University of East Anglia

Epigenetic Control of Skeletal Muscle Development: Identification and Characterization of Regulatory Elements

Emily Smith

Word Count: 24,621

Submitted for the Masters by Research degree

Email of correspondence: e.smith3@uea.ac.uk

Under the supervision of Professor A. Münsterberg and Dr. G. Mok

School of Biological Sciences University of East Anglia Norwich, NR4 7TJ

May 2021

Abstract

The musculoskeletal system is crucial for vertebrate organisms which rely on movement. During development, the musculoskeletal system arises from clusters of transient mesoderm termed somites. Somites contain the cellular precursors of skeletal muscle, tendons, cartilage, some bone, and the dermis. The proliferation, specification, and differentiation of the different musculoskeletal cell lineages is controlled by gene regulatory networks, where changes in gene expression patterns reflects the differentiation of somitic cellular lineages. The changes in the transcriptome of somites undergoing differentiation can also be linked to the dynamic changes in chromatin accessibility. Accessible regions of chromatin can be identified using ATAC (Assay for Transposase Accessible Chromatin), and many accessible regions have been found scattered throughout the epigenome. These regions have been linked to epigenetic regulatory mechanisms, including the cis-regulatory elements known as enhancers. By utilising both ATAC-sequencing and RNA-sequencing datasets previously generated in the Münsterberg lab, this project aimed to identify novel enhancers involved with somite cellular lineage differentiation that govern the gene regulatory networks in differentiating somites. The RNA dataset was conducted on somites removed from stage fourteen chicken embryos, which helped to identify genes involved in the differentiation of musculoskeletal lineages. Genes identified as having differential activity in the more differentiated somites compared to the pre-somitic mesoderm included ALX1, UNCX, KLHL31, MEOX2, FOXO1, SCX, MKX, TNC PAX1 and PAX9 genes. The ATAC-sequencing dataset was utilised to find regions of open chromatin near these genes of interest, which may indicate a putative enhancer region. Twenty-nine different putative enhancer regions were identified by analysis of the ATAC-sequencing dataset. To determine the spatio-temporal activity of these putative enhancer regions, they were cloned into fluorescent reported constructs, which were injected and electroporated into stage three chick embryos. Of the ten enhancer constructs tested, one enhancer associated with the *Foxo1* gene showed fluorescent reporter activity in blood islands and sporadically throughout the inter-somitic blood vessels and the dorsal aorta. Genes identified from the RNA sequencing that had yet to be characterised in the chicken embryo were visualised spatially throughout development using whole mount *in-situ* hybridisation techniques, generating gene expression profiles. An expression profile for the fork-head protein Foxo-1 was generated using *in-situ* hybridisation and cryosectioning techniques, along with the identification of the Scleraxis gene in younger embryos than previously reported. Overall, this project was able to use ATAC-sequencing to identify one cis-regulatory element associated with the Foxo-1 gene and generate a corresponding expression profile.

Contents

Chapter 1: Introduction

The musculoskeletal system is important for many vertebrate functions, including locomotion, respiration, feeding, postural support and communication. Proper formation of the musculoskeletal system relies on the coordinated development of tendons, cartilage, bone, and muscle, the majority of which arise from the somites (Kalcheim and Ben-Yair, 2005). It is crucial that we properly understand the genetic regulation involved in muscle, tendon, and cartilage development, as improper development of these tissues can result in several muscular disorders, complicating post-natal life.

1.1 Gastrulation

The somites arise from the mesodermal germ layer, which is specified during the gastrulation process. During vertebrate gastrulation, cells in the epiblast start to ingress to the interior of the embryo and form the primitive streak, which is first visible as a thickening of the epiblast. As these cells enter the primitive streak, it elongates towards the head end, and defines the posterior-anterior axis of the embryo. Cells migrate through the primitive streak, which results in the formation of the three germ layers: the mesoderm, the endoderm, and the ectoderm (Vasiev *et al*., 2010), see figure 1.1. The endoderm layer is formed by epiblast cells that migrate through the primitive pit and displace the hypoblast cells; cells

of the endoderm give rise to epithelial linings, the liver and pancreas. The mesoderm layer lies between the epiblast layer and the newly formed endoderm, and ingress at the primitive streak. Cells from a mesodermal lineage will give rise to the notochord, the musculoskeletal system, and the circulatory system. The ectodermal layer does not invaginate and become parts of the nervous system and the epidermis of the skin (Kiecker *et al*., 2015).

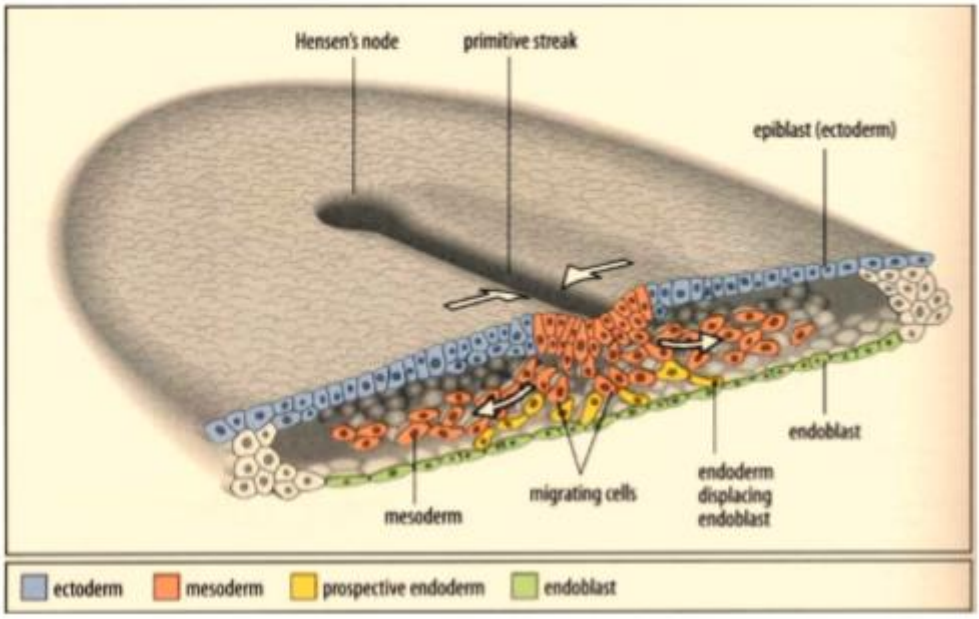

Figure 1.1: Diagram displaying vertebrate gastrulation. During this process, cells of the epiblast undergo an epithelial to mesenchymal transition (EMT) and ingresses at the primitive streak. The migrating cells form the mesodermal and endodermal layers, while those that maintain their epithelial state form the ectoderm layer. Figure from Wolpert et al., 2015.

1.2 Somite Development

Somites arise from the mesodermal compartment and are a common feature in the phylotypic stage of all Chordata. The somites are transient structures which form bilaterally on either side of the notochord, and give rise to multiple crucial body structures, including the skeletal muscle, tendons, bone, dermis and cartilage. Somites form from the paraxial mesoderm, an area of mesoderm that flanks the neural tube. The formation of somites (somitogenesis) involves the sequential segmentation of the paraxial mesoderm to form pairs of somites. In chick, one pair of somites forms about every 90 minutes, with a total of 52 somite pairs forming.

It has been demonstrated that the pre-somitic mesoderm (PSM) is specified within the primitive streak of the chicken and mouse embryo (Cambray and Wilson, 2007). Lineage analysis of somites in chicken embryos demonstrated that the progenitors that give rise to the medial half of the somite are found in the Hensen's node, an enlarged group of cells located in the anterior portion of the primitive streak, whereas progenitors that give rise to the lateral half of the somite are derived from cells found posterior to the node in the primitive streak (Selleck and Stern, 1991). These progenitors then migrate out of the streak, into the PSM, and are characterised by the expression of *Tbx6* (White, 2003). PSM continues to leave the streak as somites are formed, so there is a continual supply of pre somitic mesoderm at the caudal end of the embryo until the tail bud forms, where ingression stops (Dequéant and Pourquié, 2008).

The precise timing of somite segmentation is driven by the clock and wavefront mechanism. The first evidence for this model was the periodic expression of the transcription factor *C-HAIRY1* in the PSM (Pourquié, 2011), which suggested an oscillator associated with segmentation. The *Notch* target *Lunatic fringe* was then identified as another cyclic gene oscillating with *HAIRY* (Evrard *et al*., 1998). The segmentation clock drives the expression of the 'clock' genes, which cause the cells of the PSM to

undergo a mesenchymal to epithelial transition (MET). This results in the most anterior portion of the PSM budding off, generating an epithelial somite (Stern and Piatkowska, 2015). Other clock genes that play a role in this segmentation process include components of the *wnt*, *FGF* and *notch* signalling pathways. The Notch pathway is negatively regulated through *Lunatic Fringe* (Dale *et al*., 2003, Dequéant and Pourquié, 2008), and has shown to be crucial for clock gene oscillations and somite formation (Ferjentsik *et al*., 2009).

The determination front is an area in which the PSM can begin segmentation. This area is defined by a posterior-anterior gradient of *Fgf8* and nuclear β-catenin expression with an anterior-posterior gradient of retinoic acid (Moreno and Kintner, 2004). *Fgf8* is expressed at high levels in the posterior PSM, maintaining an undifferentiated state in these cells. When the *FGF* pathway is activated by MEK, the expression of RA-synthesising enzyme *raldh* is inhibited, preventing the synthesis of RA. Thus, the gradient of FGF and RA is determined, with *FGF* negatively regulating RA synthesis (Vermot and Pourquié, 2005). Figure 1.2 below shows this molecular determination front between *Fgf8* and RA.

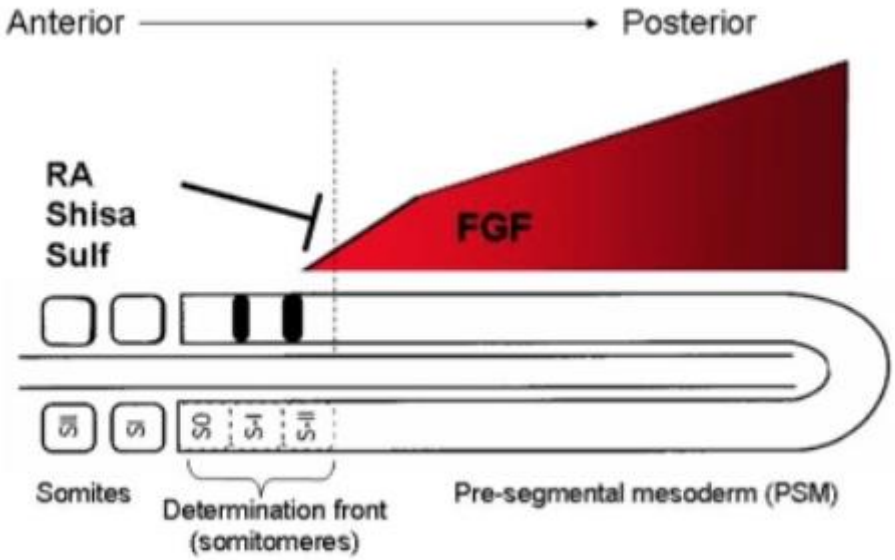

Figure 1.2: A simple diagram showing the posterior to anterior gradient of FGF establishing the determination front. Only when FGF drops beneath a certain threshold value will cells undergo MET to form somites. This threshold defines the determination front, marked by expression of lunatic fringe in chick, and mesp2 in mouse (as shown by the black lines). Adapted from Pownall and Isaacs, 2010.

Somites form once *lunatic fringe* is free from inhibition by *FGF* and notch signalling is activated, triggering a genetic cascade. *Lunatic fringe* expression is restricted to the anterior half of the prospective somite. This is one example of rostro-caudal patterning in somites, which is consistent with the re-segmentation theory. The re-segmentation theory arose from an observation that vertebrae form from the fusion of the rostral part of one somite, and the caudal end of another (Brand-Saberi and Christ, 2000). Several other genes, such as *UNCX4.1, TBX18 and C-hairy1* can be found in either the rostral or caudal half of the prospective somite (Tanaka and Tickle, 2004). The anterior-posterior separation happens later in somite differentiation, in a region termed the fissure of Von Ebner, which is found in the sclerotome. While both the sclerotomal halves will give rise to the vertebral body, the caudal half will form the vertebral pedicle, while the rostral half will give rise to the intervertebral disks (Goldstein and Kalchiem, 1992). Figure 1.3 below demonstrates how re-segmentation leads to the formation of the mature vertebrae segments.

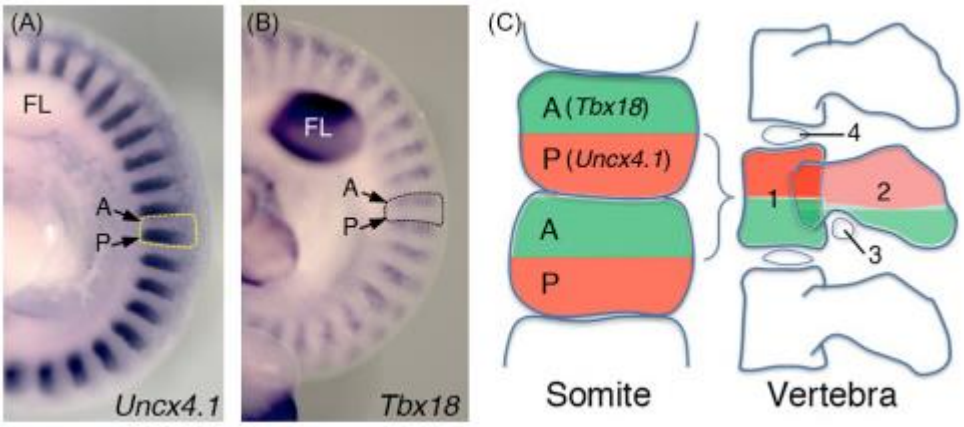

Figure 1.3: The anterior-posterior patterning of somites in the sclerotome, and how these interact to form vertebrae. A) The posterior part of the sclerotome is marked by Uncx4.1, while the anterior portion is marked by expression of TBX18 (B). C) A simple diagram showing re-segmentation of somites. Each vertebra is formed via the fusion of the posterior half of one somite and the anterior half of another (Baldock, 2016) regions depicted are: 1) vertebrae 2) the pedicle 3) Inferior vertebral notch 4) intervertebral discs.

1.2.1 Somite Differentiation and Early Patterning

Once the somite buds of from the PSM, it becomes an epithelial somite. Epithelial somites have an epithelial wall surrounding the central cavity, called the somitocoel, which contains loose mesenchyme (Scaal and Christ, 2004). This epithelial somite then begins to mature as it is exposed to inductive signals from surrounding tissues that pattern the somite (Aoyama and Asamoto, 1998). Several key changes occur after exposure to inductive signals; the ventral part of the somite undergoes EMT and forms a ventral mesenchyme, called the sclerotome, while the dorsal part of the somite remains epithelial and is termed the dermomyotome (Scaal and Christ, 2004). The dermomyotome also separates into the myotome and the dermatome, which give rise to separate tissues. As the somite continues to mature, interactions between the myotome and the underlying sclerotome give rise to a third somitic compartment, the syndetome, and the somite then described as a fully differentiated somite (see figure 1.4). These signalling pathways act to produce and differentiate somites from PSM from stage 7 to stage 25, where somitogenesis is terminated (Tenin *et al*., 2010).

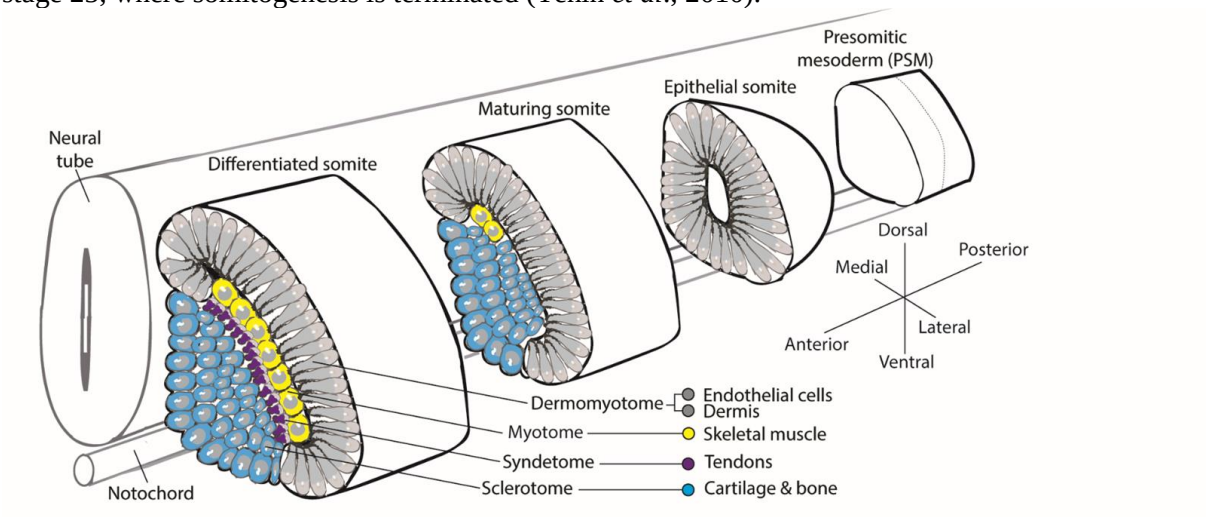

Figure 1.4: A schematic diagram showing the progressive differentiation of somites. The most posterior region is the presomitic mesoderm (PSM), which generates a new pair of epithelial somites. The somite then begins to mature, with ventral regions undergoing an EMT to give rise to two different compartments of the somite, the dermomyotome and the sclerotome. The most differentiated somite is the most anterior one, containing the sclerotome, syndetome, myotome and dermomyotome. These different regions give rise to different cell lineages (Adapted from Buckingham et al.,2003).

There are several genetic signals that trigger the formation of the sclerotome and the dermomyotome. Sonic hedgehog (*Shh*) and noggin signals from the floor plate and the notochord promote de-epithelialisation and a mesenchymal fate, forming the sclerotome in the ventral part of the somite (Borycki *et al*., 1998). Exposure to *Shh* induces *Pax1* and *Pax9* expression in the sclerotomal cells, which then triggers a genetic programme for further specification (Yusuf and Brand-Saberi, 2006) (see section 1.2.4). The sclerotome cells migrate to near the notochord and later produce the vertebrae and the ribs (Scaal and Christ, 2004). *Shh* signals from the notochord are also sufficient and necessary to activate *MyoD* gene expression in the myotomal region, beginning the muscular differentiation pathway (Borycki *et al*., 1998). Canonical Wnts such as *Wnt4*, *Wnt6* and *Wnt8c* signalling from the overlying ectoderm have been recognised as major dorsal signalling factors in the establishment of the epithelial dermomyotome. They maintain the epithelial organisation of the dorsal part of the somite, inducting expression of *Pax3* and *Pax7*, forming the dermomyotome (Münsterberg *et al*., 1995, Yusuf and Brand-Saberi, 2006).

