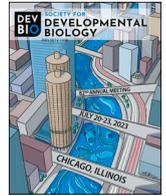


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Developmental Biology

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

Original research article

Transcriptomics and chromatin accessibility signatures define the cervical-thoracic boundary along the vertebrate axis

Shannon A. Weldon^a, Emily L. Smith^a, Magdalena Schatka^a, Leighton Folkes^a, Gi Fay Mok^a, Ashleigh Lister^b, Iain C. Macaulay^b, Wilfried Hearty^b, Andrea E. Münsterberg^{a,*} 

^a School of Biological Sciences, University of East Anglia, Norwich Research Park, NR4 7TJ, UK^b The Earlham Institute, Norwich Research Park, Norwich, UK

ARTICLE INFO

Keywords:

ATAC-Sequencing
RNA-Sequencing
Axis development
HOX genes
lncRNA

ABSTRACT

In vertebrate embryos, somite pairs form on either side of the neural tube along the main body axis. Somites generate the tissues of the musculoskeletal system, including cartilage of the vertebral column and ribs and skeletal muscles of the trunk and limbs. The detailed anatomy of somite-derived tissues varies along the axis, with unique features most easily visible in the vertebral column. Here we investigate the genetic control of this regionalization, which drives the subsequent cell differentiation programmes, focusing on the cervical to thoracic (C-T) boundary. Using ATAC-sequencing and RNA-sequencing, we establish molecular profiles of somites, in particular the chromatin landscapes and transcriptional programmes, that define this anatomical transition. Differential analysis highlights candidate *cis*-regulatory elements (CRE), and *in silico* footprints identify coverage of transcription factor (TF) binding sites associated with differentially expressed genes. Electroporation of citrine reporters *in vivo* validates the activity of CREs associated with key HOX genes, HOXC6 and HOXC8. HOXC6 footprints indicate its role in regulating a trio of differentially expressed SOX transcription factors, SOX5, SOX6 and SOX9, which are involved in chondrogenesis. In addition, the differential analysis identifies several lncRNAs, including one that is located within the HOXC cluster. CRISPR-on experiments suggest HOXC6 regulates its expression and therefore we name it lncRNA-HOXC6TA, however, its function in the thoracic region is currently unknown. Our study provides valuable datasets and illustrates how they can be mined to gain further insights into the regulatory mechanisms underlying the C-T transition along the vertebrate body axis.

1. Introduction

During vertebrate embryogenesis, somites emerge as transient mesodermal structures on both sides of the neural tube. The process of somitogenesis is conserved across vertebrates and generates the segmented body plan. It is governed by waves of gene expression, which oscillate across the presegmented mesoderm tissue at the posterior of the embryo, producing a new somite pair. Genes with oscillating patterns of expression include components of the Wnt, Notch and fibroblast growth factor (FGF) signalling pathways. In addition, there are opposing signalling gradients along the axis, comprising Wnt, FGF and Retinoic acid (RA) pathways, which determine the formation of the next segment (Benazeraf and Pourquie, 2013; Dequeant and Pourquie, 2008).

Somites progressively differentiate along the axis. This generates a maturation gradient as shown by the expression of lineage markers. As

embryogenesis proceeds somite differentiation accelerates and marker gene expression is detected at earlier somite stages (Berti et al., 2015; Borman and Yorke, 1994; Maschner et al., 2016; Mok et al., 2015). Molecular analysis showed that this earlier gene activation correlates with a change in the permissiveness of the chromatin landscape (Ibarra-Soria et al., 2023).

As they mature, somites go through a stereotypical process of morphogenesis forming different compartments, the mesenchymal sclerotome ventrally, the epithelial dermomyotome dorsally and the myotome in between. In response to signals from adjacent tissues, cells differentiate to generate the musculoskeletal system, including skeletal muscle, cartilage, tendons, and vertebrae (Brent and Tabin and 2002; Christ et al., 2007; Christ and Scaal, 2008). To better understand the gene regulatory processes underlying these cell differentiation programs, we previously characterised the dynamic changes of the

This article is part of a special issue entitled: Avian model systems published in *Developmental Biology*.

* Corresponding author.

E-mail address: a.munsterberg@uea.ac.uk (A.E. Münsterberg).

<https://doi.org/10.1016/j.ydbio.2025.06.012>

Received 29 August 2024; Received in revised form 3 April 2025; Accepted 12 June 2025

Available online 13 June 2025

0012-1606/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

transcriptomes and chromatin accessibility along the main body axis in chick, starting with presegmented mesoderm tissue (PSM) to early, maturing and differentiating somites (Mok et al., 2021). Furthermore, we used single cell RNA-sequencing of somites and surrounding tissues to identify the discrete cell lineages within the embryonic trunk. We combined this with RNA-tomography, which generated spatially resolved expression data, to discover candidate genes with potential roles in cell differentiation and the emergence of the avian body axis (Mok et al., 2024).

Depending on their location along the embryonic axis, somites have unique regional identities. These produce distinct axial domains, including occipital, cervical, thoracic, lumbar, sacral, and caudal, which are characterised by anatomical features such as limbs and vertebral elements (Weldon and Munsterberg, 2022). How cellular differentiation programs are adapted to generate different morphologies is not fully understood. It is known that the influence of collinear *HOX* gene expression on determination of axial identity is critical (Deschamps and Duboule, 2017). *HOX* expression is initiated prior to somite formation and these transcription factors control the timing of ingression of PSM precursors through the primitive streak (Iimura and Pourquie, 2006). In both, PSM and in the adjacent neural tube, their expression is initiated by temporal and directional activation within each cluster. This begins during gastrulation and is also detected in progenitor cell populations located caudally in the embryo (Wymeersch et al., 2021). This pattern ultimately instructs anatomically distinguishable vertebrae subtypes of the vertebral column (Chal and Pourquie, 2017; Scaal, 2016, 2021). Similarly, the competence of somites to generate either migratory or nonmigratory hypaxial muscle precursors depends on axial identity conferred by *Hox* genes (Alvares et al., 2003). The progressive opening of *HOX* clusters is associated with changes in chromatin structure (Ibarra-Soria et al., 2023; Soshnikova and Duboule, 2009). In chick this has been visualised using chromatin accessibility within PSM and somites (Mok et al., 2021, 2024). Differentially accessible chromatin detected across all four *HOX* clusters reflects the organisation of genes within clusters and their collinear expression. Footprints for transcription factors involved in *HOX* gene regulation and patterning, such as *CDX1/2* and RA receptors, were found in intergenic regions (Mok et al., 2021).

Here we focus on axial positions that represent prospective cervical and thoracic levels, the cervical to thoracic boundary (C-T), which generates vertebrae with anatomically discrete features such as ribs. The C-T boundary is demarcated by *HOXC6* and *HOXC8* in many vertebrates, including mouse and chick (Burke and Tabin and 1996; Gaunt, 1994), alligator (Mansfield and Abzhanov, 2010) and snakes (Cohn and Tickle, 1999; Leal and Cohn, 2018; Woltering et al., 2009). Overexpression of *HOX6* paralogues in PSM produces ectopic ribs in cervical and lumbar regions, indicating that *HOX6* can determine thoracic identity throughout the vertebral column. Conversely, ectopic *HOXA10* suppresses rib formation and the absence of *HOX10* function leads to ectopic ribs in all posterior vertebrae (Wellik and Capecchi, 2003; Carapuco et al., 2005). Both *HOX6* and *HOX10* interact with an enhancer element that regulates expression of the myogenic regulators, *MYF5* and *MRF4/Myf6*, in the hypaxial myotome, which in turn activates *FGF* and *PDGF* signalling to promote rib formation (Vinagre et al., 2010). Polymorphisms in this enhancer element modulate responses to rib-suppressing and rib-promoting *HOX* proteins leading to an expanded rib cage in some species (Vinagre et al., 2010), including in snakes (Guerreiro et al., 2013).

Mutation analysis shows that combinatorial expression of *HOX* genes at a given anterior-posterior position is an important determinant of axial identity. However, the interplay between *HOX* genes and other contributing factors in cellular differentiation events, and their role in the emergence of specific anatomical traits remains incompletely understood. To address this we generated molecular profiles, transcriptomes and chromatin accessibility, of somites across the C-T boundary. Differential analysis of these datasets highlights ATAC-peaks

associated with either cervical or thoracic regions. We validate *cis*-regulatory elements for *HOXC6*, *HOXC8* and *HOXC5*, and identify footprints and candidate factors potentially involved in the transcriptional activation of these *HOX* genes. We locate *HOXC6* footprints associated with differentially expressed *SOX* transcription factor genes and discover several differentially expressed lncRNAs. One of these lncRNAs resides in the *HOXC* cluster and its expression is activated after *HOXC6* mis-expression. Our study contributes to a better understanding of the regulation of the C-T transition, provides a resource and proof-of-principle for future mining of the datasets generated.

