signatures and accessible chromatin that occur in differentiating paraxial mesoderm along the anterior–posterior axis. We define differentially accessible chromatin regions within HOX genes that are associated with regional identities. Footprint analysis shows differential occupancy and coverage of binding sites along the axis for several TFs, including HOXA10, HOXA11, CDX2, LEF1 and RARA. CDX2 and LEF1 are both involved in similar processes during axis extension. However, network analysis shows that CDX2 and LEF1 footprints are associated with different expressed genes and there is little overlap in the genes they interact with. Correlating accessible chromatin with nearby expressed genes identifies cis-regulatory elements (CREs). We focus here on enhancers located upstream of TCF15 and MEOX1 and validate these in vivo, using electroporation of fluorescent reporters into gastrula-stage chick embryos. Time-lapse imaging shows the onset of enhancer activation in paraxial mesoderm and mutation of candidate TF motifs or epigenome modification leads to loss of gene expression and phenotypic changes. The MEOX1 CRE is evolutionary conserved in amphibians and human. Altogether our data characterises the accessible chromatin and gene expression landscapes in paraxial mesoderm, at different stages of somite maturation.

**Results**

**Transcriptional profiling of developing paraxial mesoderm.** To conduct genome-wide transcriptome analysis during the spatio-temporal transition of paraxial mesoderm, we collected pre-somatic mesoderm (PSM), ES, maturing somites (MS) and differentiated somites (DS) from Hamburger–Hamilton stage 14 (HH14)\(^27\) chick embryos in triplicate (Fig. 1a). At this stage, the five most posterior somites are epithelial, but in MS cells in the ventral part undergo EMT, the dorsal dermomyotome lip forms in the epaxial domain adjacent to the neural tube and myogenic cells begin to transition into the early myotome. Differentiating somites are compartmentalised, with a primary myotome beneath the dermomyotome and a sclerotome ventrally\(^5\),\(^28\).

After harvesting, tissues were processed for RNA sequencing (RNA-seq) (Fig. 1). Principal component analysis (PCA) showed that PSM, ES, MS and DS samples cluster into three distinct groups, with MS and DS samples clustering together (Supplementary Fig. 1a). Differential gene expression analysis comparing PSM and ES revealed up-regulation of 713 genes and down-regulation of 583 genes; comparing ES and MS revealed up-regulation of 145 genes and down-regulation of 155 genes; and comparing MS and DS revealed up-regulation of 53 genes and down-regulation of 26 genes. Comparisons between samples confirmed that the greatest differential was observed between PSM and any of the somite samples, followed by the number of differentially expressed genes between the most recently formed somite epithelium and the most differentiated somites (ES versus DS) (Supplementary Fig. 1b).

Previously described somite TFs, such as NKX6-2, NKX3-2, ZIC1 and HESS, were highly enriched in ES compared to PSM, as well as the gap junction protein GJA5 (Connexin 40). Marker genes important for myogenic (MyoD1 and ACTC1) and chondrogenic (PAX9, FST) cell lineages were enriched in MS compared to ES. Markers for chondrocytes (Chondromodulin, CNMD), bone homeostasis (Leucine-rich repeat containing, LRRC17) and cartilage (Keratan sulphate proteoglycan Keratocan, KERA) were identified (Fig. 1b–d). Myogenin (MYOG), a TF involved in differentiation of muscle fibres was enriched in DS compared to MS, as was expression of the neural crest cell (NCC) TF SOX10, due to NCCs migrating through the rostral half of differentiating somites. Other genes highly expressed in DS include the serine protease inhibitor, SPINK5; Troponin T2 (TNNT2) and Myomesin (MYOM1), encoding important proteins of the contractile sarcomere; CREB, a negative regulator...
of the endoplasmic reticulum stress response and ZFPM2, a zinc finger TF (Fig. 1d).

The functional clustering by gene ontology (GO) terms of differentially expressed genes across all four stages reveals enrichment of biological processes involved in cell differentiation in DS and MS versus PSM and ES samples. Genes involved in myoblast differentiation, cartilage condensation, skeletal muscle fibre development and myotube cell development were up-regulated (Fig. 1e). Further analysis of genes involved in positive regulation of myoblast differentiation shows that they display dynamic expression across the four groups and include classic markers for different stages of paraxial mesoderm differentiation. Genes differentially expressed in somite compartments include in the dermomyotome and myotome: MYF5, MYF6 and MYOD1, whilst MYOG and KLHL31 are associated with differentiated muscle. Classic markers for chondrogenesis and cartilage condensation within the sclerotome include the TFs, SOX9, PAX1 and PAX9, and the extracellular matrix component, COL11A1 (Fig. 1f). Functional clustering of differentially expressed genes also reveals enrichment of signalling pathways involved in anterior–posterior pattern formation. These pathways are expressed in an opposing fashion and include the FGF and Wnt signalling pathways, which are highly expressed in PSM and the retinoic acid (RA)
signalling pathway, which is more highly expressed in somite samples (Fig. 1g).

We next used weighted gene co-expression network analysis\(^29\) to characterise gene co-expression clusters across the four samples of the top 400 differentially expressed genes. We identified 11 clusters based on k-means clustering. The heat map shows the gene expression levels across the four different samples, and the t-SNE plot illustrates the dimensional distribution of the different clusters (Fig. 1h, i). The top three GO terms associated with clusters include “anatomical structure morphogenesis” (Supplementary Fig. 1). Clusters B and I comprise genes that increase or decrease in expression across the spatiotemporal series, from PSM to DS (Fig. 1, k). Cluster I features components of FGF (FGF13, FGF7, FGF10, FGF18, SPRY1), BMP (BMP2, BMP4, BMPER, NOG) and WNT (WNT8C, WNT5B) signalling pathways in addition to classic PSM markers such as MESP2, RIPPLY2, MSGN1 and MESPI, known to be important for somitogenesis. Cluster B features markers of cellular differentiation programmes such as ZIC1, ZIC4, MEOX1, TCF15 and include the myogenic regulatory factors, MTOD1 and MYF5.