The dorsal *Wnt* signals and the ventral *Shh* work synergistically to develop the myotome, and begins the muscular differentiation pathway (Borycki *et al*., 1998). Cells from the medial and lateral parts of the dermomyotome begin to express the myogenic factors *Myf5* and *MyoD*, giving rise to the underlying myotome, which goes onto produce the epaxial and hypaxial muscles. Canonical *Wnt* signalling can directly induce genes such as *Paraxis*, which maintains the epithelial organisation of the dorsal somite (Linker *et al*., 2005). *Shh* signals from the notochord with *Wnt* signalling regulate *MyoD* gene expression through Gli2 and Gli3 in the myotomal region (Borycki *et al* 2000). As the formation of the myotome requires both dorsal and ventral signals, it has been suggested that *Shh* and *Wnt* family members act through a concentration gradient that is established along the dorsoventral axis. This gradient is also maintained by the presence of competitive *Wnt* and *Shh* inhibitors, present in the prospective sclerotome and dermomyotome respectively. For example, the secreted frizzled-related protein 2 (*Sfrp2*) is a *Wnt* antagonist upregulated by *Shh.* It is expressed in the sclerotome, induced by *Shh*, and prevents the dorsalising effect of *Wnt* signalling (Lee *et al*., 2000). Only one candidate *Shh* antagonist has been discovered, the Growth-arrest specific gene1 (*Gas1*), which is inducible by *Wnt* proteins, (Lee *et al*., 2001). Once the myotome has separated from the dermomyotome, Neurotrophin 3 (*NTF3*) expression maintains the dermatome, which gives rise to dermis tissue (Brill *et al*., 1995). *Wnt, Shh and BMP* signals patterning the early somite is shown below in figure 1.5.

Bone morphogenetic proteins, (*BMPs*) are another key family of proteins that work to pattern the somite. They play a variety of differing roles in the patterning of the somite. *BMP4* is expressed in the dorsal neural tube where it indirectly induces the expression of *Wnt11* in the dorsomedial lip of the dermomyotome (Marcelle *et al*., 1997). This is the region that then forms the myotome and begins to express *MyoD*. However, other work suggests that *BMP4* negatively regulates myogenesis, by preventing expression of *MyoD* and *Myf5* (Reshef *et al*., 1998, Pourquie *et al*., 1996). *Noggin* antagonises *BMP* proteins by preventing them from binding their receptors (Zimmerman *et al*., 1996), and is therefore present in the medial somite. This work suggests a dual role for *BMP-4*, depending on where its expression is present. When present in the dorsal neural tube, it indirectly induces expression of *wnt*-11 in the dermomyotome. However, when BMP4 is expressed in the lateral plate mesoderm, it induces the differentiation of dermomyotome cell into *Met* expressing cells and represses myogenic differentiation. The action of BMP signalling on somite patterning is shown below in figure 1.5 and 1.6.

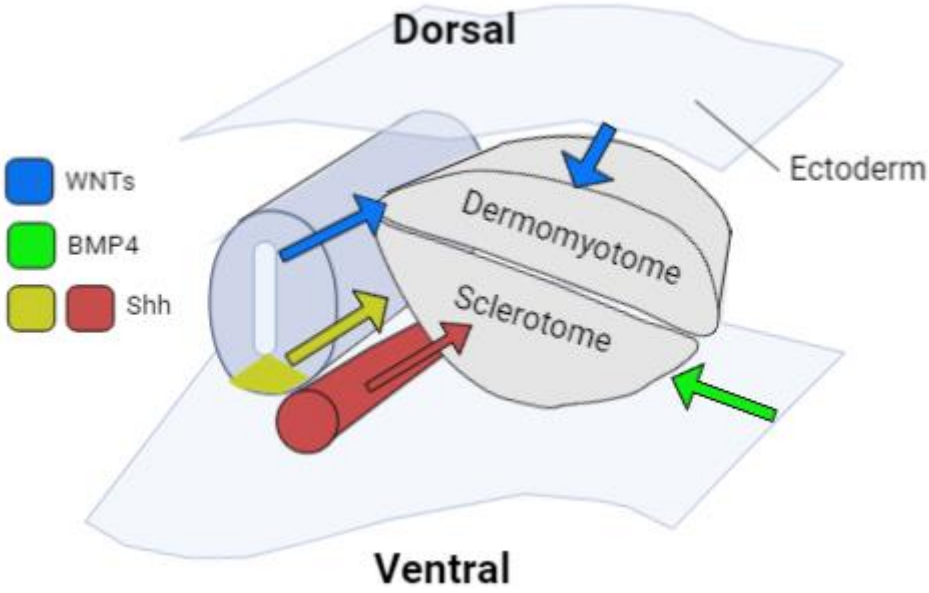

Figure 1.5: A simplistic schematic depicting a section through the somite and neural tube demonstrating some of the signalling pathways involved in the early formation of the somite. Canonical wnts from the dorsal neural tube and surface ectoderm induce the dermomyotome. Shh signals from the neural tube floor plate and notochord along with BMP4 signals from the lateral plate mesoderm maintains the sclerotomal fate.

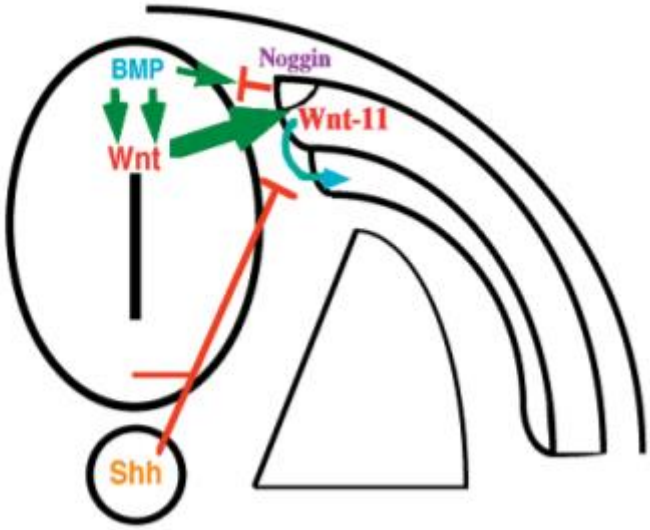

Figure 1.6: A model of BMP patterning of the dermomyotome. BMP regulates Wnt1 expression in the dorsal neural tube, which then induces expression of wnt11 in the medial lip. BMP is antagonised and therefore prevented from signalling by Noggin in the medial somite. Wnt-11 expression allows the present muscle progenitors to become the myotome (blue arrow). Shh represses Wnt-11, maintaining the dorsal expression of wnt-11. Figure from Marcelle et al., 1997.

1.2.2 Myogenesis

Muscle forms from the myotome compartment within somites, and myogenesis is mainly regulated by four key Myogenic Regulatory Factors, *Myf*5, *MyoD*, *MyoG*, and *Mrf4* (Scaal and Christ, 2004). *MyoD* and *Myf*5 are muscle specific transcription factors that lie at the top of the transcriptional regulatory network, with disruption of this network resulting in complete failure of skeletal muscle formation (Rudnicki *et al*.,1993). Myogenin (*MyoG)* acts during terminal differentiation of committed myoblasts (Berkes and Tapscott, 2005). *Mrf4* is expressed both in postmitotic maturing cells and undifferentiated proliferating cells, acting as both a determination gene and a differentiation gene (Kassar-Duchossoy *et al*., 2004). Most types of skeletal muscles express a subset of these four genes during development, but

different muscle differentiation pathways require different upstream and downstream signals (see figure 1.7).

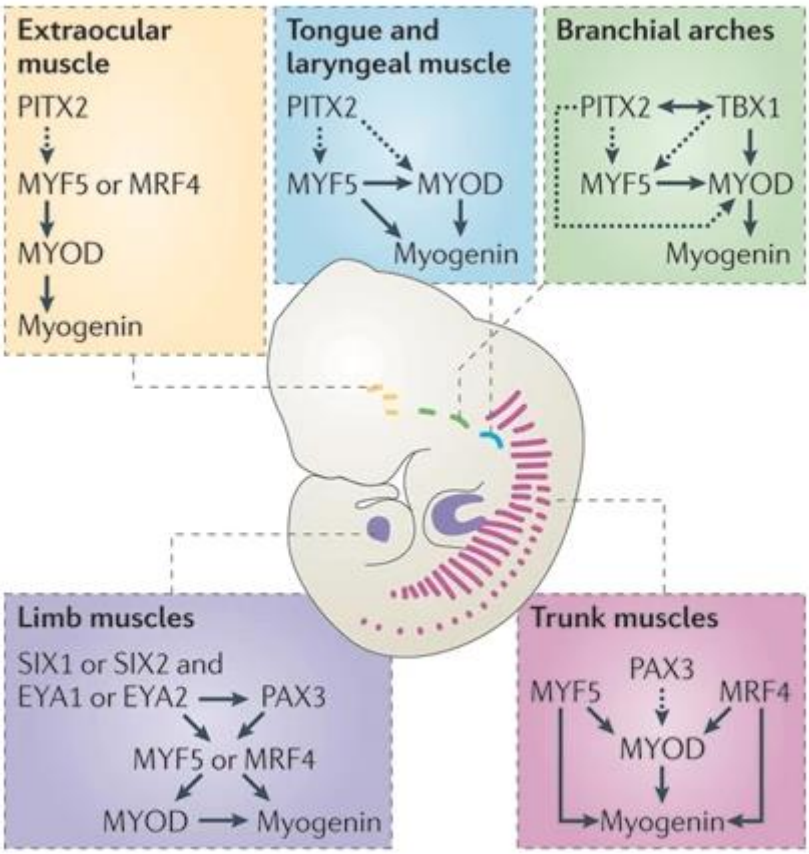

Figure 1.7: A diagram depicting the different genetic pathways for muscle development in differing muscle types. Image from Braun and Gautel, 2011. All skeletal muscle cells express a sub- set of these myogenic regulatory factors (Myf5, MyoD, MyoG and Mrf4).

The *Wnt*, *Shh* and *BMP* signalling pathways all play a role in myogenesis. Mouse knockout mutants for *wnt1* lacked parts of the dermomyotome, and the expression of the early sclerotome transcription factors *pax1* and *Pax9* (Ikeya and Takada, 1998). It has been shown that different *wnt* proteins act to induce expression of different myogenic regulatory factors; *Wnt1* increases *Myf5* expression levels, yet *Wnt7a* or *Wnt6* induced expression of *MyoD* (Tajbakhsh *et al*., 1998). This has been linked with the *wnt* receptors *frizzled*, and the canonical and non-canonical wnt signalling pathways. *Fzd7* is expressed in the hypaxial portion of the somite, where it binds *wnt7a* and signals through the non-canonical pathway. *Fzd1* and *Fzd6* are expressed in the epaxial region of the somite, where they are bound by *Wnt1* and signal through β-catenin dependent pathway (Borello, 2006). Sonic Hedgehog also plays a role in for specification of muscle progenitors. *Shh* or their receptor *Smoothened* knockout mice display a reduction of *Myf5* in the myotome, and improper formation of the sclerotome (Zhang *et al*., 2001). Chicken gain of function assays showed that ectopic expression of *Shh* increases the levels of the sclerotomal marker *Pax1* and inhibited expression of *Pax3* in the dermomyotome, a step that is crucial for commitment of myotomal cells (Borycki *et al*.,1998). *BMP* signalling negatively regulates myogenesis, by inducing expression of *Pax3* which inhibits *Myf5* and *MyoD* induction (Pourquie *et al*., 1995).

The paired box transcription factors *Pax3/7* are both induced by *BMP*, *Wnt* and *Shh* signalling. In mouse epaxial muscles, *Pax3* activates *Myf5* through the expression of *Dmrt2*, which in turn binds to the *Myf5* enhancer region (Sato *et al*., 2010). Ectopic expression of *Pax3* in the paraxial mesoderm showed it can induce *MyoD*, *MyoG* and *Myf5* in the absence of inducing tissues (Maroto *et al*., 1997).

Pax7 is a homologue of *Pax3*, with each gene being able to compensate for the other in knock down experiments (Relaix, 2004), but in experiments with a double knockout, these mutants displayed a severe disruption of muscle development, with few differentiated cells (Relaix *et al*., 2005). *Pax3* regulates key components of the *FGF* signalling pathway, that was shown to be involved in the expansion of the myogenic progenitor cell pool (Lagha *et al*., 2008). All these findings suggest that *Pax3* and *Pax7* are responsible for increasing the size of the muscle precursor cell populations, making them likely to differentiate towards a myogenic fate. In hypaxial muscles, it has been shown that *Pax3* directly induces expression of *Myf5*. Regulation of *Myf5* comes from a 145bp regulatory element 57.5kb upstream of the *Myf5* locus. This regulatory region contains a *Pax3* binding site, and mutation of this site leads loss of *Myf5* expression in mouse embryos (Bajard, 2006).

Sine oculis-related homeobox (*Six*) proteins also play a role in the early stage of myogenesis. *SIX1* and *SIX4* also regulate the transcription of *Myf5* by binding to the same regulatory element (Giordani *et al*., 2007). *SIX1* mouse mutants die at birth, showing muscle hypoplasia in several key muscles, including the limbs (Laclef, 2003). In mice lacking both *SIX1* and *SIX4*, there are no muscle progenitor cells in the limb buds, and no muscle forms (Grifone, 2005). Furthermore, these mice show no expression of *Pax3* in the dermomyotome, suggesting that *SIX* proteins lie upstream of *Pax* proteins (Grifone, 2005). These mutants also display a reduced expression of *MyoD* and *MyoG*. *Six* proteins are clearly important for the early induction of myogenic cells.

Another group of proteins that promote myogenesis from the pool of undifferentiated cells are the myotome enhancer factors (MEFs). One of the better known MEF families, MEF2 proteins are transcriptional activators, regulated by the MAP kinase signalling pathway (Potthoff and Olson, 2007). MEF2 was first identified as a muscle gene expression regulator (Gossett *et al*., 1989), and has mostly been studied in *Drosophila*. It is expressed in early mesoderm, and throughout different muscle lineages, with a main role in myoblast differentiation. Several enhancers control the expression of *Mef2*; expression in the early mesoderm requires an enhancer that is activated by *twist* (Cripps *et al*., 1998), whereas the proteins MAD and Meda act on a separate enhancer to control its expression in somites (Nguyen and Xu, 1998). *Mef2* can also auto regulate itself in the late muscle differentiation stage, by activating a distal autoregulatory enhancer (Cripps *et al*., 2004). In *Drosophila* embryos where *Mef2* is mutated, there is a complete block in myoblast fusion and expression of muscle differentiation markers (Ranganayakulu *et al*., 1995). This set of proteins are critical in the differentiation of muscle and lineage specification.

Micro-RNAs (miRNAs) also play a key role throughout the differentiation of the myotome. Micro-RNAs are transcribed from small non-coding regions of DNA and regulate gene expression post-transcriptionally. Many miRNAs are expressed specifically within muscle tissue, and are involved in myogenesis (Boutz *et al*., 2007). Chromatin immunoprecipitation followed by microarray (ChIP-chip) analysis showed that *MyoG* and *MyoD* bind upstream of miR-1 and miR-133 (Rao *et al*., 2007). *Myf5* is specifically required for miR-1 and miR-206 expression, which was absent in the somites of *Myf5* mutant mice (Sweetman *et al*., 2008). miRNAs tend to be involved in myogenic differentiation and are up regulated after differentiation (Mok *et al*., 2017). It has been suggested that miR-1 accelerates myoblast differentiation by downregulating expression of histone deacetylase 4, a known repressor of muscle differentiation, and knockdown of miR-1 slows myogenic differentiation (Chen *et al*., 2005). Various other miRNAs have been shown to be a crucial aspect of myoblast differentiation, overall adding to the complexity of the muscle development pathway.

There are many key steps in muscle differentiation in the embryo. It is worth studying and understanding these processes, as not only is it important from a developmental point of view, but also for postnatal human muscle disorders and injuries, as a similar genetic programme that is deployed in developing embryos is also used in adults when repairing broken muscle fibres. Mutations in genes that are involved in muscle development can be identified in embryos. Once gene regulatory network that the mutation disrupts is known, which can again be identified by experimentation in embryos, work can begin on finding therapies for individuals who suffer from a muscle disease phenotype.

1.2.3 Chondrogenesis

Chondrogenesis is the process that leads to the formation of chondrocytes, which in turn give rise to cartilage and produce proteoglycans of the extra-cellular matrix. These cells arise from the sclerotomal region within somites, which are marked by the expression of several genes, including *Pax1*, *Pax9* and *Nkx3.1* (Monsoro-Burq, 2005). Sclerotomal cells give rise to many different anatomical structures, including the axial skeleton, ribs, and vertebrae as well as cartilage (Christ, *et al*., 2004). The ventral sclerotome is the origin of vertebral bodies and intervertebral discs (Christ, *et al*., 2004). Populations of cells from the dorsolateral sclerotome migrate to and engulf the neural tube, giving rise to chondrocytes and eventually cartilage (Brent *et al*., 2004). The ribs and vertebrae are formed by a process called endochondral ossification, where the growing cartilage forms a scaffold which is then replaced by bone to form the growing skeleton (Ortega *et al*., 2004).

The genetic programme that controls the differentiation of the sclerotome is complex. Initially, Shh signalling from the floor plate, along with *BMP* and noggin signalling defines the sclerotomal section of the somite (Monsoro-Burq, 2005), marked by expression of *Nkx3.2* (Rainbow *et al*., 2014). *Shh* also drives the expression of *Pax1* and *Pax9*, two key sclerotomal markers (Cairns, 2008). In *Pax1*/*Pax9* double mutant mice, the ventral parts of the vertebra are severely deformed, with the mutants completely lacking the vertebral bodies, intervertebral discs and the proximal part of ribs (Peters *et al*., 1999). Although *Pax1* is critical for the formation and maintenance of the sclerotome, *Pax1* inhibits further differentiation into chondrocytes (Takimoto *et al*., 2013), and is downregulated during chondrogenic differentiation.