2. Results and discussion

2.1. Differential profiles of somite transcriptomes across the cervical to thoracic boundary

To investigate the molecular basis underpinning the generation of anatomically distinct derivatives at cervical versus thoracic axial levels, we mapped individual somite transcriptomes across this boundary. We dissected the final six somites from chick embryos at stage HH14 (Hamburger and Hamilton, 1992). This stage embryo has 22 somites, including cervical level somites (6–19) and thoracic level somites (20–22) (Weldon and Munsterberg, 2022). New somites continually form and progressively differentiate and based on agreed nomenclature (Christ and Ordahl, 1995) the final six somites represent somite stages I - the most recently formed and least mature, to somite stage VI - residing more anteriorly and more differentiated. To account for the differences in somite maturation, we dissected the final three cervical-level somites (17–19) from a younger embryo (HH13), when they are at the epithelial stage of differentiation and thus comparable with the epithelial stages (I-III) of the final three thoracic somites 20–22 harvested at HH14 (Fig. 1A).

Differential gene expression analysis was conducted for three contrasts (Fig. 1B and C). Comparing cervical with thoracic somites from the same embryo at stage HH14 - representing somite stages I-VI, revealed 1105 genes differentially expressed, 466 genes were up-regulated and 639 down-regulated in thoracic somites (differential D1). These genes should include those associated with the C-T transition, but also genes involved in somite maturation. A second contrast compared cervical somites from HH13 with thoracic somites from HH14 embryos (differential D2). These somites are all at the epithelial stage of maturation, somite stages I-III, and represent C-T axial level differences. This analysis identified 6082 differentially expressed genes, 3299 genes were up-regulated and 2783 down-regulated in thoracic somites. The third comparison examined cervical somites 17–19, either at epithelial stage (HH13) or more mature stage (HH14) (differential D3). This revealed 4995 genes differentially expressed in cervical level somites, these are likely associated with maturation and 596 of these genes were overlapping with D1 (Fig. 1C). Many more genes were identified as differentially expressed in D2 and D3, compared to D1. The reason for this is not clear, however, samples were either derived from different embryos (D2, D3), or from the same embryo (D1).

To refine the list of genes of interest, we selected those identified in both D1 and D2 and removed those present in D3. This identified a set of 343 differentially expressed genes predicted to be responsible for cervical versus thoracic axial level specific patterning but not for somite maturation (Fig. 1C). This subset consisted of 153 up-regulated and 190 down-regulated genes, including genes associated with relevant GO-terms such as anterior posterior pattern specification (*HOXC8*, *HEY1*, *HOXB8*, *HER2*), bone and cartilage development (*CALCR*, *FGF18*, *BMP7*, *PRP3*, *FGF8*, *BMP2K*, *CREB3L2*, *CDH2*, *NFIB*, *SOX9*, *RARB*, *SULF2*, *SNAI1*, *TWIST1*, *ASF1A*, *CD276*, *ITGB3*, *FBN2*, *CD81*, *COL1A2*, *DHRS3*), or skeletal muscle development (*PDLIM1*, *ASF1A*, *KRT24*, *ANXA1*, *MEOX2*, *CD81*, *ITGB1*, *DMD*, *BMP7*, *SDC4*, *GPC1*, *EYA1*, *CDH2*, *MAFF*). Genes associated with regulation of signalling pathways included: Notch signalling (*ADAM17*, *GSX2*, *HEY1*, *BMP7*, *BMP2K*,

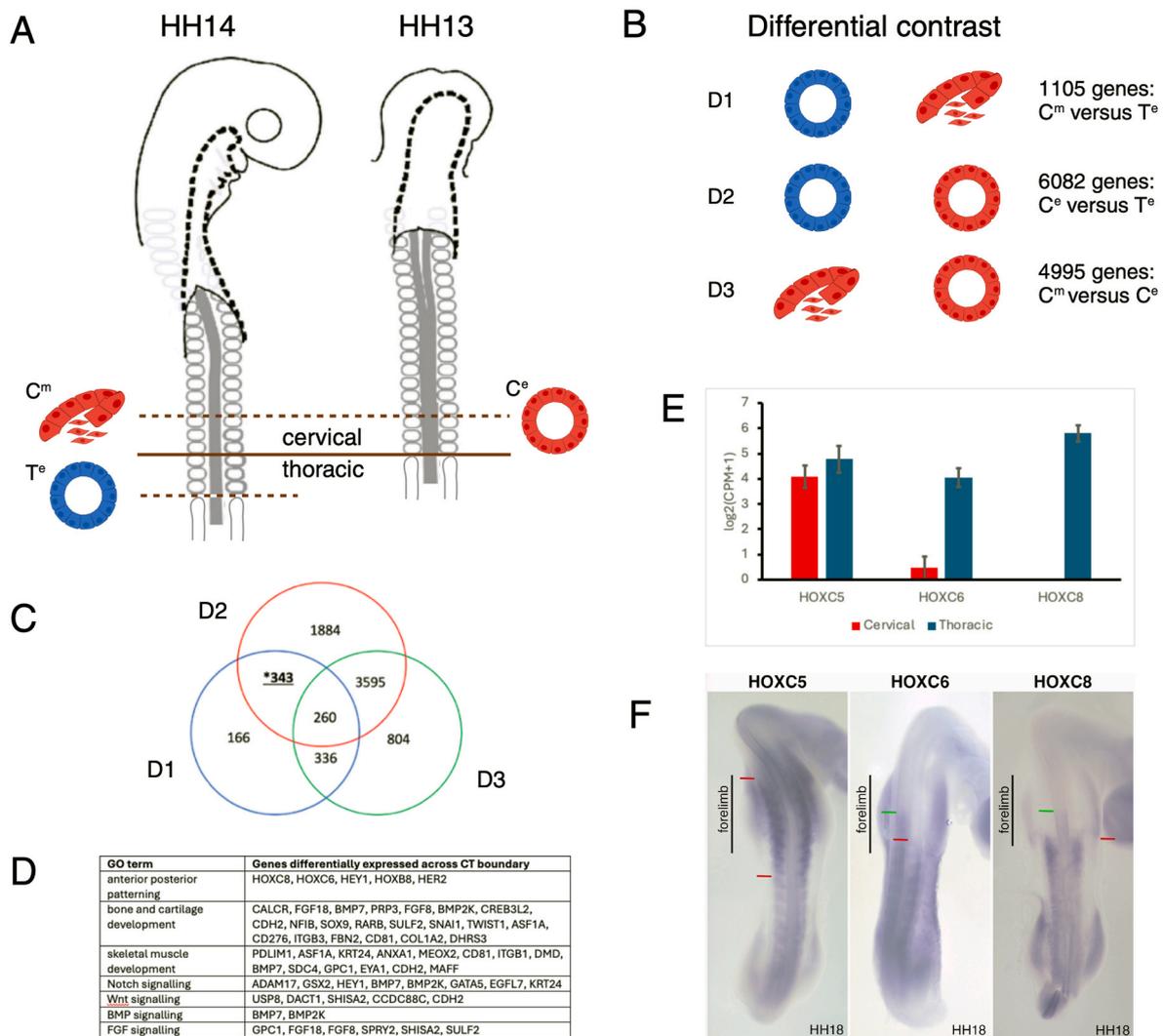


Fig. 1. Transcriptional profiling of somites across the C-T boundary. (A) Schematic representation of the experimental setup for library generation to produce three contrasts. Line indicates C-T boundary at somite level 19/20. An additional dotted line indicates somite 17, the anterior expression boundary of *HOXC5* and the first of six individual somites used for RNA-Seq and ATAC-Seq in this project. C^m = cervical maturing, C^e = cervical epithelial, T^e = thoracic epithelial (B) The contrasts examined are shown. Differential 1 (D1) compares somites across the C-T boundary at different stages of development - from epithelial to maturing and identifies genes associated with regionalization and differentiation. Differential 2 (D2) compares epithelial somites from different HH stage embryos spanning the C-T boundary to identify genes involved in position. Differential 3 (D3) compares cervical level somites at different stages of maturation. (C) Venn diagram of all differentially expressed genes, up or down, across the three comparisons. The starred intersection between D1 and D2 comprises 343 genes predicted to be mainly responsible for C versus T regional differences. (D) Selected GO terms relevant to skeletal system development, associated differentially expressed genes are shown. (E) RNA-sequencing shows log₂ fold change of expression of *HOXC* cluster genes in cervical (red) or thoracic (blue) level somites. (F) In situ hybridisation of the same genes: *HOXC5*, *HOXC6* and *HOXC8* in chick embryos stage HH18. Vertical black line marks the forelimb, red lines mark expression boundary in somites, green lines mark expression boundary in neural tube. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

GATA5, EGFL7, KRT24), Wnt signalling (USP8, DACT1, SHISA2, CCDC88C, CDH2), BMP signalling (BMP7, BMP2K) and FGF signalling (GPC1, FGF18, FGF8, SPRY2, SHISA2, SULF2) (Fig. 1D). The gene lists are included as supplementary information.