**Profiling chromatin accessibility dynamics in paraxial mesoderm along the anterior–posterior axis.** To identify genomic regulatory elements that control paraxial mesoderm and somite differentiation programmes, we used Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)\(^31\). This mapped chromatin accessibility across the paraxial mesoderm along the axis, in PSM, ES, MS and DS (Fig. 1a). Distinct chromatin accessibility profiles were evident at different stages of somite development, indicative of the dynamic progression of axial development. PCA showed a high reproducibility between biological triplicates of each sample type (Supplementary Fig. 2–g–i), but dynamic changes in chromatin accessibility were observed between them. Using DiffBind\(^32\),\(^33\) we show the densities and clustering of differentially accessible chromatin regions (peak sites) (rows), as well as the sample clustering (columns) for PSM against ES, ES against MS, and MS against DS. We identified differentially accessible peaks with differential densities showing clusters of peak sites with distinct patterns of chromatin accessibility levels for PSM against ES (Fig. 2a), ES against MS (Fig. 2b), and MS against DS (Fig. 2c). MA plots show the highest number of differentially accessible peaks is evident when comparing PSM and ES (\(n = 27,692\), Fig. 2d). The number of differentially accessible peaks is lower when comparing ES against MS, and MS against DS (\(n \approx 4670, n = 1965\), Fig. 2e, f). This is in line with the transcriptome data, where greater differences were seen between PSM and ES compared to the differences observed between different stages of somite maturation.

The genomic distribution of accessible regions was similar in all four sample types: between 39 and 42% were in intergenic regions, ~10% were in introns, 0.5% in exons, 2–3% at the TSS and 43–46% of accessible regions were within a 50 kb region upstream of the TSS which includes the promoter (Fig. 2g). Functional terms associated with predicted TF binding sites that were enriched in accessible peaks in DS compared to PSM included cell fate specification and terms related to morphogenesis or skeletal myogenesis (Fig. 2h). Consistent with the latter, we identified >200 binding sites for myogenin (MYOG) that are located within accessible chromatin peaks within 2 kb of genes differentially expressed in DS, where skeletal muscle differentiation occurs (Fig. 2i). The MYOG motif is well conserved across mouse and human, thus is likely to be conserved across avian species also. Other enriched TFs identified include bHLH proteins (TCF12, ASCL1, ARNT1), of which TCF12 is expressed in skeletal muscle, is part of the canonical Wnt pathway and implicated as a transcriptional repressor in colorectal cancer\(^34\). Specificity proteins, Sp1, Sp2, Sp3 and Sp8, are zinc finger proteins known to interact with bHLH proteins such as MyoD\(^35\).

Furthermore, Sp8 is a downstream effector of the Wnt pathway in neomesodermal stem cells\(^36\). Sp1 and Sp3 bind to GC and GT boxes and can be displaced from these sequences by KLF16, a Krüppel-like zinc finger protein for which binding sites are also enriched and increased in PSM (Fig. 2j). Other zinc finger TFs include ZNF384, ZNF740 and ZNF263, which are involved in the regulation of cell differentiation genes including those relevant to musculoskeletal development. For example, ZNF384 regulates extracellular matrix genes MMP1, MMP3, MMP7 and COL1A1\(^37\); ZNF740 recruits the chromatin regulator HDAC1 to the SMAD4-DNA complex and prevents the recruitment of the transcriptional activators CREBBP and EP300\(^38\). The binding motifs for ZNF740 are increased in PSM (Fig. 2j). ZNF263 is involved in adipogenesis\(^39\). The Ewing sarcoma RNA binding protein 1 (EWSR1) regulates gene expression, cell signalling, RNA processing and transport. Chimeric proteins resulting from chromosomal translocations between EWSR1 and various TF genes\(^40\),\(^41\) are involved in tumorigenesis such as Ewing sarcoma in bones and bone connective tissues. Furthermore, the binding motif for retinoic acid receptors (RXRA) is enriched. Enrichment analysis of differentially accessible regions identified additional TF motifs that were increased in number in either PSM (Fig. 2j) or DS (Fig. 2k). In PSM this included motifs for TFP2C/TFP2B and ZIC3/ZIC4, whose functions in axial elongation and/or musculoskeletal development are currently unknown. In DS, this included motifs for FOXO1/FOXO3 and MEOX1.

**Identification of differential footprints during somite development.** To further interrogate the accessible chromatin landscape during somite development, we used HINT-ATAC\(^42\) to discover differential TF footprints in regions of open chromatin identified in PSM, ES, MS or DS. Initially we focussed on the CDX2 TF, which is a readout for posterior WNT signalling and has been implicated in defining neomesodermal progenitors (NMP)\(^43\). CDX2 is essential for axial elongation\(^22\) and is highly expressed in the PSM (Fig. 3a). Consistent with high levels of WNT signalling activity in the PSM, HINT-ATAC identified a greater number of CDX2 footprints in open chromatin in this region when compared to ES, MS and DS (Fig. 3b–d). Similarly, LEF1, a transcriptional effector for canonical WNT signalling, is highly expressed in the PSM (Fig. 3e). LEF1 is also expressed in somites, where it becomes restricted to the myotome\(^44\),\(^45\). We identified a greater number of LEF1 footprints in the PSM when compared to any of the somite samples, consistent with the more restricted expression of LEF1 in the latter (Fig. 3f–h).

A reverse coverage pattern was observed for TFs involved in somite differentiation. For PAX3, an important TF that regulates the myogenic programme and highly expressed during somite development, HINT-ATAC revealed an increase in the number of PAX3 footprints in ES open chromatin when compared to PSM. The number of PAX3 footprints increased further in MS and DS when compared to PSM (Supplementary Fig. 3a–d) suggesting there is a greater coverage of bound sites in maturing and differentiating somites consistent with the role of PAX3 in myogenic progenitors in the dermomyotome. Another key somite TF, TWIST2 (also known as DERM01), is important for EMT during somitogenesis. TWIST2 was highly expressed in the paraxial mesoderm and expression increased as somites differentiate (Supplementary Fig. 2e). The number of genome-wide TWIST2 footprints were very similar in PSM and ES...
(Supplementary Fig. 2f), however, the number of footprints increased in MS and DS when compared to PSM (Supplementary Fig. 2g, h). Retinoic acid receptor alpha (RARA) is a nuclear receptor, which acts as a transcriptional repressor in absence of ligand but a transcriptional activator when RA is present (see[46] for review). RARA is highly expressed in somites (Supplementary Fig. 3i). HINT-ATAC identified fewer RARA footprints in PSM compared to MS and DS (Supplementary Fig. 3k, l).