Pax1 and *Pax9* act upstream of *Nkx3.2* at the start of sclerotomal formation and activate the expression of chondroblast markers (Rodrigo, 2003). *Nkx3.2* represses the myogenic marker *Pax3* which is expressed in the dermomyotome, maintaining the mesenchymal state of the sclerotome, also blocking transcription of *Sox9* repressors (Zeng, 2002). In *Nkx3.2* mutants, the sclerotome forms, induced by *Pax9*, but does not further differentiate, demonstrating the key role for this gene (Rainbow *et al*., 2014). However, whilst *Nkx3.2* is critical for chondrocyte cell fate specification and proliferation, it is a negative regulator of chondrocyte maturation, and *Nkx3.2* misexpression results in delayed chondrocyte maturation, by repressing *Runx2* (Provot, 2006). In humans, a recessive mutation of *Nkx3.2* results in spondylo-megaepiphyseal-metaphyseal dysplasia (SMMD). SMMD is a rare autosomal recessive skeletal dysplasia, with a phenotype of short stature, with a short and stiff neck and impaired ossification of the vertebral bodies (Simon *et al*., 2012). In the mouse, knockouts of *Nkx3*.2 results in embryonic lethality in mice, preventing further study.

One of the markers induced by *Nkx3.2* and *Pax1/9* is *Sox9*, a pivotal transcription factor expressed throughout developing and adult cartilage. Its importance was highlighted when it was discovered that mutations in *Sox9* lead to the human early lethal syndrome known as Campomelic Dysplasia (Lefebvre and Dvir-Ginzberg, 2016). *Sox9* binds to enhancers that activate various cartilage genes, increasing expression (Lefebvre and Dvir-Ginzberg, 2016). Genes such as *COL2A1 and Tenascin-C* are then upregulated by *Sox9*, leading to cartilage formation.

1.2.4 Tendon development

As the somite matures, the tendon lineage is established within the dorsal sclerotome, in an area termed the syndetome (Brent *et al*., 2003). The syndetome arises at the same time as the myotome emerges from the dermomyotome. The basic helix-loop-helix (bHLH) gene *Scleraxis* is the main genetic marker of early syndetome (Cserjesi *et al*., 1995), and is induced by *Fgf4* signals from the central myotome. The *Fgf4* proteins bind to their receptor *FREK*, which has a restricted expression pattern to the cranial and caudal borders of the myotome (Edom-Vovard *et al*., 2002).

Previous studies have also indicated that the transforming growth factor-β (TGFβ) is required for the early signalling pathways for tendon development and throughout tendon development (Tan *et al*., 2020). TGFβ, is a potent inducer of *Scx*, and disruption of TGFβ signalling resulted in a complete failure

of tendon development (Pryce *et al*., 2009). Overexpression experiments with the members of the TGFβ superfamily resulted in ectopic new tendon formation in rats (Wolfman *et al*., 1997), suggesting that TGFβ proteins are crucial for the development and formation of tendons.

1.3 Genetic Loci and Chromatin Structure

To enable gene expression, the transcription start site (TSS) of a gene must be accessible. The length of DNA in the nucleus is far greater than the size of the nucleus in which it is stored, and for DNA to fit into the nuclear compartment, it is packaged down 10,000-20,000-fold into chromosomes. Chromosomes form pairs, and there are varying numbers of chromosomes present in different species; for example, there are 23 pairs of chromosomes in human, 20 pairs in mouse and 39 pairs in chicken. Chromosomes are tightly packaged to contain the vast amount of genetic material. In eukaryotes, chromosomes are composed of chromatin fibre, which contains nucleosomes. A nucleosome consists of 147 base pairs of DNA that is wrapped around a set of eight histones, termed an octamer, which are further described below. Chromatin fibres are further condensed by proteins into a structure called chromatin, which is present in most cells that contain a nucleus. The folding of the chromatin fibre in the cell nucleus is tightly linked to biological function and gene expression and characterizing the conformational and dynamical properties of chromosomes has become crucial to further understanding how genes are regulated. The packaging of DNA is described below in figure 1.8.

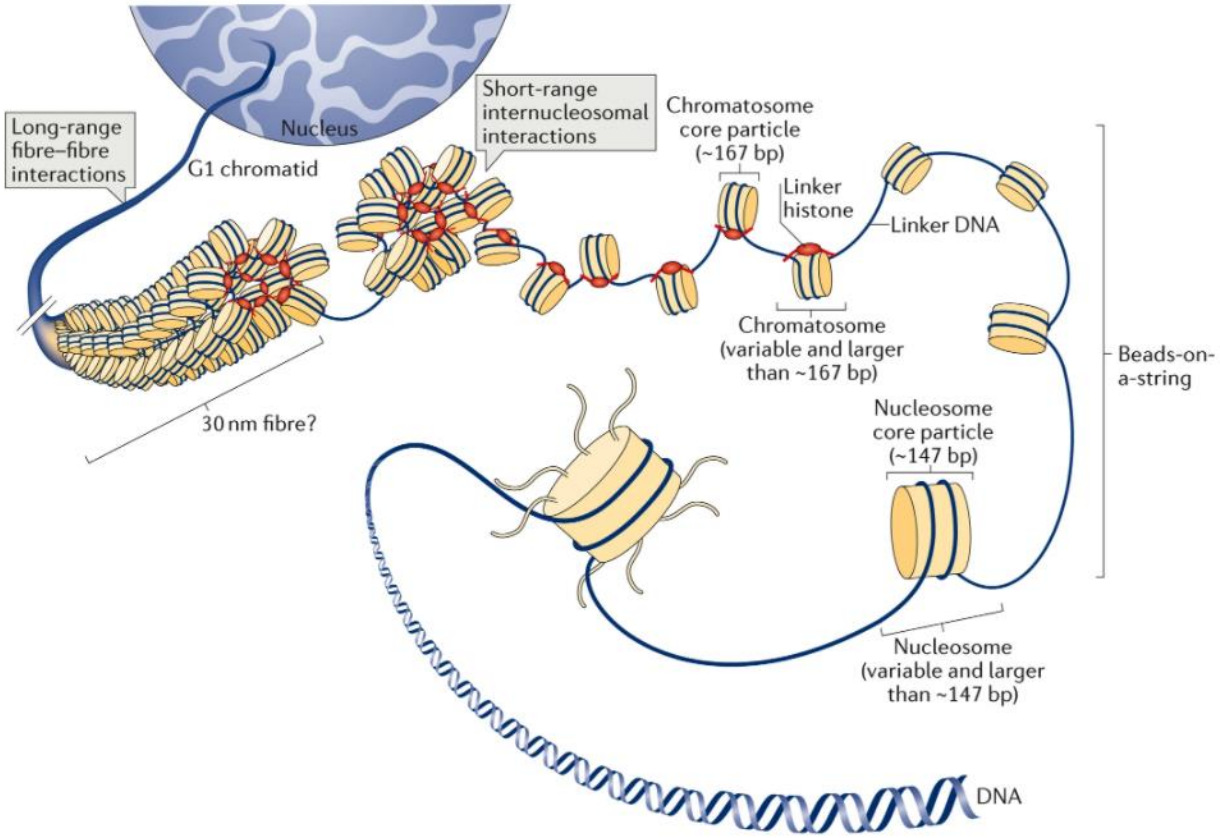

Figure 1.8: The packaging of DNA into a chromosome. The DNA double helix is wrapped around histone octamers, resulting in closed chromatin, and is described as a nucleosome These nucleosomes are further packaged into the chromatin fibre, roughly 30nm in width, which is all packaged into the chromosome. (Fyodorov et al., 2017).

Histones are a family of proteins that associate with DNA in the nucleus and help it to condense into chromatin. Five major families of histones exist: H2A, H2B, H3 and H4, which make up the core histone and H1/H5 which are known as linker histones. The core histones exist as dimers, which then come together as two H2A-H2B dimers and a H3-H4 tetramer to form one octameric nucleosome core. The linker histone H1 binds the nucleosome at entry and exit sites of the DNA, locking it into place, and it

is here where most epigenetic modifications occur, to release the DNA from the nucleosome, and grant the transcription machinery access (this is talked about in more detail in 1.4.3).

Genes are transcribed from their genetic loci, the physical region of a gene on a chromosome. It is important we understand how transcription works to elucidate the mechanisms by which it is controlled. In its most basic form, the transcription machinery first finds an accessible promoter and assembles a protein complex, which recruits RNA-polymerase II (Kornber *et al*., 2007). DNA helicase is also recruited and unwinds double stranded DNA so that RNA polymerase II can use single stranded DNA as template. This produces a single stranded messenger RNA molecule (mRNA), which is then translated by the ribosomes, to produce a string of amino acids. Genetic loci contain introns (non-coding regions of DNA) and exons (coding regions of DNA). Introns are removed by RNA splicing post transcriptionally, meaning they are not present in the final mature mRNA.

Genetic loci are further organised in the genome into topologically associated domains (TADs). TADs are fundamental units of three-dimensional nuclear organisation, which restricts the activity of other genetic loci from acting upon its own (Dixon, Gorkin and Ren, 2016). They are hundreds of kilobases to several million bases in length and are evolutionarily conserved (Dixon *et al*., 2012). They are important for two reasons; they have a self-association property, as cis-regulatory regions and genetic loci can only act upon other genetic loci within their TAD, and an insulating property from neighbouring TADS. Previous studies on various diverse cell types indicated that TADs are fairly invariant, as opposed to other chromosomal features (Dixon *et al*., 2012), meaning they are more likely to be involved with chromosome organisation rather than specific gene activation. When TADs are disrupted, disease can occur as other cis-regulatory elements act outside their restricted domain (Lupiáñez *et al*., 2015, Hnisz *et al*., 2016), and multiple regulatory elements have been shown to affect the activity of multiple loci within a TAD (Symmons *et al*., 2014). One implication for this is that distal regulatory interactions have the potential to be non-specific within a TAD domain. It is important to acknowledge the importance of TAD domains and chromatin structure during the search for cis-regulatory elements.

1.4 Epigenetic control

Epigenetics, first defined by Waddington in 1942, was used to describe the changes in phenotype without changes to the genotype (Waddington.,1942). In the following years, the concept of epigenetics helped explain phenomena such as X inactivation and imprinting. After a multitude of studies produced in the late 1900s, we now understand epigenetics as a mechanism to regulate genetic expression of key genes within the genome, which can be heritable both trans-generationally and by daughter cells. There are several different epigenetic mechanisms by which gene expression is regulated, which are key to the development of cell lineages. These individual mechanisms include DNA methylation, non-coding RNAs and histone modifications.

1.4.1 DNA Methylation

The role of DNA methylation in regulating gene expression was proposed during the 1970s (Holliday and Pugh., 1975). The principal epigenetic tag found throughout differentiated mammalian cells is the covalent attachment of a methyl group to the C5 position of cytosine rings (Bird., 1986). Methylation tends to occur in CpG sites, where a cytosine residue is followed by a guanine nucleotide. The cytosine in CpG dinucleotides is methylated to form 5-methyl cytosines (Bird., 1986). In mammals, seventy to eighty percent of CpG sites are methylated (Jabbari and Bernardi., 2004), indicating methylation plays an important role in regulation of genetic expression.

Methylation of the cytosine residues is often associated with transcriptional repression (Weber *et al*., 2007), especially if these methylated sites are promoters or other regulatory regions. It is thought that DNA methylation can suppress transcription through several mechanisms. Firstly, several studies have demonstrated that methylation can directly block binding of transcription factors (Aoyama *et al*., 2004), or recruit other factors, such as MeCP2, which binds to methyl CpG. MeCP2 then recruits other transcriptional repressors, such as histone deacetylases (HDAC), which promotes chromatin condensation, preventing transcription of genes (Nan *et al*., 1998). DNA methylation is regulated by

the enzymes DNA methyltransferases (DNMTs), which add methyl groups, and allows DNA methylation to remain stable and thus be passed from cell to cell, and trans-generationally (Chen *et al*., 2003). Methylation is also regulated by TET enzymes, which lead to the removal of a methyl group by successive oxidation.

During embryonic development, methylation patterns follow two distinct rounds of epigenetic reprogramming (Zeng and Chen, 2019). The first wave of methylation reprogramming occurs during gametogenesis in the primordial germ cells, where nearly all methylation marks are removed from the genome (Zeng and Chen, 2019). At later stages of germ cell development, an increase in de-novo methylation results in the establishment of sex-specific germ cell methylation patterns, including methylation marks at imprinted loci. After fertilisation, there is an erasure of DNA methylation, but methylation marks are retained at imprinting control regions. In the implantation stage, another wave of methylation reprogramming occurs. Methylation appears at sites near to genes involved with general developmental processes, and a reduction in methylation near to genes with tissue specific functions (Weber *et al*., 2007). Recent studies looking at DNA methylation in promoter regions of DNA suggests that only a few CpG sites within the promoter regions are methylated, inhibiting only a small set of genes (Weber *et al*., 2007). Many of these transcriptionally repressed genes are germline specific, suggesting that methylation plays an important role in the suppression of key genes during development.

1.4.2 Non-Coding RNAs

A non-coding RNA (ncRNA) is a functional RNA molecule that is transcribed from regions within the epigenome but is not translated into a functioning protein (Cech and Steitz, 2014). They are abundant, regulating a broad spectrum of cellular processes. There are several different classes of known ncRNAs, including snoRNAs, scaRNAs, siRNAs snRNAs, lncRNAs and miRNAs, and are highly conserved throughout all forms of life (Cech and Steitz, 2014). ncRNAs are involved in RNA splicing, where they are components of the spliceosome (Kishore, 2006). They also work as a post translational control mechanism through gene silencing or transcribing (Rinn and Chang, 2012), or as a hormone (Knol, *et al.*, 2015). However, one of the main roles of ncRNAs is that they act as key regulators of chromatin structure in eukaryotes. Small RNAs can target gene expression through RNA interference (RNAi) pathways, which mediate histone or methylation events that can prevent transcription (Rinn and Chang, 2012). NcRNAs play a huge role in regulating gene expression, along with a huge variety of other cellular functions, and are a key player in epigenetic regulation.

1.4.3 Histone Modifications

Histone modifications are covalent post translational modifications to histone proteins; these modifications can impact gene expression by altering chromatin structure (Mariño-Ramírez *et al.*, 2005). The main histone modifications include histone acetylation and deacetylation and histone methylation. Histone acetylation is the addition of an acetyl group from acetyl coenzyme A and is controlled by the enzymes histone acetyltransferases (HATs), and histone deacetylases (HDACs) (Roth *et al.*, 2001). Histone acetylation is generally associated with transcription, as the acetyl group neutralises the lysine's positive charge, weakening the interaction with DNA (Bannister and Kouzarides., 2011). Histone methylation mainly occurs on the side chains of arginine and lysine residues, and lysine residues may be mono- di- or tri- methylated (Bannister and Kouzarides., 2011). Methylation of histones generally correlates with a decrease in transcription of genes, as it aids in tightening the histone tail, restricting accesses to DNA, however, methylation of some residues in histones are associated with transcriptional activation (Jambhekar *et al.*, 2019). Histone modifications are a classic form of epigenetic control, by modifying the chromatin accessibility dynamically.

Chromatin remodelling complexes can also configure the chromatin to permit or repress transcription. In mammals, the Switch/Sucrose Non-Fermentable (SWI/SNF) chromatin-remodelling complex can perform this. There have been two distinct SWI/SNF complexes described, each containing a unique domain; the BAF or PBAF domains. These domains contain recognition motifs that can bind to histone tails, activating or repressing them, resulting in open or closed chromatin (Alver *et al*., 2017).

1.5 Cis-Regulatory Elements (CRE)

The majority of the human genome consists of regions that do not code for any proteins. These non-coding regions were described as 'Junk DNA' - DNA regions with no purpose. But, in 1961 a study unveiled the regulatory control mechanism behind the *lac* operon within noncoding regions of DNA (Jacob and Monod, 1961), and since then many more have worked to characterise these CREs (Doane and Elemento, 2017). CREs are typically non-coding regions of DNA which contain binding sites for transcription factors (Wittkopp and Kalay, 2011). Promoters and enhancers are the best understood, but CREs also cover silencers and operators.

It has recently been observed that mutations within CREs are sometimes responsible for a disease phenotype. When screening of several human diseases showed no distinguishable mutations in the coding regions of their DNA, scientists turned their attention to non-coding regions of DNA. With the recent advances in sequencing, it was possible to sequence the whole genome, and this suggested that mutations within CREs were likely responsible for the disease phenotype. For example, mutations in an enhancer regulating expression of the SOX10 transcription factor (TF) were shown to contribute towards Hirschsprung disease (Lecerf *et al*., 2014), and recurrent mutations in the promoter regions of genes such as PLEKHS1 was found to be responsible for tumour progression in cancers (Weinhold *et al*., 2014). Many more diseases such as diabetes and cohesinopathies were found to contain a mutated variant within their associated CRE (Friedensohn and Sawarkar, 2014, Lee and Young, 2013). However, it is unknown how some of these CREs function, and identifying them is difficult. Therefore, it is important to focus efforts on identifying and characterising these CREs, to gain better understanding of gene regulatory networks and the control of gene expression.

1.5.1 Promoters

Promoter regions are sequences of DNA that recruit RNA polymerase II to the transcription start site (TSS) of an encoded gene (Kadonaga, 2011). They are typically located directly upstream of the gene they are regulating and contain a conserved promoter sequence called the TATA box, which allows for binding of transcription factors to initiate formation of RNA polymerase transcription complex, which promotes transcription (Butler, 2002). Eukaryotic promoters include a core promoter, which in turn contains the TSS, RNA polymerase binding sites, and a TATA box (or other general TF binding sites). In humans, 70% of all discovered promoters contain CpG islands, which are often methylated to prevent expression of a gene (Deaton and Bird, 2011), a key example of epigenetic regulation. Figure 1.9 below gives an overview of promoter properties.

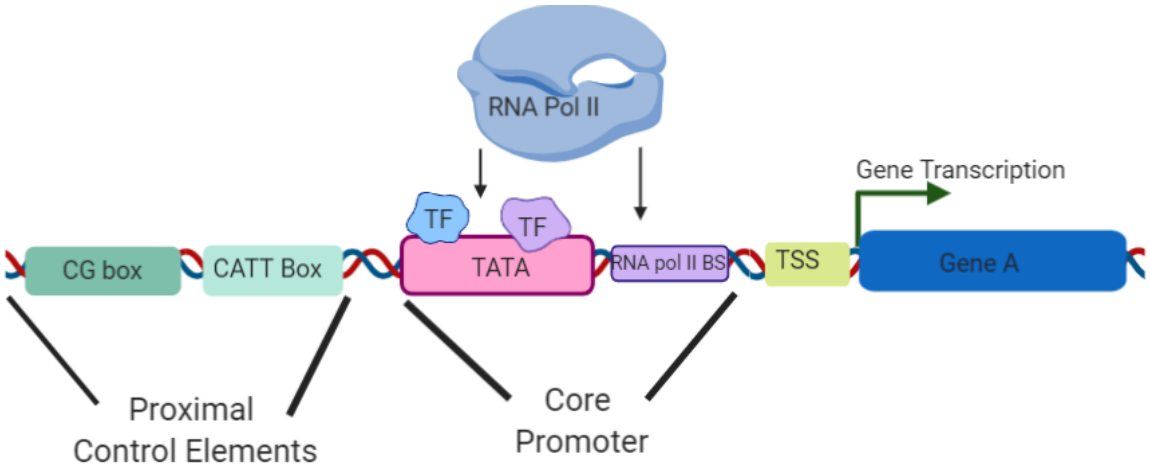

Figure 1.9: A basic diagram depicting the main regions of a eukaryotic promoter region. The core element contains a TATA box and an RNA Polymerase II binding site, which recruits and binds RNA pol II to the DNA near the transcription start site

of the gene. This allows for transcription of the gene to initiate. Proximal promoter region is also displayed, which shows two other motif boxes for transcription factor binding.