Among the genes upregulated in thoracic somites were mid-level *HOX* genes, with *HOXC8* and *HOXC6* emerging as the most differentially expressed (Fig. 1E). In contrast *HOXC5* was expressed at similar levels in cervical and thoracic somites. Wholemount in situ hybridisation in HH16 chick embryos (Fig. 1F) illustrates the association of *HOXC6* expression with the first thoracic somite, somite 20, which aligns with the posterior boundary of the forelimb. *HOXC8* is expressed from the next thoracic somite, somite 21. In contrast, *HOXC5* is expressed equally in somites across the C-T boundary. In the neural tube, the boundaries of *HOXC6* and *HOXC8* expression are shifted to more anterior regions compared to the boundaries in paraxial mesoderm. This is

consistent with the published expression boundaries for these genes in mouse and chick (Burke et al., 1995; Nishimoto et al., 2014).

Hierarchical clustering and heatmap visualization reinforced PCA findings, effectively segregating cervical and thoracic somites. Gene ontology (GO) for functional terms highlighted biological processes linked to anatomical structure morphogenesis and anatomical structure development for both up-regulated and down-regulated genes (Supplementary Fig. 1C and D). Genes and pathways associated with anterior-posterior pattern formation, bone morphogenesis and muscle differentiation were identified. This included classic markers for chondrogenesis and cartilage condensation, such as the transcription factors, *SOX9*, *SOX5*, *SOX6* and *PAX1*, as well as bone morphogenetic protein 7 (*BMP7*). The three *SOX* family members are referred to as the 'chondrogenic trio'. *SOX9* is a transcriptional activator required for chondrogenesis, and *SOX5* and *SOX6* are closely related DNA-binding

proteins that promote its function (Liu and Lefebvre, 2015). Together with SOX9, PAX1 regulates expression of genes involved in chondrogenesis, including for example *Bapx1* (Rodrigo et al., 2003; Yamashita et al., 2009). Signalling pathways expressed in a graded fashion along the axis included FGF and Wnt, which were more highly expressed in thoracic samples (Aulehla and Pourquie, 2010). In contrast, the Notch signalling pathway, which is involved in somite boundary formation (Dale et al., 2003), was more highly expressed in cervical samples.

Further hierarchical clustering analysis partitioned differential genes based on similar expression patterns spanning the C-T boundary (Supplementary Fig. 2). To reduce complexity, we removed the two somites overlapping the boundary (19, 20) and compared cervical somites 17 and 18 with thoracic somites 21 and 22. This analysis yielded five distinct clusters. Notably, clusters 3 and 5 exhibited expression patterns akin to *HOXC6* and *HOXC8/9*, respectively. Genes within these clusters were significantly associated with anterior-posterior patterning and anatomical structure development, including for example *HAPLN1*, a gene known for its structural support in cartilage formation and *PDGF*, a growth factor expressed in skeletal cells or osteoblasts.

2.2. Differential chromatin accessibility of somites across the C-T boundary identifies cis-regulatory elements for *HOXC* cluster genes

To identify regulatory elements associated with differentially expressed genes we performed ATAC-seq using matching somites. As before, libraries were prepared on six individual somites (17–22) spanning the C-T axial boundary. Four biological replicates per somite were sequenced. PCA showed that samples segregated according to axial level origin (Supplementary Fig. 3A and B). Data analyses revealed distinct chromatin accessibility profiles across C-T somites, with 27,583 differential peaks reflecting dynamic patterns that correlate with developmental progression along the axis (Supplementary Fig. 3C). We observed characteristic profiles with peaks surrounding gene loci and focussed on the *HOXC* cluster genes, *HOXC6* and *HOXC8*, which are associated with the C-T transition (Fig. 1E and F) across species; both

genes were highly differentially expressed. Classic experiments showed that heterotypic grafting of thoracic presegmented mesoderm into a cervical location leads to ectopic rib formation (Kieny et al., 1972). This observation suggested that level-specific morphogenetic capacity of the cervical and thoracic somitic mesoderm is determined before metamorphosis occurs, however, the molecular mechanisms were not clear at the time. Here we grafted presegmented thoracic level tissue into cervical regions (Huang et al., 2000) and show that expression of *HOXC6* was activated in the graft, in an ectopic location (Fig. 2A). This demonstrates that PSM cells are committed to a particular axial identity and is in contrast to bipotential neuromesodermal progenitor cells (NMP) in the tail bud, which when grafted heterotypically can reset their *HOX* identity to match the new environment (McGrew et al., 2008). Activation of *HOXC6* expression in a PSM-graft of thoracic-level origin is consistent, as it demarcates the anterior thoracic boundary in many vertebrates (Burke et al., 1995; Gaunt, 1994; Mansfield and Abzhanov, 2010; Woltering et al., 2009).

The consecutive transcriptional activation of *HOXC5*, *HOXC6* and *HOXC8* genes across the C-T boundary suggests the presence of specific CREs for each gene. Thus, we examined accessible chromatin peaks flanking the *HOXC5*, *HOXC6* and *HOXC8* loci, both upstream and downstream, focusing on differential peaks that showed sequence conservation in other vertebrates. Candidate CREs were cloned into a citrine reporter plasmid containing the herpes simplex virus thymidine kinase (HSV-TK) minimal promoter (Mok et al., 2021; Williams et al., 2018). Citrine reporters were electroporated *in ovo* into the neural tube. Compared to microinjections of individual somites, the hollow centre of the neural tube is easy to microinject, with the plasmid solution spreading along its length. Since the anterior expression boundaries are well characterised, we decided to use this approach, to validate CREs for *HOXC5*, *HOXC6* and *HOXC8*. We observed discrete boundaries mimicking endogenous patterns of the cognate genes (Figs. 1–3, Supplementary Fig. 4). Therefore, the neural tube is a valid system to test reporter activity.

A differential peak of 342 bp, which is more accessible in thoracic

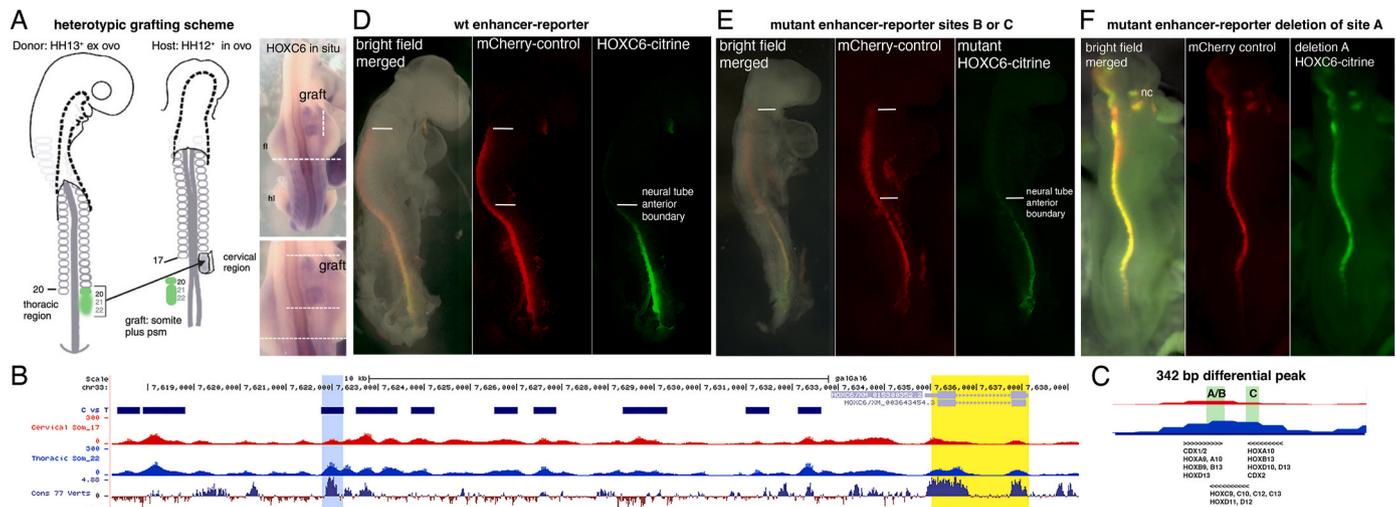


Fig. 2. Identification of a cis-regulatory element for *HOXC6*. (A) Schematic of heterotypic graft. One somite (numbered 20) and the following presegmented mesoderm (PSM) of thoracic tissue equivalent to two somites in length were removed from a HH13⁺ donor embryo and transplanted in place of cervical level somite 15 plus the following PSM in HH12⁺ host embryo *in ovo*. Two days post-op the ectopic thoracic-level derived PSM has produced somites, which express *HOXC6* in the cervical environment (vertical dashed line). This is also shown in higher magnification, horizontal dashed line indicates limit of endogenous *HOXC6* expression. (B) To identify a regulatory element for *HOXC6* differential ATAC-peaks flanking the *HOXC6* locus (yellow) were cloned and validated. Blue box indicates the putative cis-regulatory element. Black bars indicate other differential peaks called in this region. (C) Magnified image of CRE sequence with transcription factor binding sites, overlapping in region A/B and in region C (green). (D) The *HOXC6* citrine reporter shows an anterior cut-off in neural tube at forelimb level, compared to the mCherry control plasmid, which is not restricted (n = 10/10). (E) Mutagenesis (substitution) of individual TF binding sites located in B or C did not affect anterior boundary of *HOXC6* enhancer citrine reporter expression in neural tube, representative embryo shown for n = 6/6 each. (F) After deletion of region A, the *HOXC6* citrine reporter no longer showed an anterior cut-off and citrine expression extended along the length of the neural tube including the neural crest, nc (n = 8/8). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