The inverse coverage patterns identified for CDX2, LEF1 versus RARA were consistent with the opposing expression patterns for WNT and RA pathway components within paraxial mesoderm along the anterior–posterior axis (Fig. 1g). To further dissect the
roles of CDX2 and LEF1 in posterior axis elongation we determined genes associated with either CDX2 or LEF1 footprints in accessible regions within 10 kb upstream or downstream. GO terms for these genes were overlapping and include: anatomical structure morphogenesis/development, metabolic process and regulation, for both CDX2 (Fig. 3) and LEF1 (Fig. 3). We next performed STRING analysis, using a threshold of 0.700, to obtain a protein–protein interaction (PPI) map for genes identified with CDX2 (Fig. 3k) or LEF1 (Fig. 3l) footprints in accessible regions within 10 kb. This revealed genes with strong PPIs, including those associated with enriched biological processes such as embryonic morphogenesis for CDX2 and animal organ morphogenesis for LEF1. LEF1 footprints correlated with CDX2 consistent with CDX2 being regulated by the Wnt signalling pathway. Furthermore, the phenotypical traits of CDX2 mouse mutants include posterior truncations reminiscent of those found in LEF1/TCF1 double mutants. Thus, to explore whether CDX2 and LEF1 could regulate similar genes, we examined all differentially up-regulated genes in the PSM and investigated whether there are associated CDX2 and LEF1 footprints. We found 101 genes with a CDX2 footprint and 42 genes with a LEF1 footprint within 10 kb up- or downstream. Surprisingly, when comparing these sets of genes only four genes—Msgn1, Sal2, Spry1 and DDC—were associated with both CDX2 and LEF1 footprints, and the majority of correlated genes was different (Fig. 3m). Our analysis suggests that CDX2 and LEF1 are part of discrete networks acting in parallel to govern similar processes (Fig. 3k, l), but they regulate different sets of genes important for these processes.

**ChromatIN accessibility and differential TF footprints in the HOX A cluster.** We next examined the HOXA cluster, one of four HOX gene clusters imposing regional identity along the anterior–posterior axis via the colinear expression of its members. We determined how HOXA gene expression patterns correlate with the accessible chromatin landscape. RNA sequencing determined expression levels of each member of the HOXA cluster in PSM, ES, MS and DS (Fig. 4a). Their expression reflects the organisation of the genes within the cluster: the more 3′ located genes have a more anterior expression boundary compared to the genes located more 5′, which are restricted more posteriorly. Accordingly, we find that HOXA1, HOXA2, HOXA3, HOXA4, HOXA5 and HOXA6 are all highly expressed across the length of the axis: in PSM, ES, MS and DS. A small decrease in HOXA7 gene expression was detected in DS, with more pronounced decreases observed for HOXA9, HOXA10, HOXA11 and HOXA13, which were also reduced progressively in MS and ES. The colinear pattern of gene expression correlated with differentially accessible chromatin within the HOXA cluster (Fig. 4b). Accessible chromatIN regions were seen in PSM, ES, MS and DS near the promoter of HOXA1, HOXA2, HOXA3, HOXA4, HOXA5 and HOXA6. However accessible chromatIN for HOXA7 was reduced at the promoter in DS compared to PSM, ES and MS. For the more posteriorly restricted genes, HOXA9, HOXA10, HOXA11 and HOXA13 accessible chromatIN peaks were reduced in ES, MS and DS compared to PSM, which correlated with their reduced expression. We demonstrate the same relationship between gene expression and chromatIN accessibility along the anterior–posterior axis across the HOXB, HOXC and HOXD clusters (Supplementary Fig. 4a–f). In the HOXA cluster we identified footprints within accessible regions in intergenic regions. Notably, we identified footprints for TFs involved in patterning along the anterior–posterior axis, including footprints upstream of HOX1/2, LEF1 and for members of the HOX clusters themselves, as well as for some of the TFs with enriched motifs in accessible regions such as RFXA, TFAP2B/C, SP1, SP2, ZIC1/3, FOXO1/4, ZNF263 (Figs. 2i–k and 4b). To investigate the impact of the dynamic changes in HOXA gene expression along the anterior–posterior axis, we next explored the number of TF footprints for HOXA2, HOXA5, HOXA10 and HOX11 in PSM and DS (Fig. 4c–f). We observed the same number of footprints for HOXA2 and HOXA5 when comparing PSM and DS, however, a significant decrease in coverage was detected for HOXA10 and HOXA11 footprints in anterior DS compared to PSM. This reveals a strong association between gene expression levels along the anterior–posterior axis and the genome-wide coverage of HOXA binding sites.

**Identification and validation of paraxial mesoderm-specific regulatory elements.** Next, we identified differentially accessible peaks that were open specifically in PSM or in somite samples, ES, MS or DS. We hypothesise that these could represent putative enhancers. For example, differentially accessible peaks identified flanking genes highly expressed in the PSM included a peak downstream of MSGN1 present in PSM and not in ES, MS or DS (Supplementary Fig. 5a); a peak downstream of WNT8C and a peak within intron 1 present in PSM but not in somite tissues (Supplementary Fig. 5b); and peaks upstream and downstream of FGFR4 present in PSM and low or absent in somite tissues (Supplementary Fig. 5c). For the muscle differentiation gene, MYOG, a peak was identified upstream of the gene in DS, MS and interestingly also in ES, but not in PSM (Supplementary Fig. 5d). For RHD10, which is associated with RA signalling and highly expressed in somites but less abundant in PSM, a differential peak was identified in ES, MS and DS and not in PSM (Supplementary Fig. 5e). Similarly for GREM1, an antagonist of BMP signalling highly expressed in developing somites, a differential peak present in all somite samples but not in PSM was identified downstream of the gene (Supplementary Fig. 5f). In most cases, chromatIN accessibility correlated well with gene expression and in some cases it preceded transcript detection, e.g. MYOG, or high level gene expression, e.g. TCF15 (see below). The putative enhancer activities of these differential peaks remain to be confirmed experimentally, however, we validated and further characterised some somite enhancers by embryo electroporation. We focussed on TCF15 and the homeodomain TF, MEOXI, two classic markers identified in the group of genes that increased during the differentiation of paraxial mesoderm and somites (Fig. 1j; Cluster B). In addition, MEOXI binding motifs were increased in accessible regions in DS (Fig. 2k).
Fig. 3 Differential footprints identified for CDX2 and LEF1 during somite development.  

a Gene expression from mRNA-seq (error bars = SEM, n = 3) for CDX2.  

b Tn5 insertion frequency across all accessible regions containing at least one CDX2 motif, at nucleotide resolution in PSM, ES, MS and DS reveals the presence of a footprint centred on the CDX2 motif. Differential footprinting for CDX2 motif comparing PSM and ES, c PSM and MS, and d PSM and DS.  

e Gene expression for LEF1 (error bars = SEM, n = 3).  

f Differential footprinting for LEF1 motif comparing PSM and ES, g PSM and MS, and h PSM and DS.  