1.5.2 Silencers and Insulators

A silencer is a sequence of DNA that can prevent gene expression throughout differentiation and the cell cycle. Silencers can be located throughout the genome, however, are usually upstream of the gene they are acting upon, between 20bp to 2000bp, and certain silencers have been found in the 3' UTR of mRNA. Currently, there are two types of recognised silencers: classical silencers and the non-classical negative regulatory element. Classical silencers tend to work by interrupting the formation of the general transcription factor assembly, by repressing the TATA box of the core promoter region, preventing it from recruiting and binding to RNA polymerase II (Maston *et al*., 2006, Ogbourne and Antalis, 1998). As well as general transcription factor assembly, silencers are capable of cytoplasmic retention of TFs, altering chromatin structure, and can interfere with the 3' upstream UTR signal recognition (Ogbourne and Antalis, 1998). They normally bind to proteins termed repressors, including polycomb complexes PRC1 and PRC2, which remodel chromatin to prevent access of transcription machinery to the desired loci (Kanhere *et al*., 2010). Silencers that are located a long way from the promoter region it is repression will target that promoter through looping of the DNA, where the silencer element is brought into close proximity with the promoter region (Kolovos *et al*., 2012).

One example of a characterised classical silencer regulates the human thyrotropin gene-β gene. It is found in the hTSHβ promoter region, where it acts upon an AT-rich octamere-1 (OCT-1) binding domain. It is believed that the silencer recruits oct-1, which can silence transcription by interfering with TFIIH, preventing helicase activity (Kim *et al.*, 1996). NREs physically inhibit the interaction of TFs with their binding sites or interfere with specific signals that control transcription events (Ogbourne and Antalis, 1998). Silencers are an important part of genetic regulation and will also be depicted as an open peak during ATAC-sequencing, so it is important to understand them for this experimental pipeline.

1.5.3 Enhancers

In 1981, distal cell specific regions of open chromatin were identified far from any expressed genes. It was shown that these regions 'enhanced' gene expression in *cis*, and the term enhancer was first used (Banerji *et al*., 1981). It is currently believed that the human genome contains roughly 500,000 enhancer regions, which are used to regulate gene expression during development and homeostasis (Gasperini *et al*., 2020). Enhancers are defined as short sequences of DNA that contain transcription factor binding motifs that increase the likelihood of transcription of one or more genes through a *cis* regulatory mechanism (Ong and Corces, 2011). However, despite the obvious importance of enhancers to both disease biology and fundamental biology, far more work is needed to fully characterise and understand the functional importance of these regions.

There are a few distinguishable characteristics of enhancer regions. They are usually linked with certain properties of chromatin, and active enhancers are typically bound by transcription factors. Regions that flank enhancers are marked by histone modifications which allow for open chromatin complexes such as H3K4me1 and H3K27 acetylation, with a depletion of H3K4me3 (Calo and Wysocka, 2013, Jin *et al.*, 2009). The TSS of enhancers are depleted of nucleosomes, and are therefore sensitive to DNase I digestion, and many functional enhancers contain dynamic nucleosomes (Thurman *et al*., 2012). Other markers, such as the proteins Cyclic AMP-responsive element-binding (CREB) and P300 are also characteristic of some enhancer regions. (Mayuranathan *et al.*,2015) and *in-vivo* mapping of P300 binding sites in mouse accurately identified novel enhancers (Reviewed in Ong and Corces, 2011). Most enhancer regions are also bound by chromodomain helicase DNA binding protein 7, cohesin, mediator subunits and CCCTC-binding factor (CTCF) (Zentner *et al*., 2011).

As enhancers work spatio-temporally to regulate gene expression, they are found in different states depending on chromatin accessibility. Developmental enhancers undergo progressive changes that

depend on TF activity and chromatin remodelling (Peng and Zhang., 2018), and thus there are three characterised states of enhancers. Active enhancers that interact with their respective promoter have chromatin modifications that allow transcription factors to bind to the enhancer region. Prior to activation, enhancers can exist in two states; primed or poised. Many TFs can bind to enhancers and trigger nucleosome repositioning to increase accessibility to other factors (Zaret and Carroll, 2011). These TFs are termed pioneering factors and can bind to DNA that is inaccessible to other factors, and recruit chromatin remodelling complexes (Biddie *et al*., 2011). One of the most studied pioneering factors is FOXOA1, which induces chromatin remodelling (Lupien *et al*., 2008). The binding of a pioneer factor is not sufficient to form an active enhancer, and so this state is termed primed. It is believed that primed enhancers arise as a result of developmental progression, where pioneering factors that are produced by one developmental genetic process can then act on an enhancer to regulate gene expression of a gene that is required later in development. Poised enhancers have not been extensively studied but have been identified in mouse and human embryonic stem cells and are still able to interact with their corresponding promoter, in preparation for activation (Creyghton *et al*., 2010). Figure 1.10 below shows these three states of enhancers, and the chromatin modifications associated with these.

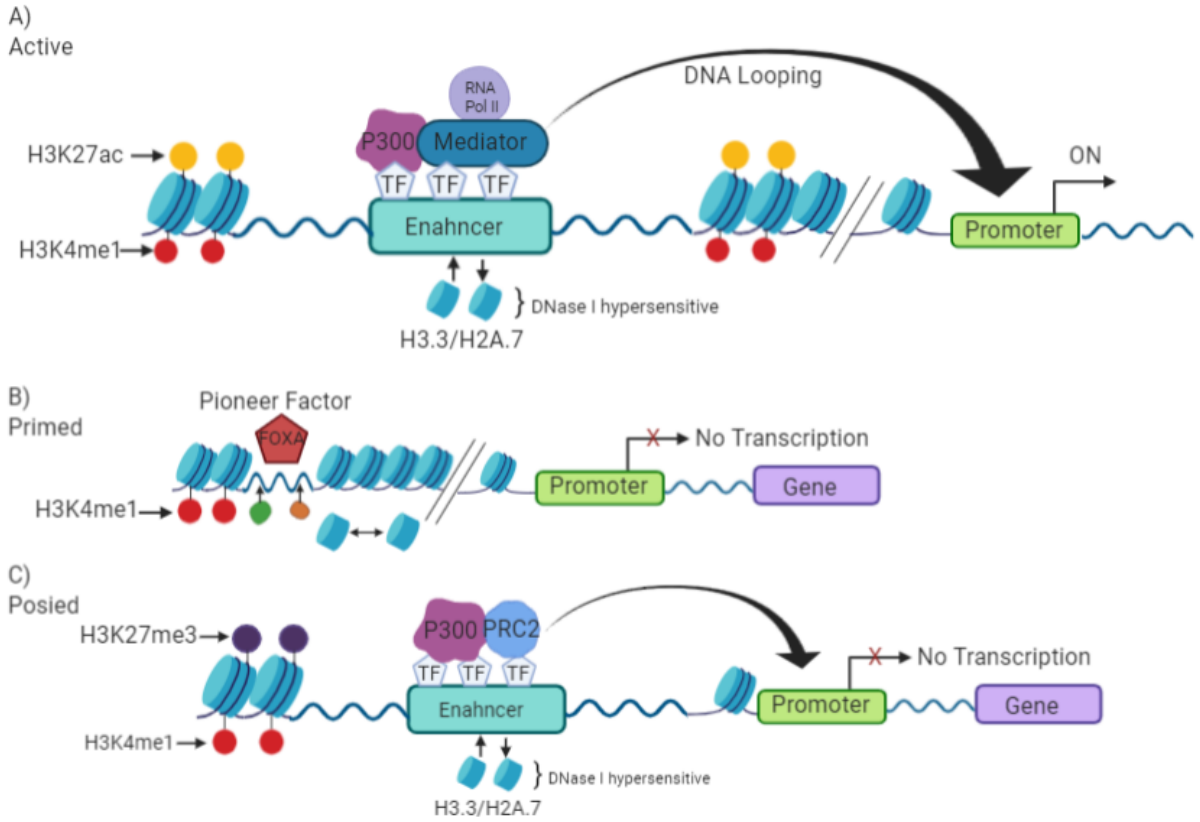

Figure 1.10: Diagram explaining the different chromatin features associated with active poised and primed enhancers. A) Active; The histones flanking the enhancer regions are marked with the modifications H3K27ac and H3K4me1. There is an incorporation of hypermobile nucleosomes containing the H3.3/H2A.7 modifications. B) Primed; this state is characterised by the presence of H3Kme1, with no nucleosome incorporation. This state does not interact with the promoter but binds the pioneer factor FOXA. C) Poised; this state is found in human and mouse embryonic stem cells. These enhancers are associated with PCR2 and interact with their promoter.

It was initially thought that enhancers worked by direct interactions with RNA polymerase II, yet now it is believed that enhancers work by an indirect interaction (Kolovos *et al*., 2012). To increase the likelihood of transcription of a gene, the enhancer must interact with its respective promoter, which could be located many base pairs away. It has been demonstrated in the β-globin locus that DNA looping brings the enhancer and the promoter regions into close proximity by 'looping out' intervening DNA (Krivega and Dean, 2012). Experiments including the chromosome conformation capture 3C (and later, 4C and 5C) confirmed this theory (de Wit and de Laat, 2012). The mechanism of DNA looping is

displayed in figure 1.11. It is currently thought that DNA binding TFs bound at both the promoter and the enhancer recruit 'looping factors' that help to form the loop (Krivega and Dean, 2012).

Genomes also contain insulators that regulate enhancers. Insulators are protein-DNA complexes that prevent enhancers from activating a gene, however, they are also involved in chromatin looping. In vertebrates, the only known insulator protein is CTCF, which recruits a protein called cohesin (Merkenschlager and Odom, 2013). Cohesin belongs to the family of SMC (structural maintenance of chromosomes) protein complexes, which are ring shaped ATPases that encircle DNA (Kagey *et al*., 2010). Recent studies have demonstrated the binding sites that regulate cohesion during genetic regulation. Normally, cohesin binds to CTCF, however, numerous and often weaker cohesion sites map to active enhancers and promoters, where it colocalizes with its loading factor Nipbl, mediator components and tissue specific transcription factors (Schmidt *et al*., 2010). The first time the cohesin-CTCF long range interactions was demonstrated to be of significance was in the mouse *lfng* locus. A CTCF binding site 60kb upstream of the *lfng* coding region interacts with two other CTCF binding sites, one in the first intron of the gene, and one 100kb downstream of the gene. These long-range interactions are required for *lfng* expression, which requires both CTFC and cohesin (Hadjur *et al*., 2009, Sekimata *et al*., 2009). Along with this, cohesion depletion was shown to disrupt promoter-enhancer interactions in embryonic stem cells (Kagey *et al*., 2010). The mechanism of looping of DNA requires insulator regions and cohesin, and this mechanism is crucial to our understanding of enhancer function, bringing the enhancer region and its associated transcription factors into close contact with promoter regions.

The transcription factors and complexes that may mediate the enhancer-promoter looping have only been identified in a few examples. In the well characterised β-globin locus, a reduction of the GATA-1, EKLF and Ldb1 transcription factors resulted in failed β-globin activation or looping between the gene and the locus control region (Vakoc *et al*., 2005). Other general transcription factors that are required in specific cases for enhancer-promoter communication include OCA-B, TFII-I, and STAB1, which when reduced leads to a compromised loop formation and transcription (Ren *et al*., 2011, Gong *et al*., 2011, Cai *et al*., 2006). There is a clear lack of understanding in how enhancer- promoter interactions are formed through the looping mechanism, and more work is needed to fully understand the repertoire of proteins and how they function to understand this phenomenon.

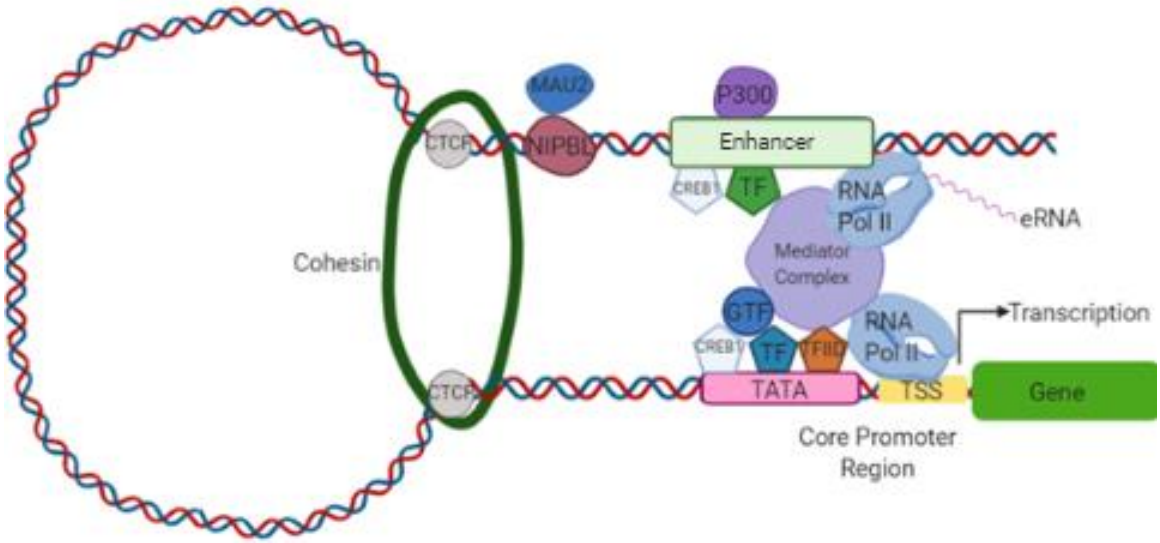

Figure 1.11: A simple diagram demonstrating the looping of DNA. Cohesin binds to CTFC factors on the DNA forming a ring and looping out the intervening DNA between the enhancer and the promoter region. Own image.

Enhancers can regulate the spatial expression of genes. It is known that more than one enhancer can mediate the expression of one gene, which gives rise to complex patterns of gene expression (Perry, Boettiger and Levine, 2011). This can result from the additive action of different enhancers with cell

type or tissue specific activities and gene expression (Barlow, 2011). There are 'primary' and 'secondary' enhancer regions, with the secondary enhancer often being described as a shadow enhancer (Barolo, 2011). It is currently thought that the presence of two enhancers with overlapping activity can provide a robustness to gene expression, establishing precise boundaries of gene expression (Frankel *et al*., 2010). In mouse embryos, *Shh* transcription is controlled by two enhancers. When activated by the two distinct enhancers, expression is restricted and robust in the floor plate. However, transcription in the limb bud is controlled by only one enhancer and gives a more sporadic expression pattern (Amano *et al*., 2009).

The identification of enhancers has proved challenging in the past for several reasons. Firstly, enhancers are scattered throughout the non-coding regions of the genome, resulting in a large search area (Gasperini *et al*.,2020). Whilst it is known that enhancers work in *cis*, they can be located many base pairs away from the promotor they are working on, bypassing their closest promoter (Spilianakis *et al*., 2005). Secondly, and whilst some enhancers have chemical properties that have been annotated (discussed above), there is no underlying feature that can identify an enhancer sequence based on DNA sequence alone. Along with this, DNA sequence conservation in these regions is often poor, as only very specific TF binding sites are required. Thirdly, some enhancers have been shown to regulate activity of two or more promoters, further adding complexity when attempting to identify and functionally characterise enhancers (Karnuta and Scacheri, 2018). Finally, the activity of enhancers is usually spatially and temporally restricted, meaning they are difficult to discover if the method used does not capture the enhancer activity in the correct tissue or at the correct developmental time point (Karnuta and Scacheri, 2018).

Despite these difficulties, many attempts have been made at identifying enhancer elements. Several different technologies have been developed to characterise enhancer regions, each with varying levels of success. Firstly, the DNA sequence of a candidate enhancer region can be studied, using sequence conservation and candidate regulatory elements for TF binding, for example, scanning the genome for regions with p300 binding sites. However, not all enhancers are conserved, and not all motifs for TF binding are known or perfectly annotated. Furthermore, the presence of a TF motif does not indicate functional binding of the TF. Therefore, predicting candidate enhancer regions based purely on sequence alone is limited.

As enhancers rely on binding of TFs and chromatin accessibility to become active, genome wide methods that determine TF binding sites or chromatin modifications are often used. These methods include FAIRE-seq, ChIP-seq, DNA adenine methyltransferase identification (DamID), DNAse-seq, and STARR-seq. These methods are further described below.

Chromatin immunoprecipitation and sequencing (ChIP-seq) for transcription factors, mediators, and cofactors such as P300 (Visel *et al*., 2009), and histone modifications such as H3K4me1 and H3K27ac (Creyghton *et al*., 2010, Cheung *et al*., 2020) is one of the main methods used to attempt to identify enhancers. The ChIP-seq approach is dependent on cross-linking of proteins to specific DNA elements, then the addition of an antibody specific for the protein-DNA complexes, followed by high throughput sequencing of recovered DNA fragments (Creyghton *et al*.,2010). Again, the downside of this is that the TF must be bound to the DNA sequence, or the histone modification must be present, meaning the enhancer region must be active. However, while ChIP-seq successfully recovers the most enhancer regions, some regions sequenced are not enhancer regions. This is because TF can bind to other regions of DNA, and as there is not one TF that binds to only enhancers, ChIP-Seq is captures some other protein-DNA interactions that are not enhancers. The use of the cofactor P300, for example, has been highly successful in identifying enhancer regions, as 58-82% of the P300 binding sites have been associated with functional enhancers (Blow *et al*., 2010).Other methods that work similarly to this include ChIP-exo, which allows the identification of TF binding sites at higher resolution due to the addition of exonuclease digestion (Rhee and Pugh, 2011), and DamID, which can be used if antibodies for the protein of interest aren't available as an alternative to ChIP-seq (Steensel and Henikoff, 2000).