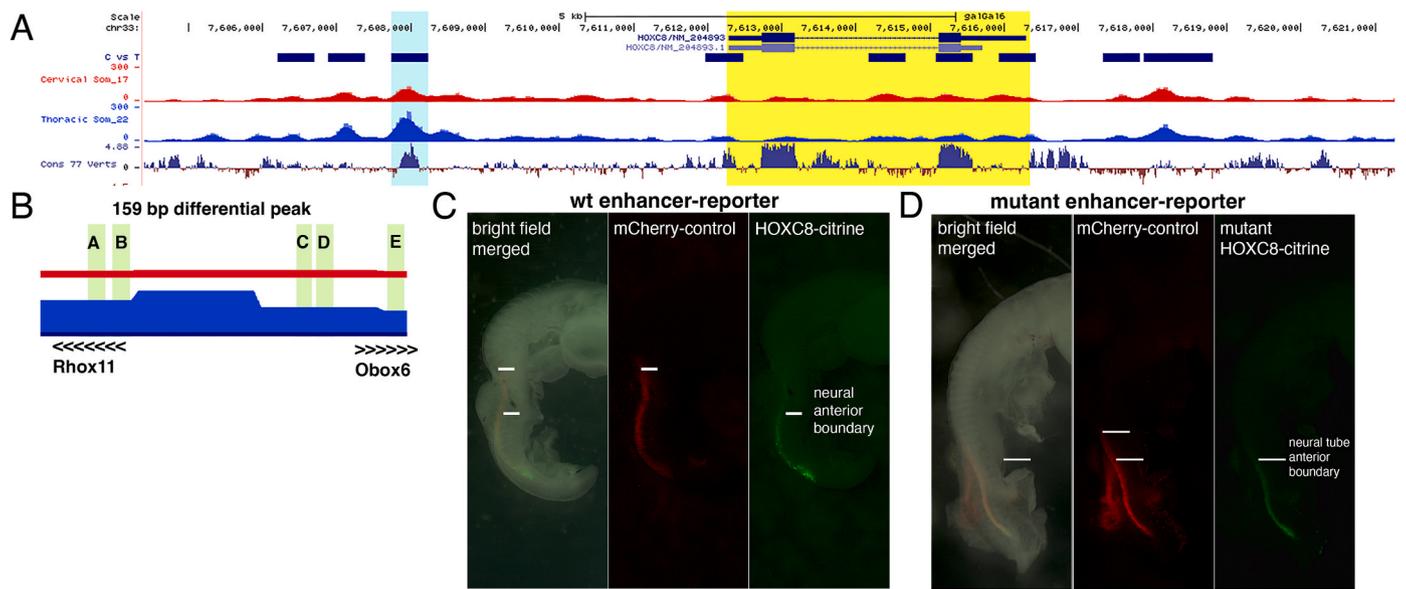


Fig. 3. Identification of a cis-regulatory element for HOXC8. (A) ATAC-Seq profile at the HOXC8 locus. Yellow shading indicates the HOXC8 gene. Blue shading indicates the CRE identified. Black bars indicate location for all differential peaks identified across the HOXC8 locus. (B) Schematic representation of putative enhancer with regions A-E (green), three of these (A, B, E) contain footprints. (C) HOXC8 enhancer citrine reporter with distinct cut-off compared to mCherry expression in neural tube (NT) ($n = 13/13$). (D) HOXC8 citrine reporter expression in neural tube (NT) following mutagenesis of individual footprints, region A ($n = 2/3$), region B ($n = 4/4$) or region E ($n = 4/4$). The example shown is a region B mutant. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

somites, was identified 12 kb upstream of *HOXC6* (Fig. 2B). HINT-ATAC (Li et al., 2019) identified genome-wide *in silico* footprints, including for CDX1, CDX2 and several mid-level and posterior HOX genes (Fig. 2C). The *HOXC6*-citrine reporter was co-injected and electroporated with a ubiquitously expressed mCherry plasmid control into the neural tube. Expression of fluorescence was monitored after overnight incubation. Citrine expression highlighted an anterior boundary in the neural tube at the forelimb level (somite 17). This correlates well with the endogenous *HOXC6* expression. Expression of the mCherry control plasmid, extended more anteriorly (Fig. 2D). Mutagenesis using nucleotide substitutions of footprint sites B or C within the *HOXC6* reporter had no effect on the anterior boundary of citrine expression suggesting that they are not essential (Fig. 2E). Another possibility is a repressive role in more posterior regions, as some of the HOX members called are expressed more posterior to the boundary captured, with the exception of *HOXC9*, which is expressed in thoracic somites. The remaining footprint within site A is for CDX1/2, which have been well documented to be involved in anteroposterior patterning and posterior axis elongation (van den Akker et al., 2002). Other mid-level HOX members called for site A are *HOXA9* and *HOXB9*, which are expressed in thoracic somites. Interestingly, a deletion of site A from the *HOXC6* citrine reporter led to anteriorly expanded citrine expression and loss of restriction to forelimb level (Fig. 2F). This suggests that site A is required to repress anterior expression of *HOXC6*. Several additional differential peaks around the *HOXC6* locus were tested. Most of these peaks were not conserved and no citrine expression was detected with those fragments. It is possible that some could be active only in somite tissue and this remains to be tested.

Next, we examined the *HOXC8* locus and identified a peak located 3 kb upstream that was conserved with other vertebrates and differentially accessible in thoracic somites compared to cervical somites (Fig. 3A). The sequence of this candidate CRE was the same as a previously identified element in mouse and chick (Shashikant et al., 1995). HINT-ATAC identified potential transcription factor footprints within this region, for RHOX11 and Obox6 (Fig. 3B). The activity of this 159 bp region was validated by *in ovo* electroporation of a *HOXC8* citrine reporter plasmid electroporated into the neural tube of stage HH10

embryos. Citrine fluorescence was detected with the anterior boundary reflecting the endogenous expression of *HOXC8* in the neural tube (Fig. 3C). Expression of the ubiquitous mCherry control plasmid extended more anteriorly. Mutations within the A/B region of this element comprising a footprint site for RHOX11 did not affect citrine expression. Similarly, mutation of the Obox6 footprint found in region E had no effect and *HOXC8* citrine-reporter expression was observed after *in ovo* electroporation (Fig. 3D). This suggests RHOX11 or OBOX6 are not essential for the regulation of this *HOXC8* CRE across the C-T boundary. The previously identified short elements (7 bp in regions C and D) did not have footprints and were not tested any further (Shashikant and Ruddle, 1996). We propose that different TFs regulate *HOXC6* and *HOXC8* as footprints called are different in each case. However, their role remains speculative.

Using the same approach, we identified a 658bp long putative CRE located 13 kb downstream of *HOXC5*. This element displayed enhancer activity in the neural tube following electroporation, with a pattern comparable to that of the endogenous *HOXC5* *in situ* expression (Supplementary Fig. 4).