Similar GO terms for genes associated with (i) CDX2 footprints and (j) LEF1 footprints. For enriched GO terms, p values were obtained from a modified Fisher exact test.  

k, l Protein–protein network analysis using STRING database. Interactions between genes identified with (k) CDX2 and (l) LEF1 footprints in an accessible region, within 10 kb upstream or downstream. Highlighted in red are genes correlated with embryonic morphogenesis in CDX2-associated genes and animal organ morphogenesis in LEF1-associated genes.  

m Venn diagram of CDX2-associated genes against LEF1-associated genes identified only four common genes—Msgn1, Sall4, Spry1 and DDC.
We examined open chromatin peaks flanking the TCF15 and MEOX1 genes within 10 kb. Identified peaks representing candidate CREs were cloned upstream of the herpes simplex virus thymidine kinase (HSV-TK) minimal promoter, driving expression of a stable Citrine reporter. Electroporation targeted the prospective mesoderm of gastrula-stage HH3+ embryos (Supplementary Fig. 5g, h), and reporter gene expression profiles were monitored until HH11.

We identified two CREs upstream of TCF15 (Fig. 5a). These sequences are chick specific and the fluorescent enhancer reporters showed spatially restricted activities. For the first element, TCF15 Enh-1 (1500 bp) we observed activity in the PSM, in all somites and in the notochord (Fig. 5b). The second element, TCF15 Enh-2 (700 bp), showed activity mainly in PSM and somites, as well as some activity in LPM (Fig. 5c, e). In situ hybridisation showed that expression of TCF15 was restricted to PSM and somites (Fig. 5g) and it is not clear at present why TCF15 Enh-1 and TCF15 Enh-2 drive ectopic reporter expression also in the notochord and LPM. Because reporter activity observed with TCF15 Enh-2 reflected more closely the spatial gene expression pattern of TCF15, we next sought to identify TFs that regulate this element. HINT-ATAC identified a TF footprint for RARA within TCF15 Enh-2, consistent with RARA expression and coverage of binding sites across the anterior–posterior axis (Supplementary Fig. 3i–l). Introducing mutations into the RARA binding site (Fig. 5a) led to loss of reporter activity in the embryo (Fig. 5f), suggesting RARA is indeed required to activate TCF15 Enh-2. To determine the potential significance of RARA-mediated regulation of TCF15 Enh-2 in vivo, we used the conventional dCas9-KRAB repressor to modify the endogenous enhancer. Two CRISPR guide RNAs (gRNA) designed to target the repressor to the TCF15 Enh-2 RARA binding site, or scrambled gRNA controls were electroporated together with dCas9-KRAB (Fig. 5g, h). Detection of TCF15 expression by in situ showed that epigenome modification of the endogenous TCF15 Enh-2 alone led to reduced TCF15 expression and conversely a drastic truncation of the body axis (n = 6/8 embryos), whilst control scrambled gRNAs/dCas9-KRAB repressor has no effect on TCF15 expression or axis.
These data suggest that RA signalling is crucial for TCF15 gene expression as RARA binding site perturbation led to loss of reporter activity and epigenome editing of the endogenous CRE resulted in disruption of anterior–posterior axis elongation. This is consistent with mouse mutants of TCF15 or mutants affecting RA signalling, in which ES formation is disrupted and the embryonic axis truncated21,53,54. Whilst it has been shown that Wnt signalling is important for TCF15 expression in early somites55, there is no evidence of direct regulatory interactions. However, we cannot exclude the possibility that Wnt signalling, via LEF/β-catenin, contributes to TCF15 expression potentially via a different CRE, which alone is not sufficient. It is also worth noting that TCF15 Enh-1 and Enh-2 are not conserved across mammalian species.

MEOX1 is an important TF for early somite patterning and differentiation23,24. We examined accessible regions of chromatin and selected an element that is evolutionary conserved at sequence level between chicken, Zebra finch, American alligator, Chinese softshell turtle, lizard, human and mouse. We identified one candidate CRE of 1095 kb, ~1 kb upstream of MEOX1 (Fig. 6a). This element displayed enhancer activity, with expression of the Citrine reporter restricted to the PSM and all somites (Fig. 6b, c). Expression in PSM was unexpected as chromatin was not accessible at that stage. It is possible that the citrine reporter is missing some repressive elements that are present endogenously. Time-lapse movies reveal Citrine fluorescence, which was first detected in the prospective paraxial mesoderm cells of a HH6 embryo. At HH7 signal was detected in the first somite and subsequently in all newly formed somites, as well as the PSM and prospective paraxial mesoderm cells. Overall the pattern was consistent with MEOX1 gene expression detected in situ (Fig. 6d, f and Supplementary Movie 2). We identified two TF footprints within the enhancer, one for FOXO1 and one for ZIC3 (Fig. 6a). We next determined their requirement for the activation of fluorescent reporter expression. Mutation of FOXO1 or ZIC3 sites individually had no effect and reporter activity was still observed (Supplementary Fig. 6a, b). However, mutation of both sites led to loss of reporter activity (Fig. 6e). This suggests both TFs are able to activate this CRE and either FOXO1 or ZIC3 alone is sufficient. To investigate the significance of this element, we modified the endogenous enhancer using four gRNAs to target the dCas9-KRAB repressor to the MEOX1 Enh. Scrambled
gRNAs with dCas9-KRAB were used as control (Fig. 6f–h). Using a probe to detect MEOX1 transcripts showed that MEOX1 Enhancer perturbation led to loss of gene expression (Fig. 6g) \((n = 9/11)\). This was confirmed by RT-qPCR (Fig. 6j) and suggests the element is required.

To determine the genes potentially regulated by MEOX1 in paraxial mesoderm, we identified accessible chromatin regions within 10 kb of an expressed gene, which comprised a MEOX1 footprint (Supplementary Fig. 6c–e). STRING analysis of these putative MEOX1 regulated genes revealed PPI networks including...
genes enriched with the GO term anatomical structural development, and also included components of signalling pathways, such as Wnt and TGFBeta (Fig. 6). To confirm that some of these genes are involved in mediating the function of MEOX1 in paraxial mesoderm, we used RT-qPCR to assess their expression in normal and epigenome edited somites (Fig. 6f). First, we showed that MEOX1 expression in wild-type somites was unaffected after introducing control gRNAs together with the dCas9-KRAB repressor. However, electroporation of sgRNAs targeting the dCas9-KRAB repressor to the MEOX1 enhancer led to suppression of MEOX1. Expression of the closely related MEOX2 gene was not affected, similarly PAX3 expression remained unchanged. Next, we assessed a number of genes involved in chondrogenesis and setting up the polarity of the sclerotome. Epigenome editing of the endogenous MEOX1 enhancer led to down-regulation of Uncx4.1, TBX18, FAT4 and TGFB2, which are all associated with a MEOX1 footprint, indicating that they could be direct targets. Expression of NKK3-2, which is not associated with a MEOX1 footprint within 10 kb upstream or downstream of the gene, was also inhibited after negative regulation of MEOX1 by epigenome editing of the MEOX1 enhancer.