Predictions using genome accessibility have also been used. In general, the regulatory elements selectively localise in the accessible chromatin (Thurman *et al*., 2012). This is because accessible chromatin is required for transcription factor binding, whereas condensed chromatin restrict binding of TFs and transcriptional regulators to the promoter and enhancer regions. There are several techniques that can be used to digest nuclear DNA. In the past, experiments using DNase I (Chen, *et al*., 2018), micrococcal nuclease (MNase) or MPE-FE (Ishii, *et al.*, 2015) coupled with sequencing would identify regions of accessible chromatin. An alternative to DNase-seq; formaldehyde-assisted identification of regulatory elements (FAIRE-seq), determines open regions of chromatin by depleting histone-bound closed DNA after crosslinking with formaldehyde (Simon *et al*., 2012). Although active enhancers are found in regions of accessible chromatin, not all accessible regions correspond to an enhancer; for example, the TSSs within a promoter are nearly always accessible.

Assay for transposase accessible chromatin (ATAC) sequencing is a relatively new method that also works to identify open chromatin regions but is faster and more sensitive than DNase-seq and other predecessors. ATAC-seq uses the transposase 5 (Tn5) enzyme found in bacteria which recognises open regions of chromatin, where it cleaves and tags DNA with sequencing adapters. These tagged regions of DNA are then purified, PCR amplified and sequenced using next generation sequencing (NGS) (Buenrostro *et al*., 2014). Sequencing reads can then be used to map regions of TF binding sites and nucleosome positions, and the number of reads for a particular region correlates with how open the chromatin is. The hyperactivity of the Tn5 transposase makes the ATAC-sequencing protocol simple and time efficient, requiring only 500-50,000 cells (Buenrostro *et al*., 2014). Figure 1.12 below depicts the process of ATAC-sequencing. More recently, the new method single-cell-ATAC-seq (scATAC-seq) has been described, which relies on using fluorescence-activated cell sorting (FACS) and nono-well-based approaches (Buenrostro *et al*., 2015, Mezger *et al*., 2018). This is useful in studying heterogenous cell populations at a single cell resolution and could be extremely useful in depicting the regions of open chromatin in developing embryos.

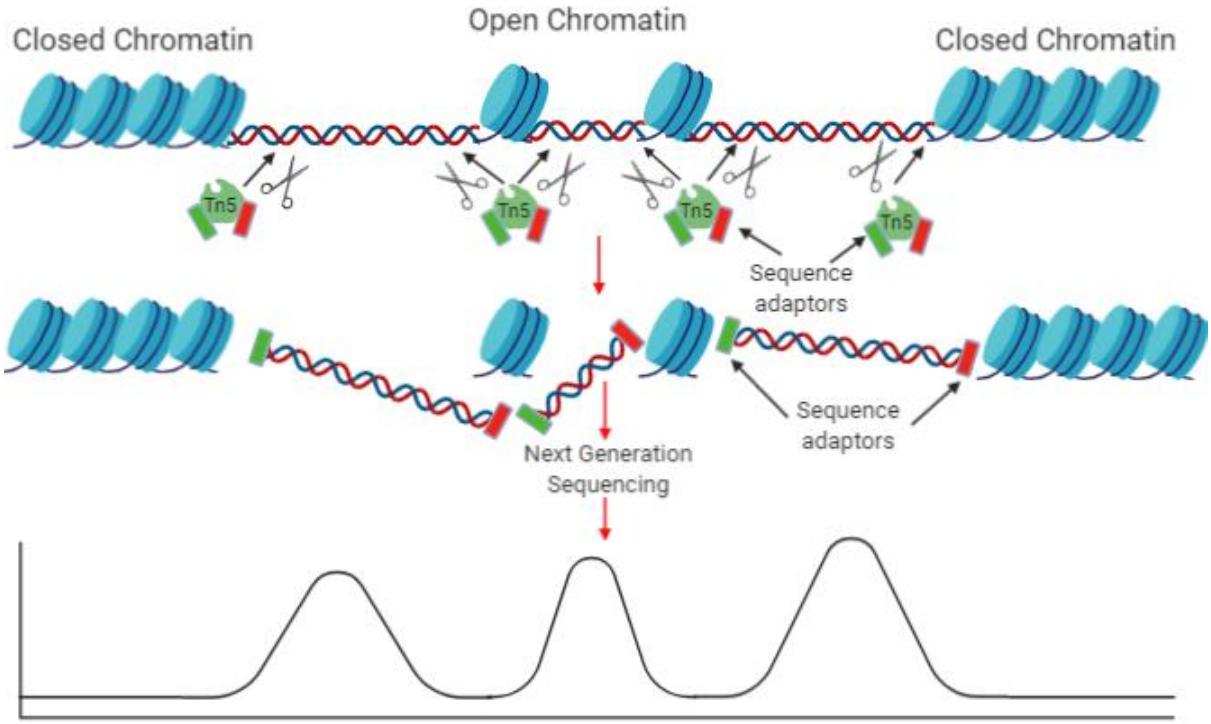

Figure 1.12: Diagram depicting the process of Assay for Transposase Accessible Chromatin. The Tn5 transposase enzyme binds to regions of open chromatin and cuts, adding sequence adaptors to the end. Next Generation Sequencing is used to read these segments, which are then mapped to the genome. Open regions are visualised as a peak.

Expression quantitative trait locus (eQTL) can be used to characterise distally located putative regulatory elements. It works by testing for correlation between genotypes (tested by genome sequencing) and the location of genes *in cis* (measured by bulk RNA-seq). Variants that are associated with gene expression differences are described as eQTLs. Variants residing within distally located putative enhancers can be linked to their target genes and validated in *vivo*. The eQTL framework is a powerful and novel tool in identifying and characterising enhancers (van der Wijst *et al*., 2018, Strober *et al*., 2019).

Another way to identify an enhancer is based on their looping method of interacting with their respective promoter region. By using 3C (Chromatin Conformation Capture) methods, high resolution 3D conformational maps are produced. These maps are produced by the cross-linking of cells by formaldehyde, followed by chromatin isolation and digestion. The digestion occurs through the use of a restriction enzyme that recognises and cuts regions of DNA which, mediated by a protein complex, is interacting with another region of DNA. These fragments are ligated, and the crosslinking procedure reversed. RT-PCR is then used to determine the quantity of ligation product. This directly correlates with the interaction frequency between the two ligated regions (Dekker, 2006). The resulting datasets have led to the identification of TADs, and enhancer-promoter loops. This 3C method can be coupled with biochemical assays such as ChIA-PET and DNase Hi-C to help identify regions with functional interactions. Other variations of 3C have been produced, including Circularize Chromosome Conformation Capture (4C) and Carbon Copy Chromosome Conformation Capture (5C), which can be used to gain larger and clearer conformation maps. Figure 1.13 depicts all the different methods that can be used to identify regulatory elements.

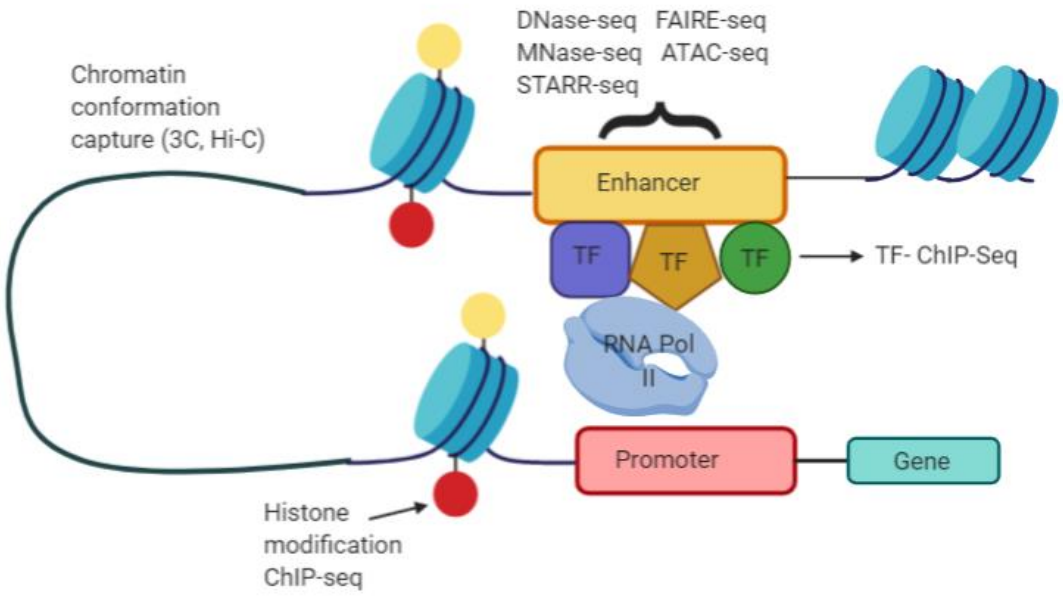

Figure 1.13: A simple diagram showing the different methods that can be used to attempt to identify enhancer regions.

There are several projects that all aim to map the functional elements within the human genome, using the techniques described above. The Encyclopaedia of DNA Elements (ENCODE) has been running since 2007 which combines a large number of sequence-based studies to discover potential elements (The ENCODE Project Consortium., 2012). Approaches used include RNA-seq, CAGE-seq (which maps TSS of transcribed genes), ChIP-seq, FAIRE-seq, DNase-seq and an assay to find methylation sites (RBBS assay). Other programmes that do similar things include FANTOM5, VISTA ENhancer Browser, Enhancer Atlas and HEDD (Peng and Zhang, 2018). However, only 26% of the candidate enhancers predicted by the data generated in ENCODE and only 50% in VISTA Enhancer Browser

displayed functional enhancer activity with reporter assays (The ENCODE Project Consortium., 2012, Kwasnieski, *et al*., 2014). Accurately validating and characterising enhancer elements remains a challenge, and clearly more work needs to be done to help identify key markers of enhancers and their functional significance.

1.5.4 Characterisation of Enhancers

There are several different ways to test enhancer activity and function. One of the most common ways is imaged based enhancer testing in developing embryos, where the candidate enhancer DNA is placed upstream of a minimal promoter and a reporter gene. The abundance and localisation of the reporter transcript displays the enhancer activity. The reporter gene can include an enzymatic reporter (luciferase or β-galactosidase), fluorescence, such as GFP or specific antibodies (Visel *et al*., 2007, 2009). However, for this method to work, the generation of transgenic animals must occur, making it unsuitable for large scale enhancer screening. To combat this, various methods have been developed using plasmid-based systems to assess enhancer activities. Again, each candidate enhancer region is placed upstream of a minimal promoter and a reporter gene. These are then electroporated into embryos, allowing cells to take up the reporter plasmid. If the candidate enhancer construct is expressed, the reporter gene will display spatially and temporally throughout the embryo (Rada-Iglesias *et al*., 2010).

Massively parallel reporter assays (MPRAs) can test the functionality of thousands of putative enhancer regions in one experiment. A library of candidate enhancers is cloned into a reporter vector, with a minimal promoter. Each reporter gene transcript includes a DNA barcode that is associated with a particular enhancer. The relative abundance of each RNA barcode is used to quantify the activity of its cognate candidate enhancers. MPRAs can test large numbers of putative enhancers using relatively simple techniques). STARR-seq (Self-Transcribing Active Regulatory Region Sequencing) is a technique whereby candidate sequences are cloned downstream of the core promoter into 3′UTR of a reporter gene's, making the candidate sequence part of the reporter genes' transcription unit. Candidates that are active enhancers will lead result in the transcription of reporter mRNAs that are the candidates' sequences. This straightforward technique allows for assessment of millions of candidate enhancer regions in parallel by quantifying the reporter mRNAs by deep sequencing (Muerdter, Boryń and Arnold, 2015).

CRISPR screening for enhancers is a relatively new technique, due to improvements in the CRISPR methodology. CRISPR screens work with the aim of characterising massive numbers of enhancers, by delivering a library of enhancer targeting guide RNAs (gRNAs) to cells. These gRNAs will disrupt their target putative enhancer region by creating a small cut, which through inaccurate repairing of the DNA, will result in the formation of an indel, which may disrupt the regulatory element's function. The phenotypic result from these screens informs researchers which gRNAs affect gene expressions (Diao *et al*., 2016). This method has the potential to link candidate enhancer regions to their target genes.

1.6 Genes Involved in Somite Differentiation

Previous work in the lab has resulted in the generation of both RNA-sequencing and ATAC-seq datasets. They were constructed by removing somites of different levels of differentiation (Pre-somitic mesoderm, epithelial somite, maturing somite and differentiated somite) from a HH14 chick embryo and performing both RNA and ATAC-seq on these dissected regions. Genes were chosen from the RNA-sequencing that showed similar expression patterns to *MyoD* for muscle markers. For tendon and cartilage markers, genes that were found from literature to be of interest were reviewed in the RNA-sequencing dataset for high levels of expression. With each of these genes, the ATAC-sequencing dataset was analysed to identify their respective candidate enhancer regions.

1.6.1 FOXO1

Forkhead box protein O1 (Foxo1) is a protein encoded by the Foxo1 gene found on chromosome 13 in the human genome and chromosome 1 in the chicken genome (Sanchez *et al*., 2013). The FoxO

family of proteins are transcription factors, characterised by a distinct fork head domain. They are key downstream targets of the P13-Kinase/PKB signalling pathway, and their transcriptional activity is controlled by phosphorylation (Sanchez *et al*., 2013). Phosphorylated FoxO proteins are confined to the cytoplasm and are thus rendered incapable of regulating transcription in the nucleus. Foxo1 has a wide variety of roles, perhaps best known for its regulation of gluconeogenesis and fate determination of a preadipocyte to commit to adipogenesis (Nakae *et al*., 2003). Due to this, it has been identified as a potential genetic control for type II diabetes.

Foxo1 also has a role in regulation of muscle differentiation and fibre type specification, with a specific role in myoblast fusion and differentiation (Yamashita *et al*., 2016). It is believed that myogenic differentiation requires Foxo1 inhibition. Foxo1 gain of function mutants in C2C12 cells demonstrated inhibition of myoblast differentiation and blocked myotube fusion, whereas transcriptionally inactive Foxo1 cell lines displayed an increase in the expression of muscle differentiation markers, including myogenin and MyHC (Hribal *et al*., 2003). Transgenic mice overexpressing Foxo1 display less skeletal muscle mass, with paler muscles and a downregulation of genes related to structural proteins of muscles (Kamei *et al*., 2004). Whilst previous research has demonstrated the expression pattern of Foxo1 gene in other model organisms (Zheng *et al*., 2020, Baek *et al*., 2018, Villarejo-Balcells *et al*., 2011), there was no expression data for Foxo1 in chicken embryos. Although Foxo1 suppresses early myoblast differentiation, Foxo1 is required for myoblast differentiation into myotubes. This mechanism is currently poorly understood, and it would be interesting to further characterise this protein and its biological regulation.

1.6.2 UNCX

UNCX is a paired homeobox gene, which codes for a transcription factor that is involved in somitogenesis and neurogenesis (Mansouri *et al*., 1997). *UNCX* gene in *gallus gallus* is homologous to *UNCX4.1* in mammals, such as humans and mice. *UNCX* plays many roles during the development of the central nervous system (Sammeta *et al*., 2010), with high levels of expression in the olfactory sensory neurons, and in the neural progenitor cells of the dorsal neural tube, which gives rise to the spinal cord and the telencephalon (Neidhardt *et al*., 1997). However, *UNCX* is best known for its role during early somitogenesis. *UNCX* is normally confined to the caudal half of the somite and sclerotome, suggesting a role for initial rostral-caudal somite patterning (Evrard *et al*.,1998), specifically to maintain the condensation of the caudal half sclerotome (Mansouri *et al*., 2000). Loss of function experiments in mice show a disruption in the establishment of antero-posterior somite polarity, and display severe malformations of the axial skeleton, especially in the ribs (Leitges *et al*., 2000, Mansouri *et al*., 2000). In knockout mouse embryos,expression of *Pax9* is lost in somites after the first three somites form, suggesting that *UNCX* maintains expression of *Pax9* in the lateral sclerotome.

Repression of *UNCX* expression in the anterior somite is controlled by a regulatory network involving the transcription factor Mesp2, Ripply and MEOX1 (Takahashi *et al*., 2013), and is negatively regulated by *Tbx18* (Nittoli *et al*., 2019). Understanding the regulatory mechanisms that underlie expression of *UNCX* would give insight into the mechanisms of early somite formation and rostral-caudal patterning.

1.6.3 KLHL31

Klhl31 belongs to the family of Kelch proteins, a functionally diverse and highly conserved set of proteins, all characterised by the presence of a Kelch-repeat domain and the BTB (Bric-a-brac, Tramtack, Broad-complex) domain (Adams *et al*., 2000). Members of this family are involved in numerous cellular processes, including protein degradation, gene expression and cell migration (Shibata *et al*., 2013., Lührig *et al*., 2013). The kelch protein was first discovered in *Drosophila melanogaster*, where females carrying a mutation in the kelch gene became sterile, due to issues with cytoplasm transport throughout oogenesis (Schupback and Wieschaus., 1991). Since the discovery of Kelch, a total of 41 kelch-like proteins were identified in humans, where they are identified with numerous roles in skeletal muscle development and disease (Gupta and Beggs, 2014).

Klhl31 was first discovered in zebrafish, where there were high expression levels in skeletal and cardiac tissues (Wu and Gong, 2004), with further work describing a similar pattern of expression in humans. Work previously conducted in the Münsterberg lab showed expression of Klhl31 in the mesoderm of the anterior intestinal portal, starting at HH8, and in cardiac progenitors from HH9 (See figure 1.14). Klhl31 was then expressed in the myotome and continued strong expression patterns in all skeletal muscle tissues (Abou-Elhamd *et al*., 2009). This paper also described its regulatory pathway; rescue experiments in the myotome showed Klhl31 is downstream of Wnt-1 and Myf5. The expression of Klhl31 is initiated by the transcription factor *MEF2*, just after MyoD in developing skeletal muscles, implying the involvement of Klhl31 in early myogenic commitment and differentiation (Liu *et al*., 2014). Deletion of Klhl31 in mice results in sarcomeric degeneration and stunted postnatal skeletal muscle growth, which highlight a role of Klhl31 in the maintenance of muscle structure (Papizan *et al*., 2017). Other studies suggest that the Klhl31 protein acts as a transcriptional repressor in the MAPK/JNK signalling pathway, due to its BTB domain repressing GAL4 DNA binding domains of both the TPA-response element (TRE) and the serum response element (SRE) (Hu *et al*., 2010).