2.3. Potential downstream targets of *HOXC6* include SOX transcription factors and lncRNAs

To uncover downstream targets of *HOXC6*, we examined genes differentially expressed across the C-T boundary for differentially accessible chromatin containing a *HOXC6* footprint. This approach identified nine potential *HOXC6* target genes. This included a trio of transcription factor genes pivotal in chondrogenesis: *SOX5*, *SOX6*, and *SOX9*, which showed reduced expression in thoracic somites (Fig. 4A). Footprint coverage for all three SOX transcription factors was concomitantly reduced in thoracic samples compared to cervical. *SOX9* is a transcriptional activator required for chondrogenesis. A differential *HOXC6* footprint was detected 10 kb upstream of its promoter (Fig. 4B). *SOX5* and *SOX6* are closely related proteins that enhance *SOX9* function (Liu and Lefebvre, 2015). Their genomic loci also included *HOXC6* footprints within differentially accessible chromatin (Fig. 4C and D). Due to their roles in chondrogenesis it is plausible to speculate that these

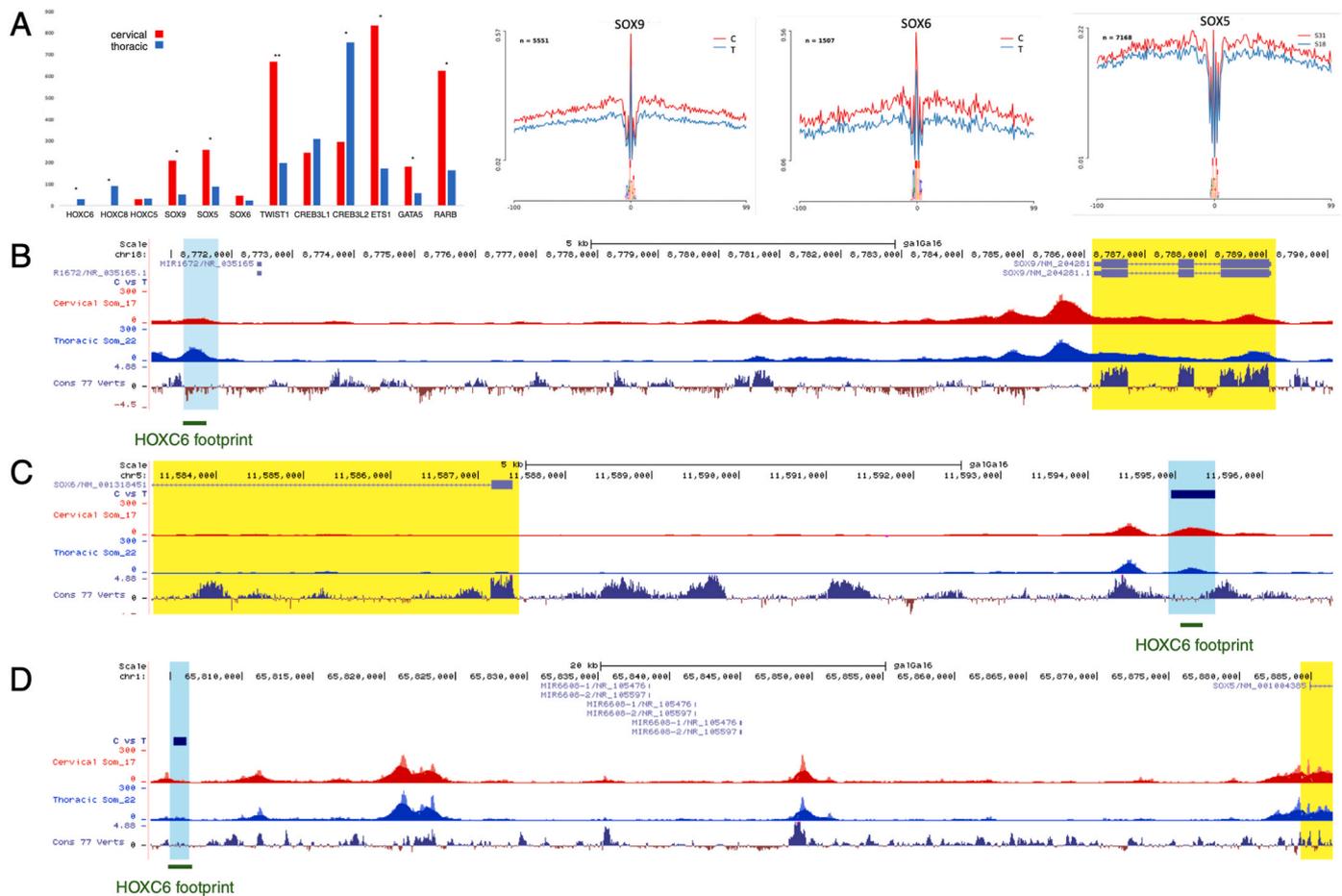


Fig. 4. HOXC6 footprints indicate target genes. (A) RNA-sequencing data for genes with differential expression across the C-T boundary. Differential read coverage for footprints for the trio of SOX transcription factors, SOX9, SOX6 and SOX5 in cervical (red) and thoracic somites (blue). (B) ATAC-sequencing profile of the *SOX9* locus, (C) the *SOX6* locus, and (D) the *SOX5* locus. Yellow boxes indicate the gene body, blue boxes indicate a region containing a HOXC6 footprint indicated below by a green bar. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

genes may be important for implementing the change in vertebrae anatomy across the C-T boundary. The other downregulated genes with a HOXC6 footprint were *TWIST1*, *ETS1*, *GATA5* and *RARB*, while Cyclic AMP-Responsive Element-Binding Protein 3-Like Proteins 1 and 2, *CREB3L1* and *CREB3L2*, were upregulated in thoracic somites. Interestingly, *CREB3L1* is a transcriptional regulator of *COL1A1* (Murakami et al., 2009) and *CREB3L1* mutations in humans are associated with Osteogenesis Imperfecta (Keller et al., 2018; Lindahl et al., 2018). *CREB3L2* is critical for chondrogenesis by activating the transcription of *SEC23A*, which promotes the transport and secretion of cartilage matrix proteins (Hino et al., 2014).

The RNA sequencing revealed previously uncharacterized lncRNAs with differential expression across the C-T axial boundary in chick embryos. Among the 343 differential genes, 18 were novel lncRNAs with 10 in the top 25 differentially expressed genes, emphasizing their potential significance for axial differentiation. The lncRNAs did not overlap with protein-coding genes and showed no predicted protein-coding potential based on analysis using predictprotein.com (Bernhofer et al., 2021). We further examined a 3941 bp lncRNA (ENSGALG0000053345) located 16 kb downstream of *HOXC4*, with a differential ATAC-peak which may represent its promoter (Fig. 5A). *In situ* hybridisation showed that this sequence was expressed in a pattern similar to *HOXC6* (Fig. 5B), raising the possibility of shared functions in determining thoracic identity. We examined whether lncRNA ENSGALG0000053345 is regulated by thoracic *HOXC* cluster genes, *HOXC6* or *HOXC8*, using a CRISPR-on approach (Gimenez et al., 2016). Short guide RNAs directed a modified CRISPR-CAS9 linked to the strong transcriptional activator, VP16,

to the promoters of either *HOXC6* or *HOXC8* using electroporation into gastrula stage embryos in EC culture (Fig. 5C and D). The specific guides led to ectopic *HOXC6* and *HOXC8* expression, detected by *in situ* hybridisation prominently in extraembryonic regions (Fig. 5C and D). In addition, CRISPR-on of *HOXC6* but not *HOXC8* resulted in ectopic expression of lncRNA ENSGALG0000053345. This suggests that this lncRNA may be a *HOXC6* target, therefore we named it HOXC6TA.

We asked whether lncRNA HOXC6TA could act as a microRNA sponge by analysing the presence of potential microRNA binding sites. We found multiple seed sequences for several microRNAs, some of which are associated with vertebrate evolution and gastrulation (Supplementary Table 1) (International Chicken Genome Sequencing, 2004; Shao et al., 2012). Whilst this is an interesting observation, the function of HOXC6TA as a sponge and the significance of the candidate microRNAs for thoracic vertebrate development remain hypothetical and require future experimental validation.

3. Conclusion

Here we generated molecular profiles of somites across the C-T boundary. This benefited from the accessible chicken embryo, a classic model for vertebrate development. We demonstrate the value of these datasets for identification of differentially expressed genes and of cis-regulatory elements. Proof-of-principle experiments validate CREs for several *HOX* genes including for *HOXC6*, which defines the first thoracic segment. *In silico* footprint analysis identified CDX1 and CDX2, both of which have been well documented to be involved in anteroposterior