As the MEOX1 Enh is highly conserved amongst amniote taxa—birds, reptiles and mammals (Fig. 7a), we next asked whether the homologous mammalian sequences are active in chick. We found that a human MEOX1 Enh, isolated from HeLa cells, was able to drive Citrine expression in somites. Activity was also detected in LPM and PSM (Fig. 7b). The human MEOX1 Enh sequence included conserved FOXO1 and ZIC3 binding sites and mutation of both sites leads to loss of reporter activity (Fig. 7c). Therefore, we propose transcriptional regulation of the MEOX1 enhancer is highly conserved in human and chick, with FOXO1 and ZIC3 binding sites required for enhancer activity. Interestingly, the MEOX1 Enh sequence was not found in fish or amphibians (Fig. 7a). However, when we injected the chick MEOX1 Enh reporter into one cell of Xenopus laevis embryos at the 2-cell stage, we observed Citrine fluorescence in the paraxial mesoderm of early neurula stages (NF stage 14), where it overlapped with MYOD (Fig. 7d and Supplementary Fig. 7b). At NF stage 25, Citrine expression was detected in mesoderm and by NF stage 33 and stage 42 Citrine was visible in elongated muscle fibres, which are somite derived (Fig. 7d). The MEOX1 Enh with mutations in the FOXO1 and ZIC3 binding sites showed no activity (Supplementary Fig. 7a). This suggests that in amphibians the MEOX1 Enh can be activated by the same regulatory mechanism, even though the CRE is not conserved in the same location in the Xenopus laevis genome.

**Discussion**

Extension of the vertebrate body axis is driven by the addition of new segments at the posterior end of the embryo. During avian gastrulation, presumptive paraxial mesoderm cells ingress through the primitive streak and their early migration patterns can be observed directly. It has been shown that HOX genes are activated in a temporal colinear fashion, just prior to ingestion when the precursors are still located in the epiblast [1,2], thus regional identity along the anterior–posterior axis is acquired at gastrula stages. Paraxial mesoderm formation continues as the main body axis forms [3] and cells are added from a bi-potential population of NMP cells found in the tailbud. In response to high levels of Wnt3a and CDX family members, these progenitors commit to mesoderm fates and give rise to neck, trunk and tail structures [4].

Here we provide detailed molecular profiles of paraxial mesoderm of cranial and trunk regions by RNA-seq and ATAC-seq [5]. Using samples from along the axis we identify differential gene expression signatures consistent with axial patterning and differentiation, including the appearance of chondrogenic and myogenic markers in more anterior differentiating somites (Fig. 1c, d, j) [5,8,19,28,51]. We show that genes correlated with cell fate specification and muscle development are enriched in differentiating somites, and that the TF binding motifs found upstream of differentially expressed genes include the motif for myogenin consistent with its role in myogenic differentiation (Fig. 2h, i). Components of signalling pathways involved in anterior–posterior axis patterning are also differentially expressed, such as FGF, Wnt and RA pathways (Fig. 1g, k) [33,36] and the motif for RXRA is enriched in differentiating somites (Fig. 2i). Furthermore, we uncover genome-wide dynamic changes in chromatin accessibility across the spatiotemporal series. Using HINT-ATAC, an improved method to predict TF binding sites with footprints [42], we show differential coverage of several binding sites along the anterior–posterior axis, including sites for RARA and LEF1, transcriptional effectors of the RA and Wnt pathways. Furthermore, we identify differential footprints for CDX2, a readout for WNT and a TF required for axial elongation (Fig. 3a–l and Supplementary Fig. 3f–l). The observed coverage patterns correlate well with gene expression and with the known functions of RA and Wnt signalling in anterior–posterior axis patterning. Interestingly, despite their similar function in posterior axis extension our network analysis shows that CDX2 and LEF1 footprints are largely correlated with different genes. Only four genes are associated with both CDX2 and LEF1 footprints, suggesting that posteriorization driven by Wnt signalling involves at least two parallel acting protein–protein networks. Three of the
When paraxial mesoderm cells transition to generate somites. This emphasises the complexity of the segmentation process, axis elongation. PSM samples correspond to thoracic axial structures. As mentioned above, CDX2 is highly expressed in consistent with the role of HOX family members. These patterns are also detected in comparison to the differences seen between somites at different stages segmented mesoderm and all somite samples, for both chromatin previously not recognised. However, its possible role in axial elongation, indicated here, was previously not recognised.

Notably, we observe the greatest differences between pre-segmented mesoderm and all somite samples, for both chromatin accessibility and differential gene expression signatures, compared to the differences seen between somites at different stages of maturation (Fig. 2d–f and Supplementary Fig. 1b and 2d–f). This emphasises the complexity of the segmentation process, when paraxial mesoderm cells transition to generate somites.

We also observe differential chromatin accessibility along the anterior–posterior axis in the HOXA cluster (Fig. 4b) as well as differential expression (Fig. 4a) and footprints (Fig. 4c–f) of HOXA family members. These patterns are also detected in HOXB, HOXC and HOXD clusters (Supplementary Fig. 4) consistent with the role of HOX clusters in the regionalisation of axial structures. As mentioned above, CDX2 is highly expressed in PSM and has greater coverage of footprints in PSM compared to somite samples (Fig. 3a–d) consistent with its role in posterior axis elongation. PSM samples correspond to thoracic axial levels and this region is defined by central HOX genes, which are regulated by CDX proteins. CDX2 footprints were found in intergenic accessible chromatin peaks within the HOXA cluster (Fig. 4b).

Predicting enhancer gene interactions remains challenging, although new computational methods are becoming available. For example, the recent activity-by-contact model indicates that very long-range interactions are rare. Thus, in our footprint analysis we selected accessible chromatin regions within 10 kb of the transcription start site, or within 10 kb downstream of the gene. This approach combined with experimental validation and time-lapse imaging in gastrula-stage chick embryos identified CREs for TCF15 and MEOX1, both of which are in close proximity of the transcription start site (Figs. 5 and 6). For TCF15 we identified two separate elements, TCF15 Enh-1 and TCF15 Enh-2, which are both active in presegmented mesoderm and somites. TCF15 Enh-1 shows ectopic activity in the notochord. This expression was not seen when both elements were combined, suggesting that TCF Enh-2 may contain a repressor of notochord expression. However, both TCF Enh-2 and the combined CREs show ectopic activity in LPM, indicating additional repressive elements are missing. Alternatively, this CRE may interact with other gene(s) and direct their expression in the LPM. Footprint analysis identifies a RARA binding site as a highly relevant candidate TF binding site. Citrine activity is lost after mutation of the RARA site. Furthermore, dCas9-Krabo epigenome modification leads to loss of TCF15 expression. Although this is restricted to the region of the embryo that is targeted by electroporation at gastrula stages, this finding suggests the endogenous CRE is essential. This element is not conserved at sequence level and might be chick specific. In contrast, the MEOX1 enhancer identified here is conserved in avians, reptiles and mammals but not in amphibian or fish (Fig. 7a). In both chick and human CREs, TF binding sites for FOXO1 and ZIC3 are required for Citrine expression (Figs. 6c and 7c). In vivo epigenome modification of the MEOX1 enhancer causes axial elongation phenotypes in embryos (Fig. 6d, h). RT-qPCR shows that genes involved in
chondrogenesis and sclerotome polarity are affected in somites where expression of MEKO1 is lost after targeting the dCas9-Krab repressor to the enhancer (Fig. 6). Furthermore, the elongation phenotypes observed are consistent with mouse mutants\textsuperscript{21,23} and human Klippel-Feil patients who display skeletal abnormalities\textsuperscript{25,26}. Thus, it could be of interest to determine whether the MEKO1 enhancer is affected in patients, who do not have a coding mutation in MEKO1.