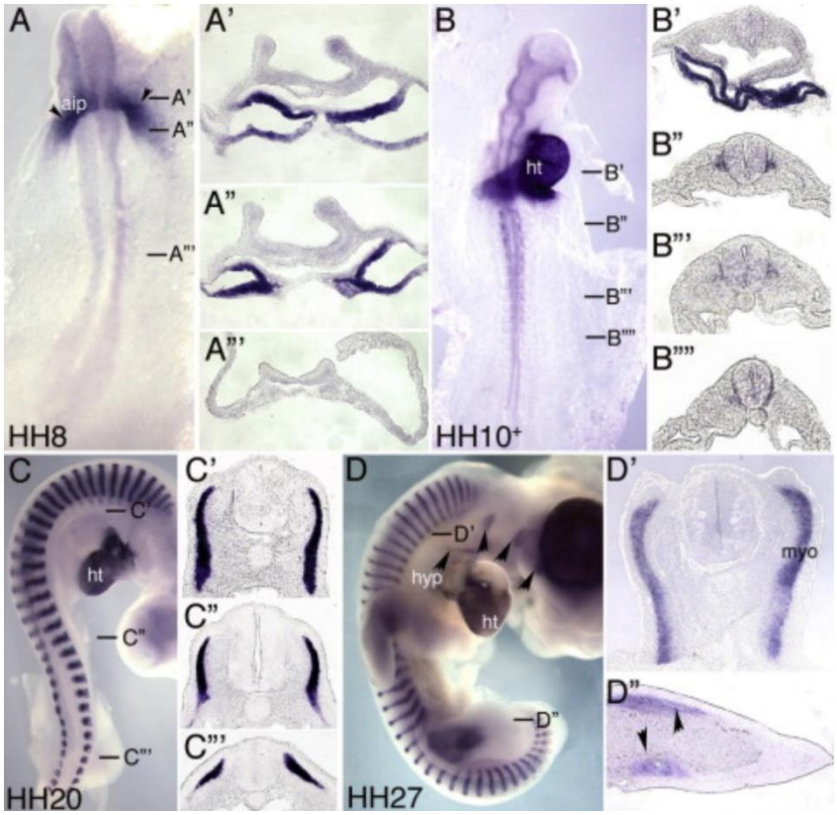

Figure 1.14: In situ hybridizations with Klhl31 antisense RNA probes on whole mount embryos, with images of cryosections. Image taken from Abou-Elhamd et al., 2009.

Other previous work done on embryos during targeted knock down of Klhl31experiments using antisense morpholinos lead to an increase in the number of mitotic cells in the myotome, whilst mis-expression of Klhl31 lead to a reduced size of myotome (Abou-Elhamd *et al*., 2015). This work suggests the importance of Klhl31 during muscle development. Experiments done to understand the underlying mechanisms demonstrated the Klhl31 interferes with β-catenin dependent *Wnt* signalling.

Klhl31 has previously been intensively studied in the Münsterberg lab due to its role in myogenesis (Abou-Elhamd *et al*., 2009., and 2015), and further characterisation of the mechanisms of its regulation by enhancer activity could prove useful in gaining a broader view of muscle development.

1.6.4 ALX1

The Alx1 gene (also known as *CART1*) belongs to the aristaless-like homeobox family of proteins, which directs the formation of structures during early embryonic development (Ettensohn, 2003). The Alx1 protein is a transcription factor, required for normal formation of the craniofacial region (Lyons *et al*., 2016). Three mutations found within this gene result in extreme frontal dysplasia type three, with individuals carrying this mutation displaying severe malformations of the face, including clefts (Uz *et al*., 2010). However, this gene also has links to skeletal muscle development. In sea urchins, Alx1 expression is restricted in the skeletal micromeres and the cellular descendants of this lineage, which become the skeletal mesenchyme (Ettensohn, 2003). In mouse embryos, expression is restricted to the craniofacial region and the limb buds (Beverdam and Meijlink, 2001).

Morpholino antisense interference experiments have suggested Alx1 expression is driven in skeletal mesenchymal cells by the ETS1 transcription factor (Damle and Davidson, 2011), and the MO- injected embryos were not able to differentiate skeletal mesoderm in sea urchins (Ettensohn, 2003). In chickens with abnormal face structures, for example, a duplicated beak, Alx1 expression is increased or altered (Buchtová *et al*., 2009). Alx1 is also responsible for the diversification of beak shape within Darwin finches, again showing the evolutionary importance of this gene (Lamichhaney *et al*., 2015).

Interestingly, in early development, Alx1 RNA expression has been detected in pre-chondrocytic mesenchymal cells responsible for the formation of cartilage, and later in the tendons and limb buds (Zhao *et al*., 1994). Embryos with an Alx1 mutation display polydactyly, and other limb deformities, suggesting a role in the formation of the limb. Alx1 may be interesting for further study due to its high expression in somites at HH14 and understanding its importance in early development will further aid our understanding of muscle and craniofacial development.

1.6.5 Scleraxis

The scleraxis (*scx*) gene codes for a protein that is a member of the basic helix-loop helix (bHLH) family (Cserjesi *et al*., 1995), which all contain a highly conserved domain, includeing a stretch of amino acids adjacent to two ALPHA-helices separated by a loop. These proteins typically form heterodimers and bind to DNA with the consensus sequence known as an E-box. Members of the bHLH family have been shown to play roles in regulation of growth and differentiation of cell lineages (Jones 2004). Other members of this family include the well characterised myogenic regulatory factors MyoD, myf5 and myogenin.

Scleraxis is expressed in the syndetome compartment of somites, which give rise to ribs and vertebrae (Cserjesi *et al*., 1995), and is also expressed in all cells of a tendon and ligament lineage (de Crombrugghe *et al*., 2000., Perez, *et al*., 2003) (see figure 1.15). The first look at the gene regulatory network controlling *Scx* induction demonstrated that removal of the dermomyotome results in no induction of *Scx*, and that fibroblast growth factors (FGFs) secreted from the myotome induce the anterior and posterior sclerotome to adopt the syndetome fate (Brent *et al*., 2003). It was found that the two important FGFs for inducing syndetome and *scx* were FGF4 and FGF8, where overexpression results in ectopic *scx* expression throughout the entire sclerotome (Brent *et al*., 2003, Edom-Vovard *et al*., 2002). Interestingly, the FGF receptor *FREK* expression is localised to the anterior and posterior myotome borders, the location of the syndetome (Kahane *et al*., 2001). Further study went on to prove that the Ets transcription factors *Pea3* and *Erm* respond to FGF signalling to induce *Scx* expression (Brent, 2004).

Scleraxis knockout experiments produce embryos with severe defects to the force transmitting and intermuscular tendons (Murchison *et al*., 2007), and a loss of distal rib development (Smith *et al*., 2005). However, the ligaments and muscle anchoring tendons remain unaltered (Murchison *et al*., 2007), suggesting that whilst scleraxis is a faithful marker of the lineage, it is not required for the complete development of all tendons.

Scleraxis has multiple downstream targets that contribute towards the regulation of tendon formation. Scleraxis can activate the collagen type 1 alpha 1 proximal promoter in-vitro (Lejard *et al*., 2007). *Scx* can positively regulate the expression of *Tendomodulin* (*Tnmd*), a marker of differentiated tenocytes (Shukunami *et al*., 2006), and *Mohawk* (*Mkx)*, along with upregulating the expression levels of proteoglycans found in the tendon matrix (Alberton *et al*., 2012).

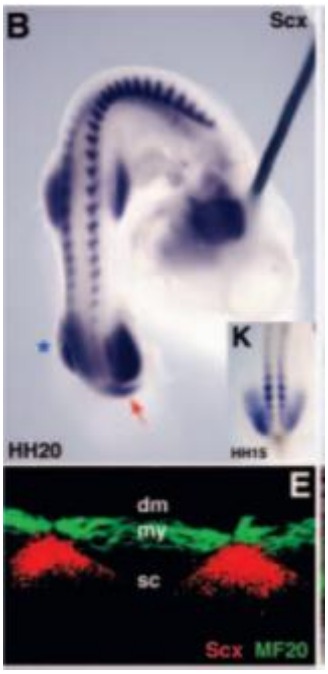

Figure 1.15: B) Whole mount in situ- hybridisation with the Scleraxis cDNA probe on a HH20 chicken embryo. E) Frontal cryosection image through a stage 24 embryo. Somites stained with MF20 antibody to visualize the myosin heavy chain protein, which is displayed throughout the myotome. Red staining shows Scleraxis transcripts. Taken from Smith et al.,2005.

Scleraxis plays an important role in tendon formation, and previous publications have attempted to locate the cis-regulatory element for scleraxis (Perez *et al*., 2003), however, were only successful in identifying a region that when deleted resulted in a lack of *scx* expression in mice. Proper characterisation of a scleraxis regulatory element in chicken development would further our understanding of the regulatory networks involved in tendon development and the gene regulatory network influencing the expression of *scleraxis*.

1.6.6 Mohawk

Homeobox protein Mohawk (*Mkx)* is a member of the TALE superclass of atypical homeobox genes, related to the iroquois family of proteins, and regulates tendon differentiation during embryological development. During mouse development, *Mohawk* is expressed in the syndetome at embryonic day 9, skeletal muscle and the sex chords of the male gonad, and the transcription of *Mkx* requires *Paraxis (*Anderson *et al., 2006)*. *Mkx-/-* mice contain hypoplastic tendons throughout the body, along with a down regulation of type 1 collagen in tendons, and drastically reduced size of tendons (Ito *et al*., 2010). *Mkx* overexpression suppresses chondrogenic differentiation while *Mkx* deficiency results in chondrogenic differentiation, suggesting that *Mkx* has many roles, such as differentiating tendons and in suppressing chondrogenic differentiation (Suzuki *et al*., 2016).

In addition to this, *Mkx* negatively regulates myogenic differentiation, through a conserved sequence named Mohawk Repressor Domain1, 2 and 3 (MRD1,2 and 3) (Anderson *et al*., 2007). *Mohawk* also represses myogenic differentiation by association with the co-repressor complex Sin3A Histone Deacetylase (Adams *et al*., 2018) and polymerase II general transcription factors, such as TFIIA1,

TFIIB and Tbp (Anderson *et al*., 2009). Binding to these transcription factors results in gene silencing, due to the inability of the RNA polymerase II formation at transcription start sites (Heinzel *et al*., 1997).

Most importantly, it has been shown that through using these repressive controls, *Mkx* is able to repress *MyoD*, a gene known for its role at the top of the myogenic differentiation pathway. It was shown that in 10T1/2 fibroblast cells expressing *MyoD* that were exposed to full length *Mxk* transcript the number of muscle specific protein myosin heavy chain were reduced drastically. It also reduced the number of myotubes than the cells expressing just *MyoD*. The expression of both *Mkx* and *MyoD* resulted in a 47% reduction of the number of MHC-positive myotubes when compared to just *MyoD* alone (Anderson *et al*., 2009). *Mohawk* plays a large role in tendon differentiation, and identifying any enhancers associated with it would further our understanding not only of the tendon differentiation pathway, but also the chondrogenic and myogenic pathways.

1.6.7 TN-C

Tenascin C (*TN-C*) is a glycoprotein that is the founding member of the tenascin gene family. It was chosen as a tendon and cartilage differentiation marker as it is expressed not only in embryogenesis, but also in adult tendons (Saunders *et al*., 2012). *TN-C* appears to play a role in early differentiation events, as one of the earliest markers of maturing tendons and cartilage and can stimulate chondrogenesis *in vivo* (Gluhak *et al.* 1996). During early limb development, Tenascin-C is selectively associated with condensing chondrogenic mesenchyme (Mackie and Murphy, 1998). *TN-C* in tendons is a direct key target of *Scx*, where it binds to a variety of other extracellular matrix proteins, such as collagen and fibronectin (Chen *et al*., 2016) Mutations within the *TN-C* gene have been associated with Achilles tendinopathy, where the body cannot repair any damage done to the Achilles tendon (Saunders *et al*., 2012), and tendon ruptures.

TN-C is an interesting gene to understand as it marks the transition into mature tendons and chondrocytes. Investigating enhancer regions for this gene could further our understanding in what regulates the transition of chondrocytes and tenocytes into mature, fully functioning cartilage and tendon cells.

1.6.8 Pax1 and Pax9

Pax1 is a member of the paired box family of highly conserved transcription factors, characterised by the presence of the DNA binding domain called the paired box. This family of proteins were originally discovered in *Drosophila*, but are conserved throughout nearly all species, and are associated with pattern formation during vertebrate development (Wallin *et al*., 1994). *Pax1* is well known for its role in the development of the ventral vertebral column (Peters *et al*., 1999), and mutations within this gene often lead to diseases associated with malformation of the spinal cord, including Jarcho-Levin syndrome (Bannykh *et al*., 2003) and Kippel-Feil syndrome (McGaughran *et al*., 2003). However, *Pax1*, along with *Pax9* plays a major role in chondrogenesis, and is highly expressed at early stages of embryo development in the sclerotome (Peters *et al*., 1999), see figure 1.16. Experiments done in embryos deficient for both *Pax9* and *Pax1* shows defects with condensation of the sclerotomal cells around the notochord, along with a decrease in sclerotomal cell proliferation, and thus chondrogenesis is not initiated (Peters *et al*., 1999).

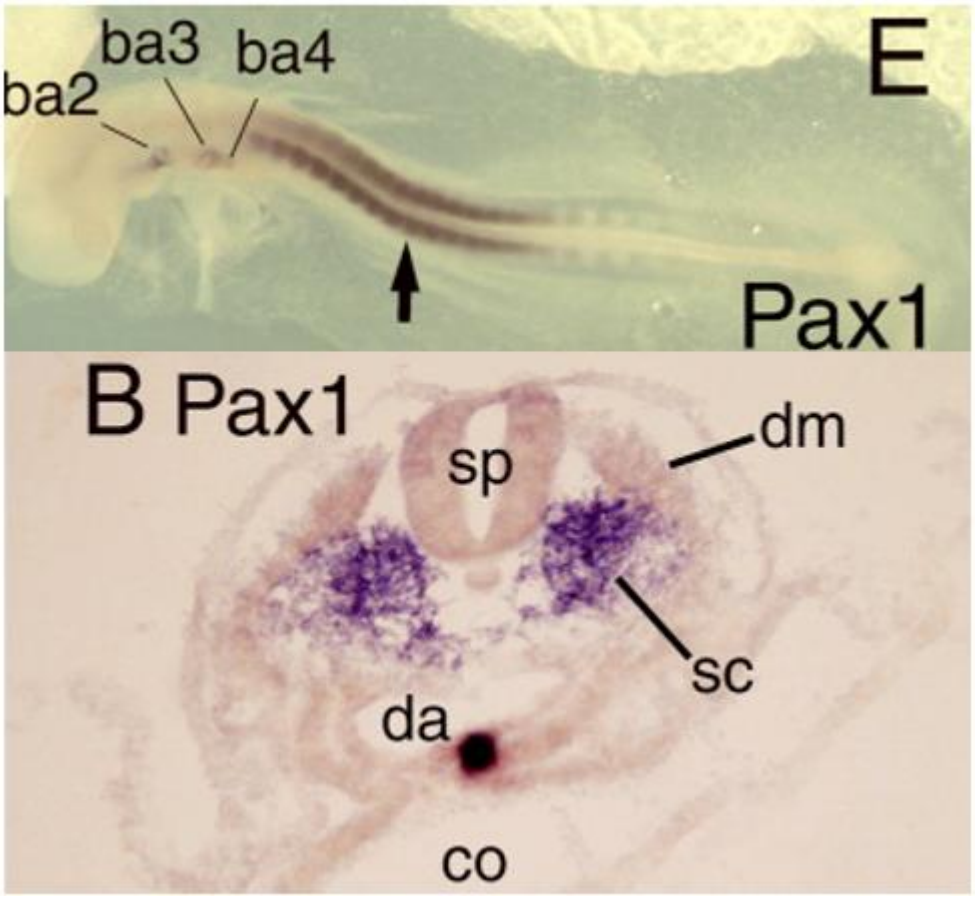

Figure 1.16: Above: Whole-mount in-situ hybridisation at HH 13 showing detection of Pax1 mRNA in the somites and the branchial arches. Below: Cryomount section of this in-situ hybridisation, showing Pax1 hybridising to the sclerotomal regions. Images taken from the Geisha database (http://geisha.arizona.edu/)

Pax1 and *Pax9* are two key genes involved in sclerotomal fate. Along with this, a downregulation of *Pax1* and *Pax9* are crucial for the formation of the syndetome. These two genes are early regulators of both cartilage and tendon fate and understanding regulation of these genes will aid our knowledge of sclerotomal and syndetome formation.

1.6.9 MEOX2

Mesenchyme homeobox 2 (*Meox2*) is a transcription factor that specifies mesodermal tissue that is fated to become the muscle lineages (Valcourt *et al*., 2007) (see figure 1.17). Knockout of *Meox2* in the mouse leads to a reduction of muscle mass and a loss of specific muscle types from the limbs (Mankoo *et al*., 1999). They also develop cleft palate due to a lack of mesenchymal cells in the region of palatal fusion (Jin and Ding, 2006). *Mef2* regulates *Meox2* expression during myogenesis and is also a gene target of the TGFβ signalling pathway (Valcourt *et al*., 2007). *Meox2* also binds to an enhancer that regulates the expression of the *Myf5* gene and is required for activation of *Myf5* activation in the myotome (Daubas *et al*., 2015).

Figure 1.17: A) Whole mount in-situ hybridisation on HH.13 and B) HH.17 chicken embryo hybridising to Meox2 mRNA. Expression of Meox2 is found in the neural tube/plate, as well as the somites. Images acquired from Geisha database (http://geisha.arizona.edu/).

1.7 The Chicken as a Model for Development

The use of chick as an experimental model for researching developmental processes has led to some major discoveries. Experiments in chicken embryos have contributed significantly to understanding embryology, including the understanding of Left/Right asymmetry in the placement of internal organs (Raya *et al*., 2004), and experiments that revealed the signalling pathways in the limb (Towers and Tickle, 2009). Chicken embryos are used for several reasons; the chicken egg is a common and available source, allowing for quick experimentation. Chick embryos take only three weeks to fully develop and can be brought to the desired developmental time point by use of an incubator. As the embryo develops ex-ovo, no harm comes to the mother, making it an excellent choice for ethical reasons. Embryos within the eggs can be manipulated *in-ovo*, or they can be removed and grown *ex-ovo*, allowing development to be properly monitored.

In the early 2000s, scientific advances have made the chick an even more attractive model for studying embryo development; the whole chicken genome was properly sequenced and expression vectors were optimised for different applications. For example, vectors were developed for misexpression experiments that could be targeted to a group of cells using electroporation techniques (Yaneza *et al*., 2002, Nakamura and Funahashi, 2001). This approach allows for quick gain of function experiments, with control over both the positioning and timing of the gene of interest. For loss of function experiments, dominant negative constructs, or fluorescein-labelled morpholino could be introduced (Sheng *et al*., 2003, Nakamura *et al*., 2004). This approach gave advantage over the previous popular *Xenopus* embryos, due to the spatial precision and developmental time point the morpholino or construct is introduced at. The chicken is an ideal model to study developmental processes and can yield quick results.