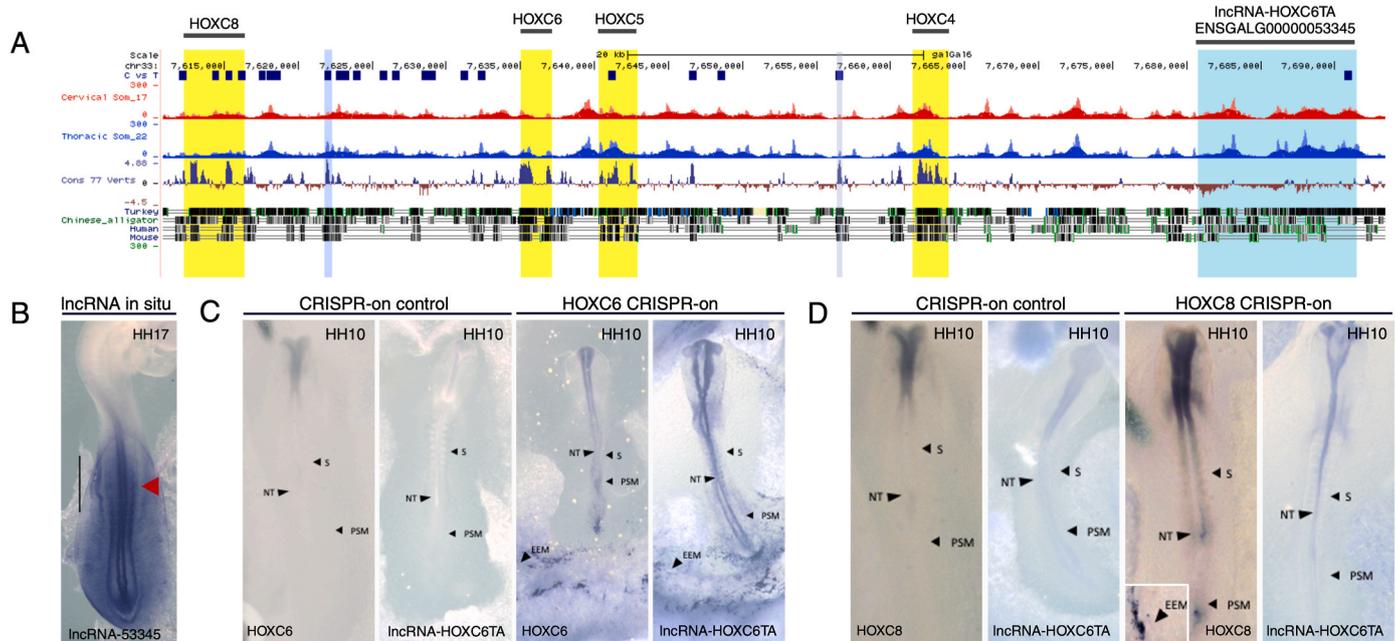


Fig. 5. IncRNA HOXC6TA is regulated by HOXC6 but not HOXC8. (A) ATAC-sequencing profile of the HOXC cluster. Yellow boxes indicate location of genes across the cluster with the gene names above, the pale blue box indicates the HOXC6 CRE, the grey box indicates the HOXC5 CRE, the turquoise box indicates the novel lncRNA identified (ENSGALG0000053345). Navy blue bars indicate location for all differential peaks identified across the genome section shown. Conservation across selected vertebrate species is indicated below the tracks. (B) In situ hybridisation detects expression of lncRNA at HH17. The red arrow indicates the level of the first thoracic somite. The emerging forelimb bud is indicated by a black line. (C) Electroporation of CRISPR-on control or of HOXC6-CRISPR-on at HH4 gastrula stages as indicated. Embryos were incubated to HH10 and in situ hybridisation detected ectopic expression of HOXC6 and lncRNA-53345, renamed lncRNA HOXC6TA. (D) Electroporation of CRISPR-on control or of HOXC8-CRISPR-on at HH4 gastrula stages as indicated. Embryos were incubated to HH10 and ectopic expression of HOXC8 but not lncRNA was detected by in situ hybridisation. Black arrows indicate embryonic tissues: neural tube (NT), somites (S), presomitic mesoderm (PSM), and extraembryonic membrane (EEM). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

patterning and posterior axis elongation (van den Akker et al., 2002). Footprints for HOXC6 are associated with several genes differentially expressed across the C-T boundary that are known to be involved in chondrogenesis, notably a trio of SOX transcription factors. We postulate that HOXC6 mediated regulation of these genes in cervical and thoracic somites may be involved in regulating the distinct morphologies of vertebrae derivatives that are generated subsequently. This hypothesis remains to be tested and future experiments should verify the roles of SOX9, SOX6 and SOX5, TWIST1, ETS1, GATA5 and RARB, CREB3L1 and CREB3L2, and their regulation by HOXC6.

4. Materials and methods

4.1. Chicken embryo dissection

Fertilised white leghorn chicken eggs were ordered weekly from Henry Stewart & Co. Ltd, UK and stored at 17°C prior to incubation. To initiate development, the eggs were then transferred to a humid incubator and kept at 38°C until they reached Hamburger and Hamilton (HH) stages, HH13 and HH14, when embryos have developed between 19 and 22 pairs of somites. Embryos were dissected into Ringer's solution, which was replaced with Dispase (1.5 mg/ml) in DMEM 10 mM HEPES pH7.5 at 37 °C for 7 min prior to treatment with Trypsin (0.05 %) at 37 °C for 7 min. The reaction was stopped with Ringer's solution with 0.25 % BSA. The somites were dissected away from neural and lateral mesoderm tissue using sharp tungsten needles and were used for RNA-seq and ATAC-seq. Gastrula stage HH4 embryos were used in EC culture for injection and electroporation. All experiments were performed on chicken embryos younger than 14 days of development and therefore were not subject to regulation by the Animal Scientific Procedures Act 1986.

4.2. RNA extraction, library preparation and sequencing and ATAC-seq

Individual dissected somites were placed into RLT lysis buffer. RNA was extracted using Qiagen RNAeasy kit (Cat no. 74104) and DNase treated (Qiagen Cat no 79254) for removal of DNA. Libraries were prepared and sequenced on the Illumina HiSeq4000 platform (75 bp paired end) at the Earlham Institute. A minimum of four biological replicates for each stage were used for analysis. For ATAC-sequencing, dissected somites were transferred immediately into 10 µl of cold lysis/tagmentation buffer in a 96 well plate for 15 min, with intermittent pipetting to attain a single cell suspension. The plate was then spun for 10 s at 500 g and tagmentation was carried out for 1 h at 37°C on a shaking thermomixer using the Illumina Nextera DNA kit (FC-121-1030). Tagmented DNA was purified following manufacturer's instructions using Qiagen MinElute kit (Cat no. 28004) and then amplified using NEB Next High-Fidelity 2X PCR Master Mix (Cat no. M0543S) and a universal AD1 forward primer with individual reverse primers per sample, for indexing. Library preparation was complete after further clean up using Qiagen PCR MinElute kit (Cat no. 28004) and Beckman Coulter XP AMPpure beads (A63880). Tagmented fragment size was assessed using Agilent 2100 Bioanalyser. Libraries were quantified with Qubit 2.0 (Life Technologies) and sequenced using paired-end 150bp reads on the Illumina HiSeq4000 platform at Novogene UK.

4.3. Enhancer cloning

Chick genomic DNA (gRNA) was extracted from HH14 embryos using Invitrogen Purelink gDNA extraction kit (Cat no. K1820-00). Putative enhancers were amplified using primers with specific sequence tails to enable cloning into reporter vector using a modified Golden Gate protocol under the following conditions: 94 °C, 3 min; 10 cycles of:

94 °C, 15 s; 55 °C, 15 s; 68 °C, 3 min, 25 cycles of 94 °C, 15 s; 63 °C, 15 s; 68 °C, 3 min; and final step of 72 °C, 4 min. Amplicons were purified using Qiagen PCR Cleanup (Cat no. 28104) and 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 min; a single step of 55 °C, 5 min; and a final step of 80 °C, 5 min. For mutagenesis of specific sites in enhancers we utilised FastCloning methodology.

4.4. Embryo preparation and ex ovo electroporation

Hamburger and Hamilton (HH3+) embryos were captured using the filter paper based easy-culture method (Chapman et al., 2001). Briefly, eggs were incubated for ~20 h, 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 Ringer's solution and excess yolk was removed. The embryo was transferred into a dish containing albumin-agar with the ventral side up and plasmid DNA (1 µg/µl) in 10 % injecting dye (Fast Green FCF, Sigma-Aldrich; Cat no./ID F7252) was injected between the vitelline membrane and embryo and electroporated used five pulses (5 V, 50ms at 100ms intervals). Thin albumin was used to seal the lids of dishes and embryos were cultured at 37 °C to the desired stage.

4.5. In ovo plasmid injection and electroporation

Eggs were incubated with the blunt end up where they were then punctured with a forceps and a large window above the embryo was created for easy accessibility. Egg shell membranes were removed and a solution of 1:500 PBS/Pen Strep (PS) with black India Ink (Winsor & Newton) was injected beneath the embryo using a 1 ml syringe with a 25Gx 5/8" needle for contrast and then 2 drops of 1:500 PBS/PS-solution was applied above the embryo. A small incision was made in the vitelline membrane above the injection site of the embryo, a micromanipulator was used to direct the needle into the neural tube for injection. Stage HH10 chick embryos were injected into the brain and at multiple sites along the neural tube. Electroporation was performed using five pulses (60 V for 50ms at 100ms intervals) using a square wave electroporator, ~3 ml of thin albumin was removed using a syringe, the egg was sealed with tape and incubated at 38°C for ~20 h.

4.6. Wholemout in situ hybridisation

Wholemout in situ hybridisation used DIGUTP labelled antisense RNA probes for HOXC5, HOXC6, HOXC8 and lncRNA and standard methods. Briefly, following fixation in 4 % PFA embryos were treated with Proteinase K, hybridised over night at 65 °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. Wholemout 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.