- Taken together we provide a resource of paraxial mesoderm samples across a spatiotemporal series. Our analysis focussed on PPI networks and CREs important for vertebrate anterior–posterior axis formation. We assess evolutionary conservation and validate in vivo function to establish proof-of-principle, which underpins further interrogation and mining of this comprehensive data set.

**Methods**

**Chicken embryos.** Fertilised chicken eggs (Henry Stewart & Co.) were incubated at 37 °C with humidity. Embryos were staged according to Hamburger and Hamilton\textsuperscript{27}. All experiments were performed on chicken embryos younger than 14 days of development and therefore were not regulated by the Animal Scientific Proce- dures Act 1986.

**Embryo dissection.** HH14 embryos were dissected into Ringers solution in silicon lined petri dishes and pinned down using the extra-embryonic membranes. Ringers solution contained with Disperse (1.5 mg/ml in DMEM). Embryos in DMEM 10 mM HEPES + PHE 7.5 at 37 °C for 5 min prior to treatment with Trypsin (0.05%) at 37 °C for 7 min. The reaction was stopped with Ringers solution with 0.25% BSA. The PSM, ES, and MS were carefully dissected away from neural and lateral mesoderm tissue using sharp tungsten needles.

**RNA extraction, library preparation and sequencing.** RNA was extracted using Qiagen RNAeasy kit (Cat no. 74104) and DNase treated with Invitrogen Purelink gDNA extraction kit (Cat no. K1820-00). Enhancer cloning

Three biological replicates for each stage were used for analysis. Paired-end 150 bp reads on the Illumina HiSeq4000 platform at Novogene UK.

**Libraries were quantified using a modi-** 

**fied GoldenGate protocol\textsuperscript{61} under the following conditions:** 10 cycles of 94 °C, 55 s; 55 °C, 3 min; 65 °C, 25 cycles of 94 °C, 5 s; 65 °C, 1 min; 72 °C, 1 min. Library preparation was complete after further clean up using Qiagen M0543S for 11 cycles as follows: 72 °C, 5 min; 98 °C, 3 min; 10 cycles of 94 °C, 55 s; 55 °C, 3 min; 65 °C, 3 min. Libraries were pooled with pTK nanotag reporter vector with T4 DNA ligase (Promega) and BsmBI (NEB) restriction enzyme. This reaction was prepared for T4-mediated ligation and BsmBI digestion under the following conditions: 25 cycles of 37 °C, 2 min; 16 °C, 5 s; a single step of 55 °C, 5 min; and a final step of 80 °C, 5 min. For generation of specific sites in enhancers we utilised FastCloning methodology\textsuperscript{82}.

**Crispr-mediated enhancer repression.** sgRNAs specific for MEKO1 Enh and TCF15 Enh-2, or a scrambled control were cloned into a chicken pU6-3 vector using standard protocols\textsuperscript{83}. For enhancer repression, sgRNAs and dCas9-Krab were electroporated ex ovo\textsuperscript{22}. All primer sequences are detailed in Supplementary Table 1.

**RT-qPCR.** cDNA was synthesised from 300 ng of RNA using a Maxima First Strand cDNA synthesis kit (Thermo Fisher Scientific). qPCR was performed on a 7500 Fast Real Time PCR machine (Applied Biosystems) using SYBR Green PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer’s instructions. Primers (see Supplementary Table 1) were designed with Primer3Plus software (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi). RT-qPCR was normalised to GAPDH mRNA. Three independent experiments each with replicate samples were performed for each RT-qPCR. The delta-delta CT method was used to analyse gene expression levels. Statistical analysis was performed using Graph-Pad Prism (Version 6) software. Mann–Whitney non-parametric two-tail testing was applied to determine p values.

**Embryo preparation and ex ovo electroporation.** Hamburger and Hamilton (HH3+) embryos were captured using the filter paper base easy-culture method. Briefly, eggs were incubated for ~20 h, a window was created using forceps, the embryo and yolk were transferred into a dish and thin albumin above and around the embryo was removed using tissue paper. A circular filter paper ring was placed on top, excised and transferred into a separate dish containing Ringers solution and excess yolk was removed. The embryo was then transferred into a dish containing albumin-sagar and ready for electroporation. The embryonic DNA was injected between the membrane and embryo to cover the whole epiblast. Electroporated using five pulses of 5 V, 50 ms on, 100 ms off. Thin albumin was used to seal the lids of dishes and embryos were cultured at 37 °C with humidity to the desired stage.

**Cryosectioning and immunostaining.** Embryos were fixed in 4% parformaldehyde (PFA) for 2 h at room temperature (RT) or at 4 °C overnight, washed 3 x 10 min in PBS. Embryos were transferred into 30% sucrose/PBS overnight at 4 °C prior to 3 x 10 min washes in OCT before final embedding of OCT in dry ice. Cryosectioning was performed at 15 μm thickness. Sections were washed in 3 x 15 min PBS and 1 x 15 min PBS/0.5% Triton X-100 prior to blocking in 5% goat serum and 5% BSA in PBS for 1 h at RT. Incubation with primary antibody for rabbit anti-rabbit IgG (1:200, Torrey Pines Biologs Cat no. TP401) at 4 °C overnight, followed by 3 x 10 min washes in PBS and incubation with secondary antibody AlexaFluor-568-conjugated donkey anti-rabbit IgG (1:500, Thermo Fisher Cat no. A21206) for 1 h at RT. Sections were washed 3 x 10 min in PBS and 1 x wash with PBS and DAPI (Sigma-Aldrich) at 0.1 mg/ml in PBS.