Aims

Enhancers are clearly important in genetic regulation and embryonic differentiation. For example, the enhancer that controls *Myf5* is essential, as mutations within this region result in a lack of muscle differentiation (Daubas and Buckingham, 2013). *Mef2* binds to many enhancer regions of muscle development genes, allowing for their activation and transcription. The aim of this project was to elucidate the control mechanisms that regulate myogenesis, chondrogenesis and the development of tendons, by identifying novel enhancers that are active in somites that may regulate genes involved in the development of the musculoskeletal system. To identify the likely candidate enhancer regions, the ATAC-sequencing dataset already established in the Münsterberg laboratory was utilised to identify regions of open chromatin near to genes of interest that may contain regulatory cis-elements. These are cloned into a fluorescent reporter construct, and then injected into developing embryos and screened for reporter activity.

Chapter 2: Methods and Materials

2.1 ATAC sequencing and analysis

Previously generated ATAC-sequencing datasets were utilised to find putative enhancer regions. ATAC-sequencing was performed as described previously (Buenrosto *et al*., 2015). The ATAC libraries were made from pre-segmented mesoderm and somites that had been removed from a stage 14 embryo at one of three defined stages of differentiation: epithelial somite, maturing somite and differentiating somite (Fig.4). The sequencing reads were mapped to the *Gallus gallus* genome and can be viewed using the UCSC genome browser. Leighton Folkes developed the bioinformatics pipeline, including a differential peak-calling algorithm. Higher levels of sequence reads indicated areas of high transposability, denoting regions of open chromatin. Putative enhancer regions were picked based on their sequence homology with other vertebrates, which can be viewed on UCSC genome browser.

2.2 RNA and ATAC sequencing/transcriptomics and analysis

The Münsterberg lab had previously generated both ATAC-sequencing and RNA-sequencing data sets. Both data sets were generated from dissected pre-somitic mesoderm (PSM), epithelial somite (ES), maturing somite (MS) and differentiated somite (DS) from a HH14 embryo. The sequencing was repeated in triplicate (Mok *et al*., 2021).

To utilise this data set to identify genes involved with early muscle differentiation in the somite, the expression pattern of *MyoG* was identified (see figure 2.1 below), and genes that shared a similar expression pattern were analysed. As *MyoG* is active in differentiating somites, due to the role it plays in early muscle differentiation, its expression profile is useful to find other genes involved in the same pathway. Generally, the other genes identified showed lower levels of RNA expression in the PSM and higher levels of RNA expression in the differentiated somite (fig.3.6 and 3.7). Genes involved with tendon and cartilage expression were first identified from literature research. The RNA sequencing was then utilised to pick genes from the literature search that showed high levels of RNA expression in the differentiated somites compared to the PSM.

The genes chosen from the RNA-sequencing dataset and introduced in 1.5 were viewed in the UCSC Chicken Genome Browser (2015), with the ATAC-sequencing data mapped to it (Mok *et al*., 2021). The ATAC-sequencing dataset had also been generated by dissecting out individual regions of PSM, singular epithelial, maturing and differentiated somites and performing the ATAC-sequencing protocol discussed in 1.53. Peaks in the ATAC-sequencing data set correspond to accessible and bound regions of chromatin and were therefore identified as candidate enhancer regions (fig.3.6 and 3.7). In addition, sequence conservation between species was considered when picking candidate enhancer regions. An example is shown in figure 3.8 below for the ALX1 locus, viewed in the USCS chicken genome browser. Some regions were also picked based on the presence of HK27ac3 histone modification present in humans as this mark is involved with active cis-regulatory regions.

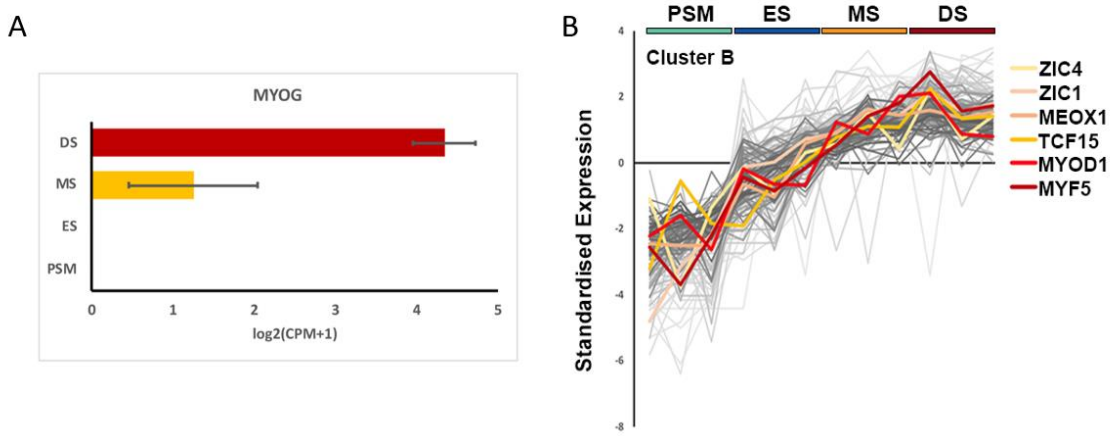

Figure 2.1: The RNA sequencing was performed on dissected pre-somitic mesoderm (PSM), epithelial somite (ES), maturing somite (MS) and differentiated somite (DS). A) The expression profile for MyoG, counted in counted in log2 counts per million plus 1. This shows high levels of expression of MyoG in the differentiated somite B) The expression levels of genes identified as the myotome cluster K-means clustering analysis on the RNA sequencing. Some genes are labelled to identify the key transcription factors found in the myotome. (Mok et al., 2021).

2.3 Visualising Gene Expression

Patterns of gene expression were detected using whole-mount *in-situ* hybridisations (WISH). This procedure elucidates the spatial and temporal regulation of gene expression in embryos, by visualising mRNA expression. Probes were designed for mRNA transcripts produced by genes found to be of interest from the analysis of the RNA-sequencing dataset.

2.3.1 Probe Preparation and Whole Mount in situ Hybridisation

Digoxigenin labelled antisense probes were synthesised by first designing primers that were specific to the cDNA of the gene of interest. Primer design used the NCBI primer design tool (Ye *et al*., 2012), and some of the parameters considered when choosing primers included a GC content of less than 60%, and primer melting temperatures between 57°C and 63°C, with an optimum of 60°C. A full list of primers can be found in table SI tables 1 and 2.

Template cDNA was synthesised from total RNA from a HH14 chick embryo. Fragments between 500bp and 1.5kb in length for the gene of interest were amplified (KAPA LongRange, Sigma-Aldrich; Cat no./ID LRHSRMKB) and purified (QIAquick PCR Purification kit; Cat No./ID: 28104). The derived PCR products were cloned into the pGEM-T vector (Promega; Cat no./ID A1360), which contains an ampicillin resistance gene (AmpR), a *lacZ* locus and Sp6, T7 and M13 priming sites (see figure 2.2). The ligation product was used to transform competent *E. coli* bacteria (Strain: DH5α). Transformants were plated onto selective LB agar plates and colonies were grown overnight. LB plates were spread with x-gal, to allow for selection of white colonies. Positive colonies were grown in 5 ml LB liquid culture overnight at 37°C. DNA was extracted from cultures using a QIAGEN mini-prep kit (Cat No./ID 27104), and 10ng/µl of DNA was sequenced at Eurofin Genomics. Returned samples with correct sequencing were checked for orientation and PCR amplified with M13 primers. Resulting PCR products were used to synthesise anti-sense RNA probed by *in-vitro* transcription using T7 and SP6 RNA polymerases. The probes were then purified using GE Healthcare Illustra probequant G-50 micro columns (cytiva; product code 28903408).

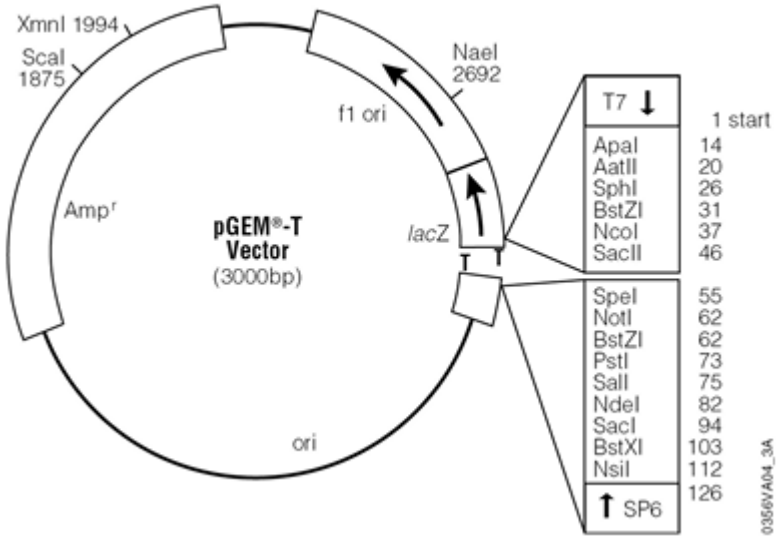

Figure 2.2: A vector map of the pGEM-T vector, which contains a LacZ site with multiple restriction sites for cloning of constructs, as well as an Ampicillin resistance gene. Image taken from Promega (https://www.promega.co.uk/).

Chick embryos at the desired stages were harvested and fixed in 4% PFA and stored for up to six months. Whole mount RNA *in-situ* hybridisations were performed as described previously (Nieto and Wilkinson., 1996). Post hybridization washes were carried out using MABT (maleic acid buffer containing 0.1% Tween-20) and target-specific probe binding was visualized using NBT/BCIP (nitro-blue tetrazolium chloride/5-bromo-4-chloro-3'-indolylphosphate p-toluidine salt) as the colour marker. The time to develop the colour reaction varied with each probe, from 2 hours to several days.

2.4 Histological Sectioning and Imaging

Fixed embryos were immersed in 30% sucrose in PBS overnight at 4°C before embedding in OCT Embedding matrix (CellpathTM, Fisher Scientific; Cat no./ID 1521277). Embedded samples were mounted in a cryostat and sectioned into 20µm slices. These were then mounted onto microscope slides. Cryosections were then washed (3x 10 min washes in dH_2O), before applying a slide cover using hydromount.

2.5 Cloning Plasmid constructs

Enhancer regions identified by ATAC-sequencing analyses were PCR amplified (KAPA LongRange, Sigma-Aldrich; Cat no./ID LRHSRMKB) from chicken genomic DNA (PureLink gDNA mini kit, Invitrogen; Cat no./ID K1820-00). Golden Gate based cloning of putative enhancer regions into a ptk-Citrine vector, which contains a *lacZ* locus (see figure 2.3) was then performed. Plasmid safe reactions were performed as this selectively hydrolyses linear double stranded DNA, removing any un-ligated plasmids and contaminating DNA. This reaction was performed by combining 10.25µl of the previous BsmBI reaction to 1.25µl plasmid safe buffer, 0.5µl 25mM ATP and 0.5µl plasmid safe enzyme. Plasmid constructs were then transformed into competent DH5-α *E. coli* cells, which were grown in 5ml liquid LB culture at 37°C. PCRs on ampicillin resistant colonies were performed to ensure the inserted sequence was the correct size. Plasmid DNA was extracted using a QIAGEN plasmid maxi prep kit (Cat no./ID 12162), ready for injection into embryos.

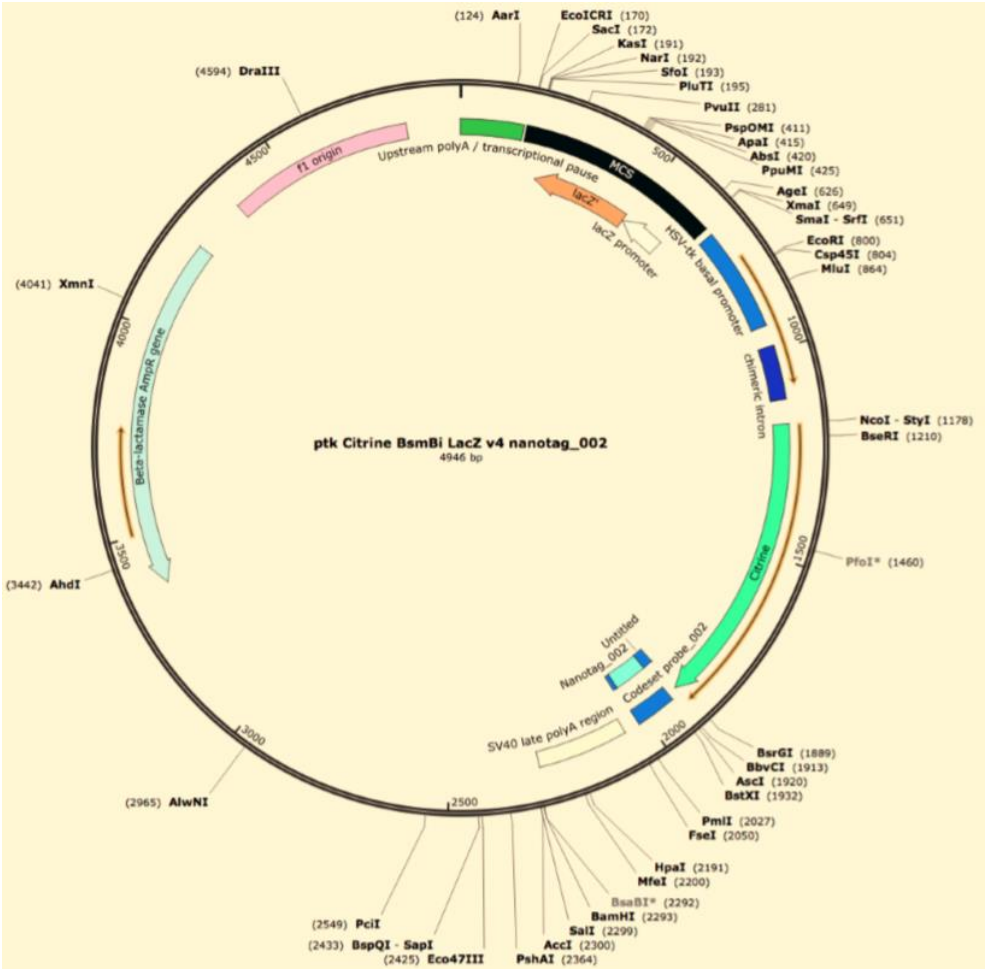

Figure 2.3: The ptk-Citrine vector map. This contains the ampr gene for initial selection of plasmids, a TK minimal promoter, a lacZ operon and the citrine gene. Putative enhancer regions would be inserted into the plasmid between the minimal promoter and the citrine reported gene, disrupting the lacZ operon.

2.6 Easy Culture

To culture gastrulating chicken embryos, early chick (EC) culture methods were used (Chapman *et al.*, 2001). Fertilised chicken eggs from Henry Stewart & Co. Ltd were incubated for 20-22 hours at 37°C. Egg contents were deposited into a petri-dish, and an area of albumen was cleared from around the embryo. A ring of filter paper with a hole in the centre was positioned over the embryo and yolk and allowed to attach to the vitelline membranes, framing the embryo. The embryo was then attached under tension to the filter ring and, after cutting around the outside of the filter ring, can be removed from the yolk. The embryo is then washed, inverted, and placed ventral-side up on 30mm petri-dishes containing an agar-albumen mixture as described by Chapman (Chapman *et al.*, 2001). Embryos were then grown at 37°C to the desired stage of development.

2.7 Plasmid injections and electroporation

HH3+ chick embryos harvested via EC method as described were transferred ventral side up to an electroporation chamber containing Ringer's Buffer and a 2mm2 platinum-based electrode. Using a pulled glass capillary and a microinjector, embryos were injected with the candidate putative enhancer reporter plasmid (1µg/µl) in 10% injecting dye (Fast Green FCF, Sigma-Aldrich; Cat no./ IDF7252) throughout the embryo proper, between the primitive streak and the epiblast. Electroporation was then performed using a $2mm^2$ positive electrode (cathode) which was placed above the embryo in the Ringers buffer solution. Five pulses (5V for 50ms at 100ms intervals) were then applied using a square wave electroporator. Embryos were returned to the agar albumen petri-dishes and grown for 20-22 hours at 37°C. They were then observed under a fluorescent stereomicroscope at stages HH10-HH14. Embryos that displayed fluorescent reporter activity were fixed in freshly thawed 4% paraformaldehyde (PFA)

and stored at 4°C for up to 4 weeks. Table 2 in the supplementary information (SI) shows all the putative enhancer regions, if they were successfully cloned into the reporter construct, the number of injections, if these were successful or not and if a positive control was used. pCIGcherry was co-injected with the fluorescent reporter construct, and MEOX1 enhancer construct was injected into different embryos but at the same time and under the same conditions as the investigated construct. SI Table 1 shows the primers that successfully amplified candidate enhancer regions which were then cloned into a reporter construct.