4.7. RNA-sequencing analysis

Four biological replicates were used for analysis. Adaptors 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 galGal6 cDNA using kallisto (v.0.44.0). Resultant quantification files were collated to generate an expression matrix. Differential expression and pathway analyses were

then conducted using the DESeq2 package in R Studio (v.1.3.959) and default settings within the iDEP (v.9.51) web interface. GO term analysis (Gene Ontology) was carried out using g:Profiler, a web server for functional profiling and interpretation of gene lists. Principal component analysis (PCA) based on expression count data revealed a discernible batch effect and replicate samples clustered based on time of dissection (1.5 h apart) rather than on cervical versus thoracic origin. The batch effect was attributed to developmental timing disparities, specifically the phase of somitogenesis. This was supported by expression fluctuations in key genes associated with the segmentation process such as *HAIRY1/2* and *LFNG* (Dale et al., 2003; Jouve et al., 2000; McGrew et al., 1998; Palmeirim et al., 1997), which showed variation between replicates. The effect was mitigated by applying the `Combat` function from the BioConductor package to yield adjusted data that eliminated batch variations and normalized the dataset (Supplementary Fig. 1A and B).

4.8. ATAC-sequencing and in silico foot printing analysis

Four biological replicates were sequenced at 30 million reads (150 paired-end). Analysis followed a previously established work flow (Mok et al., 2021). Adaptors were removed from raw paired-end sequencing reads and trimmed for quality using Trim Galore! (v.0.5.0) a wrapper tool around Cutadapt and FastQC. Default parameters were used. Quality control (QC) was performed before and after read trimming using FastQC (v.0.11.6) 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/>). Reads were mapped to the chicken genome galGal6 assembly using bowtie2 (v.2.3.4.2). The resultant Sequence Alignment Map (SAM) files were compressed to the Binary Alignment Map (BAM) version on which SAMtools (v.1.9) 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) and peaks were called using MACS2 (v.2.1.1) (parameters -f BAMPE -g mm -B -nomodel -shift -100 -extsize 200). Coverage tracks and peaks (narrow peak format) were uploaded to the UCSC Genome Browser as custom tracks for ATAC-seq data visualization. Analysis for differential accessibility was carried out in R (v.3.5.1) using the DiffBind package (v.2.8.0) with default parameter settings. Differential accessibility across samples was calculated using the negative binomial distribution model implemented in DESeq2 (v1.4.5). Computational footprinting analysis was conducted across samples using HINT-ATAC which is part of the Regulatory Genomic Toolbox (v.0.12.3) also using default parameter settings and the galGal6 genome.

4.9. CRISPR-ON primer design

Potential Cas9 target sites were identified by scanning manually the promoter regions of HOXC6 and HOXC8 for PAM sequences (NGG). Identified sequences were then surveyed genome-wide using BLAT to ensure that they had a single hit in the genome and then run through oligo-calc to identify any potential self-annealing sequences. Forward and reverse oligos for each guide recognition sequence (18-20 nt) directly preceding the PAM (NGG), but not include the PAM itself, were designed. Sequences including a G base at the 5' end were preferentially selected for, as this is required for polIII to initiate transcription from a

U6 promoter. In the case for sequences where this was not possible, a G residue was added upstream of the selected recognition sequence. Each of the sgRNA sequences used also included flanking BsmBI sites and corresponding overhangs for Golden Gate cloning into the U6 vector, where multiple fragments of DNA can be assembled by using combinations of overhang sequences in a single PCR step.

CRedit authorship contribution statement

Shannon A. Weldon: Writing – original draft, Visualization, Validation, Methodology, Investigation. **Emily L. Smith:** Writing – review & editing, Visualization, Validation, Investigation. **Magdalena Schatka:** Investigation, Validation, Visualization, Writing – review & editing. **Leighton Folkes:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Gi Fay Mok:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Ashleigh Lister:** Writing – review & editing, Methodology. **Iain C. Macaulay:** Methodology. **Wilfried Hearty:** Writing – review & editing, Supervision, Formal analysis, Data curation. **Andrea E. Münsterberg:** Writing – original draft, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Funding source

Shannon A. Weldon, Emily L. Smith and Magdalena Schatka were supported by BBSRC NRPDTP studentships. LF and GFM were supported by BBSRC grant BB/N007034/1 to AM. WH, IM, and AL were supported by the BBSRC Core Strategic Programme Grant (BB/CSP1720/1) and its components (BBS/E/T/000PR9818, BBS/E/T/000PR9819, BBS/E/T/000PR9817) and the BBSRC Earlham Institute Strategic Programme Grant Cellular Genomics BBX011070/1 (BBS/E/ER/230001A, BBS/E/ER/230001B, BBS/E/ER/230001C). IM and AL were supported by the BBSRC Core Capability Grant BB/CCG1720/1 and the National Capability (BBS/E/T/000PR9816).

Acknowledgements

We thank Grant Wheeler for helpful discussions. Graphical abstract and schematic in Fig. 2 were created with BioRender.com.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ydbio.2025.06.012>.

Data availability

The data presented in the study are deposited in the European Nucleotide Archive repository, accession number PRJEB79488. (embargoed until acceptance)

References

- Alvares, L.E., Schubert, F.R., Thorpe, C., Mootosamy, R.C., Cheng, L., Parkyn, G., Lumsden, A., Dietrich, S., 2003. Intrinsic, Hox-dependent cues determine the fate of skeletal muscle precursors. *Dev. Cell* 5, 379–390.
- Aulehla, A., Pourquie, O., 2010. Signaling gradients during paraxial mesoderm development. *Cold Spring Harbor Perspect. Biol.* 2, a000869.
- Benazeraf, B., Pourquie, O., 2013. Formation and segmentation of the vertebrate body axis. *Annu. Rev. Cell Dev. Biol.* 29, 1–26.
- Bernhofer, M., Dallago, C., Karl, T., Satagopam, V., Heinzinger, M., Littmann, M., Olenyi, T., Qiu, J., Schutze, K., Yachdav, G., Ashkenazy, H., Ben-Tal, N., Bromberg, Y., Goldberg, T., Kajan, L., O'Donoghue, S., Sander, C., Schafferhans, A., Schlessinger, A., Vriend, G., Mirdita, M., Gawron, P., Gu, W., Jarosz, Y., Trefois, C., Steinegger, M., Schneider, R., Rost, B., 2021. PredictProtein - predicting protein structure and function for 29 years. *Nucleic Acids Res.* 49, W535–W540.
- Berti, F., Nogueira, J.M., Wohrle, S., Sobreira, D.R., Hawrot, K., Dietrich, S., 2015. 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.
- Borman, W.H., Yorde, D.E., 1994. Analysis of chick somite myogenesis by in situ confocal microscopy of desmin expression. *J. Histochem. Cytochem.* 42, 265–272.
- Brent, A.E., Tabin, C.J., 2002. Developmental regulation of somite derivatives: muscle, cartilage and tendon. *Curr. Opin. Genet. Dev.* 12, 548–557.
- Burke, A.C., Nelson, C.E., Morgan, B.A., Tabin, C., 1995. Hox genes and the evolution of vertebrate axial morphology. *Development* 121, 333–346.
- Burke, A.C., Tabin, C.J., 1996. Virally mediated misexpression of Hoxc-6 in the cervical mesoderm results in spinal nerve truncations. *Dev. Biol.* 178, 192–197.
- Carapuco, M., Novoa, A., Bobola, N., Mallo, M., 2005. Hox genes specify vertebral types in the presomitic mesoderm. *Genes Dev.* 19, 2116–2121.
- Chal, J., Pourquie, O., 2017. Making muscle: skeletal myogenesis in vivo and in vitro. *Development* 144, 2104–2122.
- Chapman, S.C., Collignon, J., Schoenwolf, G.C., Lumsden, A., 2001. Improved method for chick whole-embryo culture using a filter paper carrier. *Dev. Dyn.* 220, 284–289.
- Christ, B., Huang, R., Scaal, M., 2007. Amniote somite derivatives. *Dev. Dyn.* 236, 2382–2396.
- Christ, B., Ordahl, C.P., 1995. Early stages of chick somite development. *Anat. Embryol.* 191, 381–396.
- Christ, B., Scaal, M., 2008. Formation and differentiation of avian somite derivatives. *Adv. Exp. Med. Biol.* 638, 1–41.
- Cohn, M.J., Tickle, C., 1999. Developmental basis of limblessness and axial patterning in snakes. *Nature* 399, 474–479.
- Dale, J.K., Maroto, M., Dequeant, M.L., Malapert, P., McGrew, M., Pourquie, O., 2003. Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. *Nature* 421, 275–278.
- Dequeant, M.L., Pourquie, O., 2008. Segmental patterning of the vertebrate embryonic axis. *Nat. Rev. Genet.* 9, 370–382.
- Deschamps, J., Duboule, D., 2017. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev.* 31, 1406–1416.
- Gaunt, S.J., 1994. Conservation in the Hox code during morphological evolution. *Int. J. Dev. Biol.* 38, 549–552.
- Gimenez, C.A., Ielpi, M., Mutto, A., Grosebacher, L., Argibay, P., Pereyra-Bonnet, F., 2016. CRISPR-on system for the activation of the endogenous human INS gene. *Gene Ther.* 23, 543–547.
- Guerreiro, I., Nunes, A., Woltering, J.M., Casaca, A., Novoa, A., Vinagre, T., Hunter, M.E., Duboule, D., Mallo, M., 2013. Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10682–10686.
- Hamburger, V., Hamilton, H.L., 1992. A series of normal stages in the development of the chick embryo. 1951. *Dev. Dyn.* 195, 231–272.
- Hino, K., Saito, A., Kido, M., Kanemoto, S., Asada, R., Takai, T., Cui, M., Cui, X., Imaizumi, K., 2014. Master regulator for chondrogenesis, Sox9, regulates transcriptional activation of the endoplasmic reticulum stress transducer BBF2H7/CREB3L2 in chondrocytes. *J. Biol. Chem.* 289, 13810–13820.
- Huang, R., Zhi, Q., Patel, K., Wilting, J., Christ, B., 2000. Contribution of single somites to the skeleton and muscles of the occipital and cervical regions in avian embryos. *Anat. Embryol.* 202, 375–383.
- Ibarra-Soria, X., Thierion, E., Mok, G.F., Munsterberg, A.E., Odom, D.T., Marioni, J.C., 2023. A transcriptional and regulatory map of mouse somite maturation. *Dev. Cell* 58, 1983–1995 e1987.
- Imura, T., Pourquie, O., 2006. Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingress. *Nature* 442, 568–571.
- International Chicken Genome Sequencing, C., 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.
- Jouve, C., Palmeirim, I., Henrique, D., Beckers, J., Gossler, A., Ish-Horowitz, D., Pourquie, O., 2000. Notch signalling is required for cyclic expression of the hairy-like gene HES1 in the presomitic mesoderm. *Development* 127, 1421–1429.
- Keller, R.B., Tran, T.T., Pyott, S.M., Pepin, M.G., Savarirayan, R., McGillivray, G., Nickerson, D.A., Bamshad, M.J., Byers, P.H., 2018. Monoallelic and biallelic CREB3L1 variant causes mild and severe osteogenesis imperfecta, respectively. *Genet. Med.* 20, 411–419.
- Kieny, M., Mauger, A., Sengel, P., 1972. Early regionalization of somitic mesoderm as studied by the development of axial skeleton of the chick embryo. *Dev. Biol.* 28, 142–161.
- Leal, F., Cohn, M.J., 2018. Developmental, genetic, and genomic insights into the evolutionary loss of limbs in snakes. *Genesis* 56.
- Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M., Costa, I.G., 2019. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* 20, 45.
- Lindahl, K., Astrom, E., Dragomir, A., Symoens, S., Coucke, P., Larsson, S., Paschalis, E., Roschger, P., Gamsjaeger, S., Klaushofer, K., Fratzl-Zelman, N., Kindmark, A., 2018. Homozygosity for CREB3L1 premature stop codon in first case of recessive osteogenesis imperfecta associated with OASIS-deficiency to survive infancy. *Bone* 114, 268–277.
- Liu, C.F., Lefebvre, V., 2015. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res.* 43, 8183–8203.
- Mansfield, J.H., Abzhanov, A., 2010. Hox expression in the American alligator and evolution of archosaurian axial patterning. *J. Exp. Zool. B Mol. Dev. Evol.* 314, 629–644.
- Maschner, A., Kruck, S., Draga, M., Prols, F., Scaal, M., 2016. Developmental dynamics of occipital and cervical somites. *J. Anat.* 229, 601–609.
- McGrew, M.J., Dale, J.K., Fraboulet, S., Pourquie, O., 1998. The lunatic fringe gene is a target of the molecular clock linked to somite segmentation in avian embryos. *Curr. Biol.* 8, 979–982.