**Wholemount in situ hybridisation.** Wholemount in situ hybridisation using DIG-UTP labelled antisense RNA probes for MEKO1 (a gift from Baltjinder Mankoo, King’s College London UK) and TCF15 (a gift from Susanne Dietrich, University of Portsmouth UK) was carried out using standard methods. Briefly, following fixation in 4% PFA embryos were treated with Protease K, hybridised over night at 60 °C, after post-hybridisation washes and blocking with BMB (Roche), embryos were treated with anti-DIG antibody, coupled to alkaline phosphatase (Roche). Signal was developed using NBT/BCIP.

**Live imaging of enhancer reporter.** Embryos cultured in six-cell stage cell culture plates (Falcon) were time-lapse imaged on an inverted wide-field microscope (Axiovert; Zeiss). Brightfield and fluorescent images were captured every 6 min for 20–24 h, using Axiovision software as described in ref. \textsuperscript{84}. At the end of the incubation, most embryos had reached stage HH10-11.

**Image analysis.** Sections were visualised on an Axioscope with Axiovision software (Zeiss). Wholemount embryos were photographed on a Zeiss SV11 dissecting microscope with a Micropublisher 3.5 camera and acquisition software or Leica MZ16F using Leica Firecam software. Live imaging datasets were analysed in Fiji/ImageJ.

**ATAC-seq processing.** Adaptors were removed from raw paired-end sequencing reads and trimmed for quality using Trim Galore! (v.0.5.0)\textsuperscript{85} a wrapper tool around Cutadapt\textsuperscript{66} and FastQC\textsuperscript{27}. Default parameters were used. Quality control (QC) was performed before and after read trimming using FastQC (v.0.11.6)\textsuperscript{87} and no issues were highlighted from the QC process. Subsequent read alignment and post-alignment filtering was performed in concordance with the ENCODE project’s “ATAC-seq Data Standards and Prototype Processing Pipeline” for replicated data (https://www.encodeproject.org/atac-seq/). In brief, reads were mapped to the chicken genome gaGAS assembly using bowtie2 (v.2.3.4.2)\textsuperscript{88}. The resultant Sequence Alignment Map (SAM) files were compressed to the Binary Alignment Map (BAM) version on which SAMtools (v.1.9)\textsuperscript{89} was used to filter reads that were unmapped, mate unmapped, not primary alignment or failing platform quality checks. Reads mapped as proper pairs were retained. Multi-mapping reads were
removed using the Python script assign_multimappers provided by ENCODE's processing pipeline and duplicate reads within the BAM files were tagged using Picard MarkDuplicates (v.2.18.12) [http://broadinstitute.github.io/picard/] and then filtered using SAMTools. For each step, parameters detailed in the ENCODE pipeline were used. From the processed BAM files, coverage tracks in bigWig format were generated using deepTools bamCoverage (v.3.1.2)39 and peaks were called using MACS2 (v.2.1.1)37 (parameters -f BAMPE -g mm -B -nolod SELF -shift -100 -etxsize 200). Coverage tracks and peaks (narrow peak format) were uploaded to the UCSC Genome Browser79 as custom tracks for ATAC-seq data visualisation.

**Differential accessibility and footprinting.** Analysis of ATAC-seq for differential accessibility was carried out in R (v.3.5.1)39 using the DiffBind package (v.2.8.0)42,43 with default parameter settings. Differential accessibility across samples was calculated from the negative binomial distribution model implemented in DESeq2 (v.1.4.5)44. Computational footprinting analysis was conducted across samples using HINT-ATAC which is part of the Regulatory Genomic Toolbox (v.0.12.3)42 also using default parameter settings and the galGal5 genome.

**RNA-seq differential expression analysis.** Adapters were removed from raw paired-end sequencing reads and trimmed for quality using Trim Galore! (v.0.5.0) using default parameters. QC was performed before and after read trimming using FastQC (v.0.11.6) and no data quality issues were identified after checking the resultant QC reports. Processed reads were mapped to galGal5 cDNA using kallisto (v.0.44.0)75. Resultant quantification files were collated to generate an expression matrix. Differential expression, GO term and pathway analyses were then conducted using DESeq274 and default settings within the iDEP (v.9.0)76 web interface.

GO term analysis used PGSEA method for GO Biological Process with a minimum GO term enrichment score of 0.01.

**Data availability**

The authors declare that all data supporting the article and its supplementary information findings of this study are available within the Nature Research DataMaturity database. Further information on research design is available in the Nature Communications article and its supplementary information.

**References**

1. Yang, X., Dormann, D., Munsterberg, A. E. & Weijer, C. J. Cell movement patterns during gastrulation in the chick are controlled by positive and negative chemotaxis mediated by FGF4 and FGFR. Dev. Cell. 3, 425–437 (2003).

2. Imura, T., Yang, X., Weijer, C. J. & Pourquie, O. Dual mode of paraxial mesoderm formation during chick gastrulation. Proc. Natl. Acad. Sci. USA 104, 2744–2749 (2007).

3. Benazeraf, B. & Pourquie, O. Formation and segmentation of the vertebral body axis. Annu. Rev. Cell Dev. Biol. 29, 1–26 (2013).

4. Brent, A. E. & Yamada, C. J. Developmental regulation of somite derivatives: muscle, cartilage and tendons. Curr. Opin. Genet. Dev. 12, 548–557 (2002).

5. Christ, B., Huang, R. & Scaal, M. Amniote somite derivatives. Dev. Dyn. 236, 2382–2396 (2002).

6. Gros, J., Serralbo, O. & Marcellé, C. WNT11 acts as a directional cue to organize the elongation of early muscle fibres. Nature 457, 589–593 (2009).

7. McColl, J. et al. 4D imaging reveals stage dependent random and directed cell motion during somite morphogenesis. Sci. Rep. 8, 12644 (2018).

8. Gros, J., Scaal, M. & Marcellé, C. A two-step mechanism for myotome formation in chick. Dev. Cell 6, 875–882 (2004).

9. Knutla, M. & Dribble, D. Ontogeny of T-box genes in time and space; 25 years of colinear tinkering. Science 301, 333–333 (2003).

10. Noordermeer, D. et al. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. Elife 3, e02557 (2014).

11. Neijts, R. & Deschamps, J. At the base of colinear Hox gene expression: cis-regulatory and trans-factor interactions orchestrating the initial phase of Hox cluster activation. Dev. Biol. 428, 293–299 (2017).

12. Imura, T. & Pourquie, O. Colinear activation of Hox genes during gastrulation is linked to mesoderm cell ingress. Nature 442, 568–571 (2006).

13. Imura, T., Denans, N. & Pourquie, O. Establishment of Hox vertebral identities with the embryonic spine precursors. Curr. Top. Dev. Biol. 88, 201–234 (2009).

14. Aires, R., Dias, A. & Mallo, M. Deconstructing the molecular mechanisms shaping the vertebrate body plan. Curr. Opin. Genet. Dev. 55, 81–86 (2018).

15. Tabaries, S. et al. Cdx protein interaction with Hox5 regulatory sequences contributes to Hox3 regional expression along the axial skeleton. Mol. Cell. Biol. 25, 1389–1401 (2005).

16. Chawengnakophak, K., de Graaff, W., Rossant, J., Deschamps, J. & Beck, F. Cdx2 is essential for axial elongation in mouse development. Proc. Natl. Acad. Sci. USA 101, 7641–7645 (2004).

17. Neijts, R., Amin, S., van Rooijen, C. & Deschamps, J. Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk phenotype. Dev. Biol. 321, 145–154 (2017).

18. Abou-Elhamd, A. et al. Khal1 attenuates beta-catenin dependent Wnt signalling and regulates embryo myogenesis. Dev. Biol. 402, 61–71 (2015).

19. Mok, G. F., Mohammed, R. H. & Sweetman, D. Expression of myogenic regulatory factors in chicken embryos during somite and limb development. J. Anat. 227, 352–360 (2013).

20. Bertini, F. et al. Time course and side-by-side analysis of mesodermal, pre-myogenic, myogenic and differentiated cell markers in the chicken model for skeletal muscle formation. J. Anat. 227, 361–382 (2015).

21. Burgess, R., Rawls, A., Brown, D., Bradley, A. & Olson, E. N. Requirement of the paraxis gene for somite formation and musculoskeletal patterning. Nature 384, 570–573 (1996).

22. Young, T. et al. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. Dev. Cell 17, 516–526 (2009).

23. Mankoo, B. S. et al. The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the formation, patterning and differentiation of somites. Development 130, 4655–4664 (2003).

24. Skuntz, S. et al. Lack of the mesodermal homeodomain protein MEOX1 disrupts sclerotome polarity and leads to a remodeling of the cranio-cervical joints of the axial skeleton. Dev. Biol. 332, 383–395 (2009).

25. Bayrakti, F. et al. Mutation in MEOX1 gene causes a recessive Klippel-Feil syndrome subtype. BMC Genet. 14, 95 (2013).

26. Mohamed, Y. J. et al. Mutations in MEOX1, encoding mesenchyme homeobox 1, cause Klippel-Feil anomaly. Am. J. Hum. Genet. 92, 157–161 (2013).

27. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. J. Morphol. 88, 49–92 (1951).

28. Kalcheim, C. & Ben-Yair, R. Cell rearrangements during development of the somite and its derivatives. Curr. Opin. Genet. Dev. 15, 371–380 (2005).

29. Langfèdèr, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma. 9, 559 (2008).

30. Langfèdèr, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatica 24, 719–720 (2008).

31. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).

32. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393 (2012).

33. DIFFbind v.3.0.13 (Biocconduct, 2011).

34. Lee, C. C. et al. TCF12 promotes functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. J. Biol. Chem. 287, 2798–2809 (2012).

35. Biesiada, E., Hamamori, Y., Kedes, L. & Sartorelli, V. Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac alpha-actin promoter. Mol. Cell. Biol. 15, 2577–2584 (1995).

36. Dutny, W. C. Jr., Kennedy, M. W., Chalamalasetty, R. B., Campbell, K. & Yamaguchi, T. P. Transcriptional profiling of Wnt3a mutants identifies Sp transcription factors as essential effectors of the Wnt/beta-catenin pathway in neuronesomesural stem cells. PloS ONE 9, e87018 (2014).

37. Shalamoto, T. et al. CIZ, a zinc finger protein that interacts with p30(cas) and activates the expression of matrix metalloproteinases. Mol. Cell. Biol. 20, 1649–1658 (2000).

Received: 5 February 2020; Accepted: 25 January 2021; Published online: 19 February 2021
38. Yang, Y. et al. Pokemon (FBI-1) interacts with Smad4 to repress TGF-beta-induced transcriptional responses. *Biochim. Biophys. Acta* **1849**, 270–281 (2015).
39. Ambele, M. A. A. & Tague, M. S. Identification of transcription factors potentially involved in human adipsogenin in vitro. *Mol. Genet. Genom. Med.* **5**, 210–222 (2017).
40. Delattre, O. et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*** **359**, 162–165 (1992).
41. Ghyselinck, N. B. & Duester, G. Retinoic acid signaling pathways. *Dev. Cell*** **145** (2014).
42. Yang, Y. et al. Pokemon (FBI-1) interacts with Smad4 to repress TGF-beta-mediated transcriptional responses. *Nucleic Acids Res.* **44**, W160–W165 (2016).
43. Feingold, J., Liu, T., Qin, R.; Zhang, Y. V. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
44. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
45. Core Team, R. A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2018).
46. Love, M. L., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 523–527 (2016).
48. Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018).
49. Sklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D687–D693 (2019).

**Acknowledgements**

We thank all members of the Münsterberg and Wheeler labs for helpful discussion. Dr Timothy Grocott for discussions, Ronco Saputil and undergraduate project students for assistance with enhancer analysis. G.F.M., L.F., E.M. and V.M.W. were supported by BBSRC project grant (BB/N003034/1) to A.E.M. and G.N.W. and MRC project grant (MR/R00549/1) to A.E.M.; S.A.W. and A.M.G. were supported by studentships funded by the UKRI Biotechnology and Biological Sciences Research Council Norwich Research Park Biosciences Doctoral Training Partnership to A.E.M. and G.N.W.

**Author contributions**

G.F.M. and A.E.M. conceived and designed the study. G.F.M. generated and analysed RNA-seq and ATAC-seq data, performed and analysed chick reporter expression assays, in situ hybridisation, immunohistochemistry, QPCR experiments and assisted in bioinformatic analysis. L.F. performed bioinformatic analysis on RNA-seq and ATAC-seq data. E.M. and S.A.W. assisted in live imaging, chick embryo injections and in situ hybridisation. V.M.W. assisted with cloning reporter constructs. R.M.W. and T.S.S. shared electroporation setup, plasmids and expertise in NGS. A.M.G. performed in situ hybridisations into *Xenopus laevis* and in situ hybridisation. G.N.W. discussed and supervised *Xenopus laevis* experiments. S.M. helped oversee the computational analysis. G.F.M. and A.E.M. discussed ideas and interpretation of data and wrote the manuscript with input from all authors. A.E.M. supervised the study.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21426-7.

**Correspondence**

requests for materials should be addressed to A.E.M.

**Peer review information Nature Communications** thanks Hector Escrivá and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2021