- McGrew, M.J., Sherman, A., Lillico, S.G., Ellard, F.M., Radcliffe, P.A., Gilhooley, H.J., Mitrophanous, K.A., Cambray, N., Wilson, V., Sang, H., 2008. Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. *Development* 135, 2289–2299.
- Mok, G.F., Folkes, L., Weldon, S.A., Maniou, E., Martinez-Heredia, V., Godden, A.M., Williams, R.M., Sauka-Spengler, T., Wheeler, G.N., Moxon, S., Munsterberg, A.E., 2021. Characterising open chromatin in chick embryos identifies cis-regulatory elements important for paraxial mesoderm formation and axis extension. *Nat. Commun.* 12, 1157.
- Mok, G.F., Mohammed, R.H., Sweetman, D., 2015. Expression of myogenic regulatory factors in chicken embryos during somite and limb development. *J. Anat.* 227, 352–360.
- Mok, G.F., Turner, S., Smith, E.L., Mincarelli, L., Lister, A., Lipscombe, J., Uzun, V., Haerty, W., Macaulay, I.C., Munsterberg, A.E., 2024. Single cell RNA-sequencing and RNA-tomography of the avian embryo extending body axis. *Front. Cell Dev. Biol.* 12, 1382960.
- Murakami, T., Saito, A., Hino, S., Kondo, S., Kanemoto, S., Chihara, K., Sekiya, H., Tsumagari, K., Ochiai, K., Yoshinaga, K., Saitoh, M., Nishimura, R., Yoneda, T., Kou, I., Furuichi, T., Ikegawa, S., Ikawa, M., Okabe, M., Wanaka, A., Imaizumi, K., 2009. Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. *Nat. Cell Biol.* 11, 1205–1211.
- Nishimoto, S., Minguillon, C., Wood, S., Logan, M.P., 2014. A combination of activation and repression by a colinear Hox code controls forelimb-restricted expression of Tbx5 and reveals Hox protein specificity. *PLoS Genet.* 10, e1004245.
- Palmeirim, I., Henrique, D., Ish-Horowicz, D., Pourquie, O., 1997. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* 91, 639–648.
- Rodrigo, I., Hill, R.E., Balling, R., Munsterberg, A., Imai, K., 2003. Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development* 130, 473–482.
- Scaal, M., 2016. Early development of the vertebral column. *Semin. Cell Dev. Biol.* 49, 83–91.
- Scaal, M., 2021. Development of the amniote ventrolateral body wall. *Dev. Dyn.* 250, 39–59.
- Shao, P., Liao, J.Y., Guan, D.G., Yang, J.H., Zheng, L.L., Jing, Q., Zhou, H., Qu, L.H., 2012. Drastic expression change of transposon-derived piRNA-like RNAs and microRNAs in early stages of chicken embryos implies a role in gastrulation. *RNA Biol.* 9, 212–227.
- Shashikant, C.S., Bieberich, C.J., Belting, H.G., Wang, J.C., Borbely, M.A., Ruddle, F.H., 1995. Regulation of Hoxc-8 during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression. *Development* 121, 4339–4347.
- Shashikant, C.S., Ruddle, F.H., 1996. Combinations of closely situated cis-acting elements determine tissue-specific patterns and anterior extent of early Hoxc8 expression. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12364–12369.
- Soshnikova, N., Duboule, D., 2009. Epigenetic temporal control of mouse Hox genes in vivo. *Science* 324, 1320–1323.
- van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B.I., Deschamps, J., 2002. Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* 129, 2181–2193.
- Vinagre, T., Moncaut, N., Carapuco, M., Novoa, A., Bom, J., Mallo, M., 2010. Evidence for a myotomal Hox/Myf cascade governing nonautonomous control of rib specification within global vertebral domains. *Dev. Cell* 18, 655–661.
- Weldon, S.A., Munsterberg, A.E., 2022. Somite development and regionalisation of the vertebral axial skeleton. *Semin. Cell Dev. Biol.* 127, 10–16.
- Wellik, D.M., Capecchi, M.R., 2003. Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* 301, 363–367.
- Williams, R.M., Senanayake, U., Artibani, M., Taylor, G., Wells, D., Ahmed, A.A., Sauka-Spengler, T., 2018. Genome and epigenome engineering CRISPR toolkit for in vivo modulation of cis-regulatory interactions and gene expression in the chicken embryo. *Development* 145.
- Woltering, J.M., Vonk, F.J., Muller, H., Bardine, N., Tudeau, I.L., de Bakker, M.A., Knochel, W., Sirbu, I.O., Durston, A.J., Richardson, M.K., 2009. Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the Hox code. *Dev. Biol.* 332, 82–89.
- Wymeersch, F.J., Wilson, V., Tsakiridis, A., 2021. Understanding axial progenitor biology in vivo and in vitro. *Development* 148.
- Yamashita, S., Andoh, M., Ueno-Kudoh, H., Sato, T., Miyaki, S., Asahara, H., 2009. Sox9 directly promotes Bapx1 gene expression to repress Runx2 in chondrocytes. *Exp. Cell Res.* 315, 2231–2240.