2.8 Summary of Enhancer discovery Pipeline

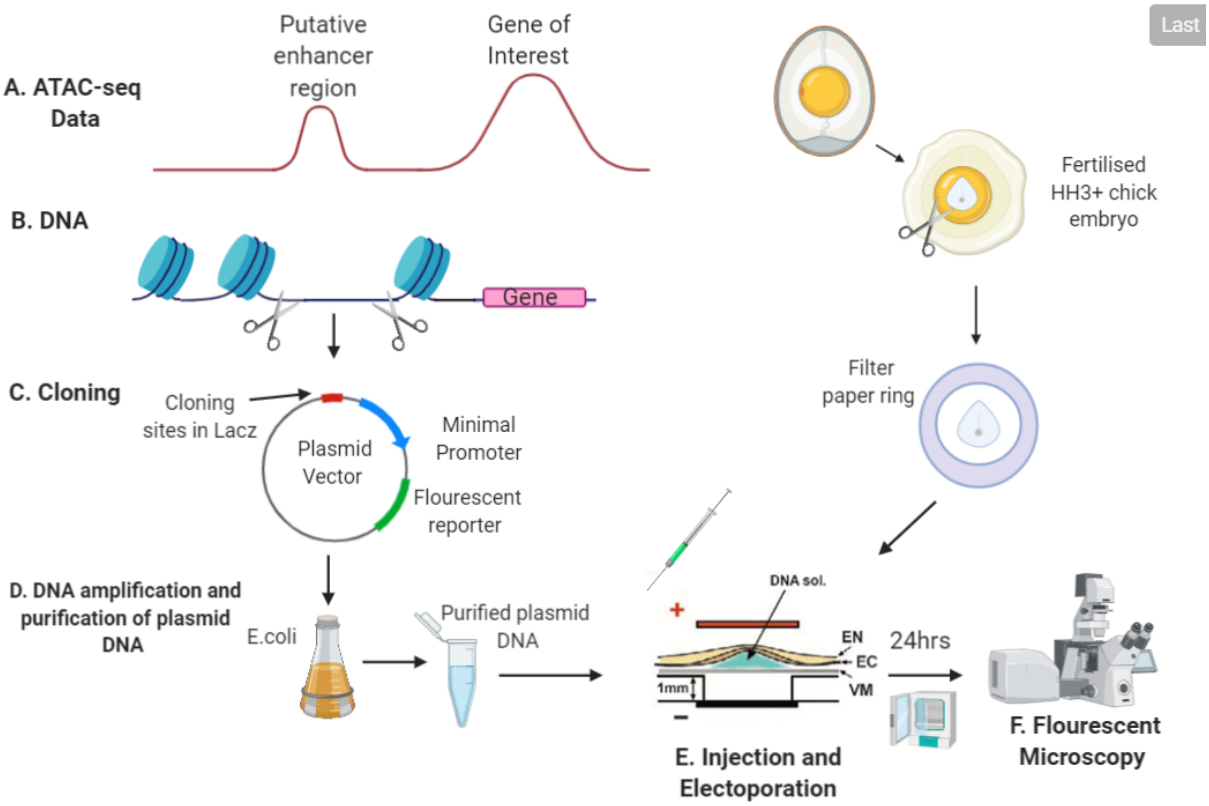

Figure 2.4: A diagram showing the experimental pipeline. A) ATAC-sequencing can identify regions of open DNA, that are devoid of nucleosomes and histones, as shown in B. The enhancer region is cloned into a fluorescent reporter plasmid vector (C), transformed into E. coli, to amplify the plasmid DNA (D). Purified plasmid DNA is electroporated into HH3+ chicken embryos (E), which are cultured using easy-culture methods and cultured in agar-albumen dishes (Chapman et al., 2001). The embryos are left to grow for 24hours in a 37°C incubator before being visualized under a confocal microscope with UV light (F), to look for expression of the GFP protein, indicating that the enhancer is also expressed in those regions.

2.9 Foxo-1 sequence conservation

The Foxo1 whole mount in-situ hybridisation showed specific expression signal in the somites (see section 3.1). As this was a novel discovery, with Foxo1 not previously being described as present in the somites at early stages, it was further investigated for sequence conservation between species, to understand if this protein is fundamental in most organisms. Firstly, a protein sequence alignment was performed using Aliview software (Larson., 2014) to identify regions of homology or regions where the sequence is different between species (see figure 3.12). From this protein alignment, a phylogenetic tree for this gene could be produced, using the MEGAX software (fig. 3.13).

Chapter 3: Results

3.1 Whole Mount in-situ Hybridisation Results

Gene expression patterns were visualised for the genes chosen from RNA-sequencing dataset to ensure they were expressed in the somites during chick development. Probes were designed for the genes Foxo1, Scx, Pax1, Tn-C, Mkx, Alx-1 and UncX (see SI Table 1 for a list of primer pairs). However, only the Foxo1, Scx and Pax1 probes were completely synthesised and showed a specific signal, restricted to the somites. As there is no other characterisation of Foxo1 in chicken embryos, in either GEISHA (A database of expression patterns for chick embryos) or in any other literature, a stage series of whole mount in-situ hybridisations for this gene was performed (see figure 3.1-3.3), to observe the spatiotemporal patterns of Foxo1 gene expression throughout development. Sections were performed on the HH14 embryo to further characterise gene spatial expression (figure 3.4).

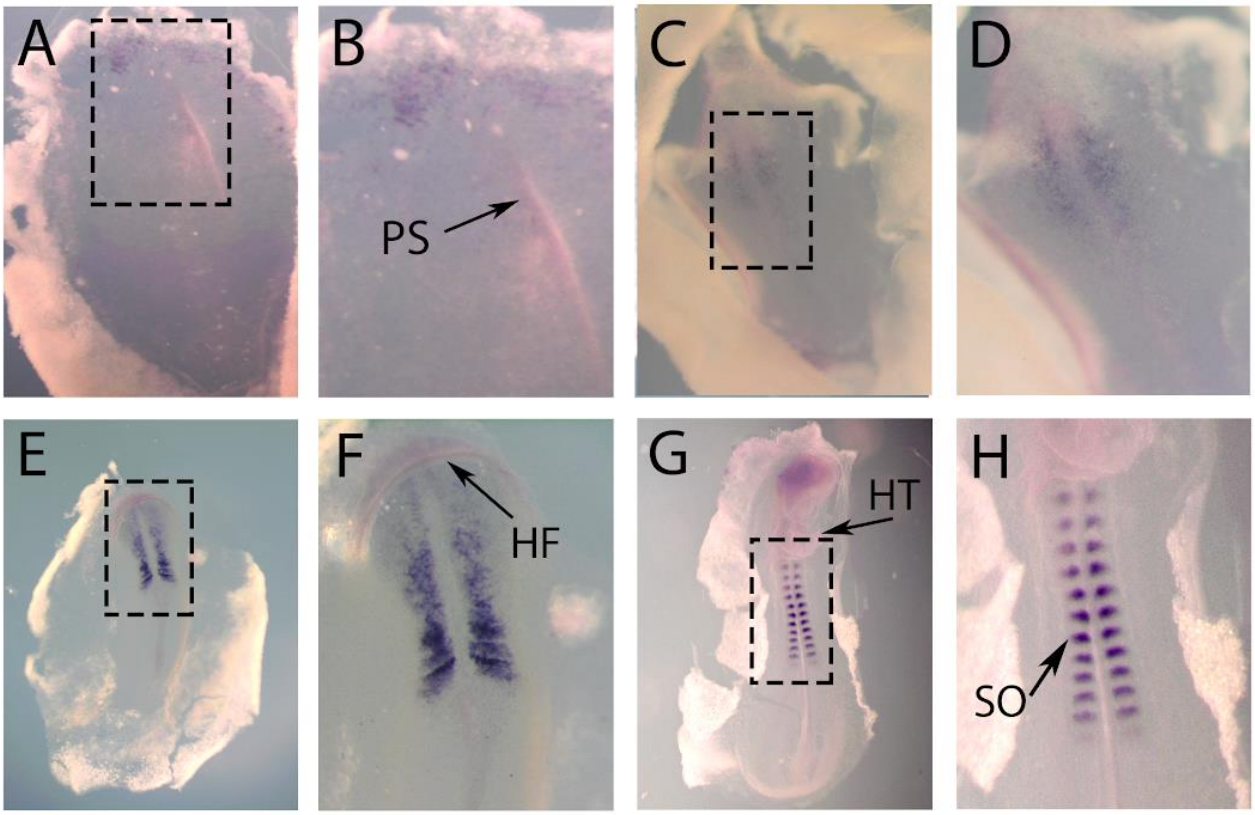

Figure 3.1: Whole mount In-Situ hybridizations showing patterns of mRNA expression of the Foxo-1 gene throughout development; A) HH.4, C) HH.6, E) HH.8 G) HH.10. B, D, F and H are close ups of the regions depicted in the boxes in A, C, E and G respectively.

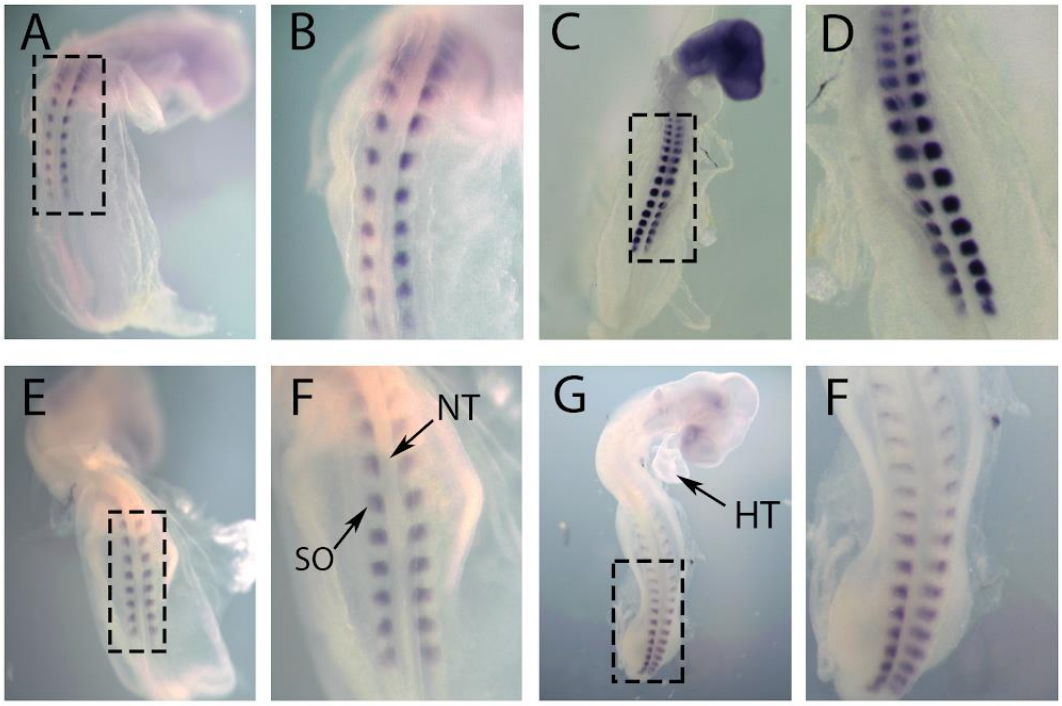

Figure 3.2: Whole mount In-Situ hybridizations showing patterns of mRNA expression of the FOXO-1 gene throughout chicken development; A) A HH12 embryo, C) A HH14 embryo, E) A HH15 embryo and G) a HH17 embryo. B, D and F show close ups of the regions depicted in boxes on the embryos in A, C and E respectively.

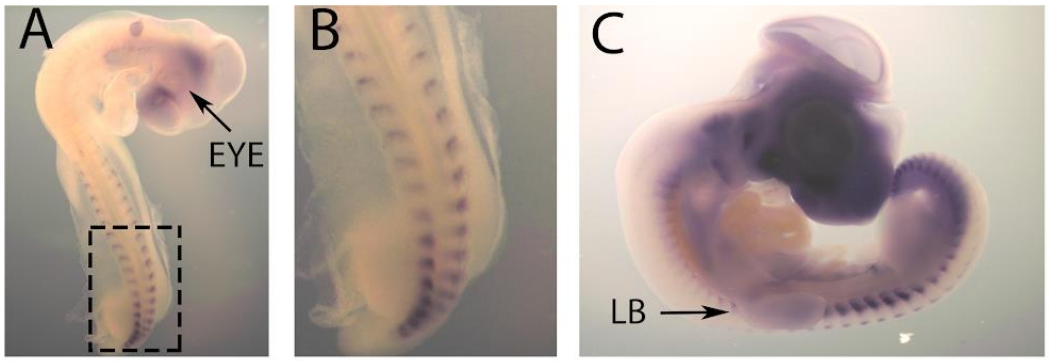

Figure 3.3: Whole mount In-Situ hybridizations showing patterns of mRNA expression of the FOXO-1 gene throughout chicken development; A) a HH18 embryo, and C) HH25 embryo. B) Close up of the boxed region in A.

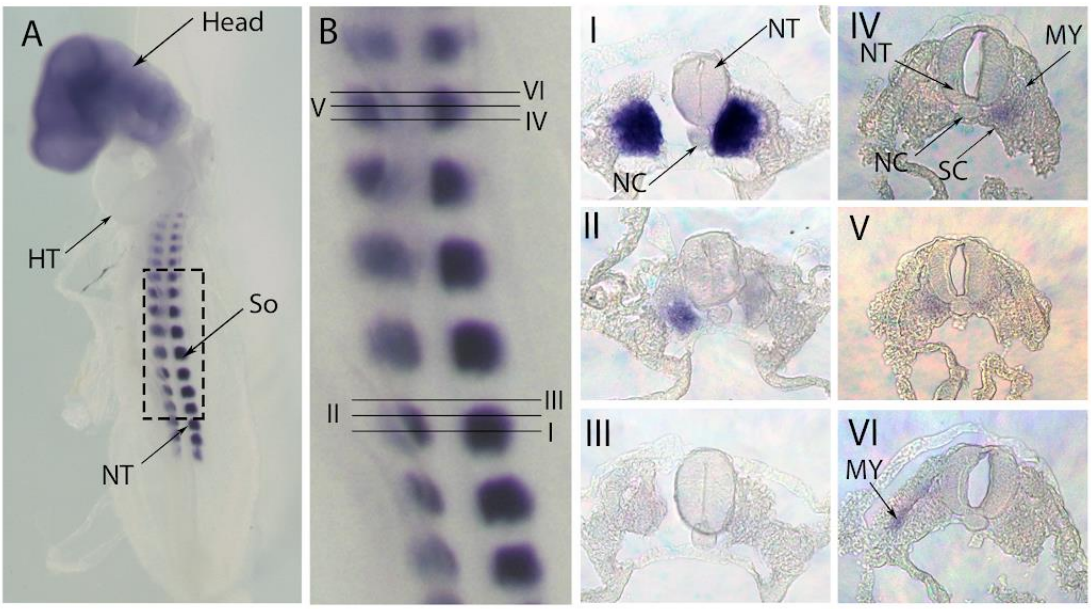

Figure 3.4: Cryostat sectioning of a HH 14 embryo into 20µm sections. A) HH14 embryo that is sectioned. B) A close up of the region in the box shown in A, and where the corresponding sections were taken from. I-VI) 20UM sections taken at 10x magnification. For abbreviations, see SI 1.1.

Figure 3.5: Whole mount in -situ hybridisations on HH14 embryos for A) PAX-1 and B) SCX.

3.2 ATAC sequencing regions and RNA-sequencing

The ATAC-sequencing was mapped to GalGal 5 on the UCSC genome browser. Figs. 3.6 and 3.7 show the locus of the gene of interest with mapped ATAC peaks below. The four different stages of somite differentiation are represented by different colours: from the PSM track in green, the ES track in blue, MS track in yellow, and the DS track in red. The expression data from the RNA-sequencing is shown alongside the corresponding gene track. Regions chosen as putative enhancer regions are highlighted in blue. A list of primer pairs that amplified these chosen regions can be found in the supplementary information (Table 1).

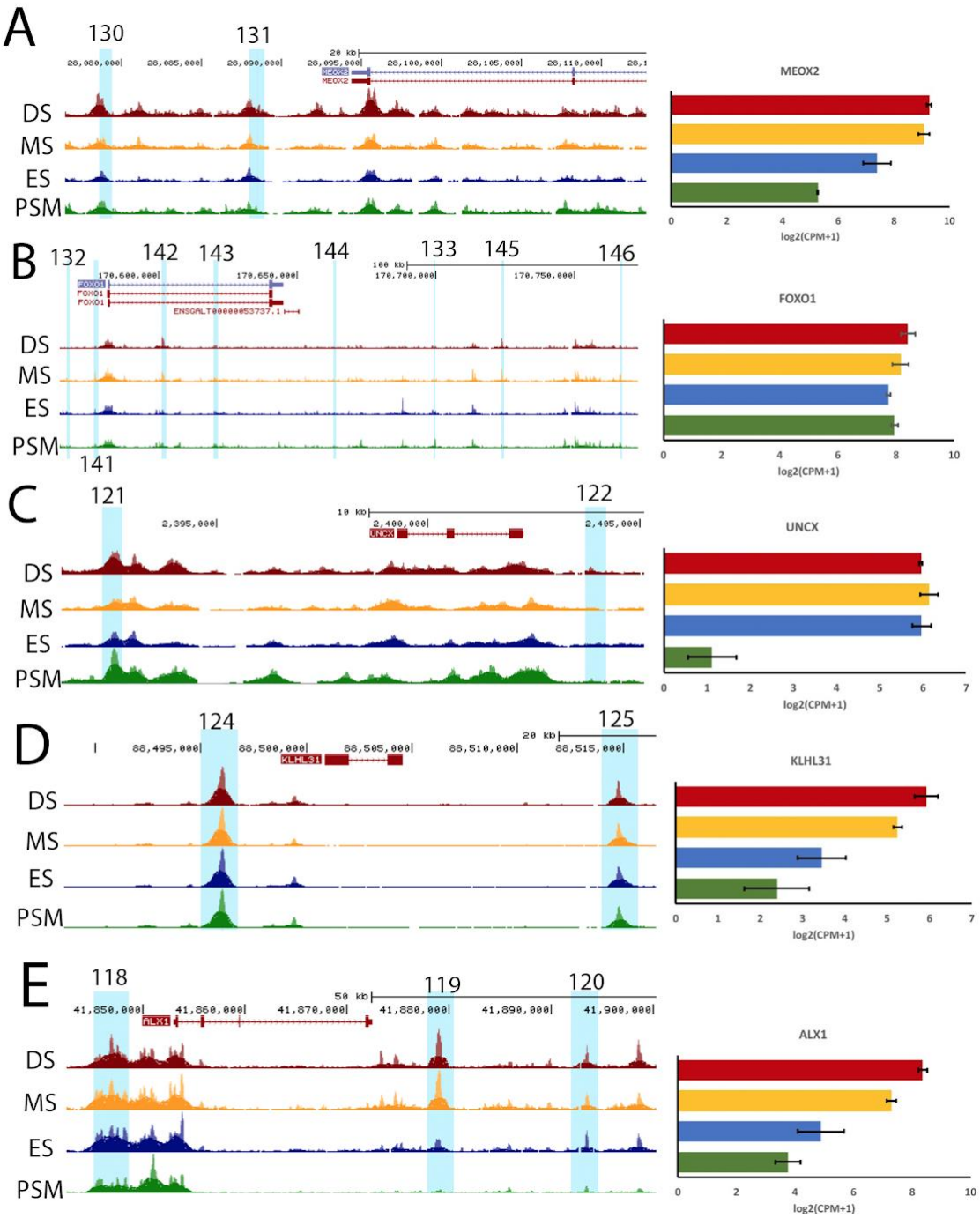

Figure 3.6: The ATAC-sequencing dataset mapped to the UCSC chicken genome browser (2015) for genes involved with muscle development. Peaks chosen for subsequent amplification and cloning are highlighted in blue. The different tracks represent different stages of somite development; DS is the most differentiate somite, MS is a maturing somite, ES is an epithelial somite, and PSM is the presomitic mesoderm. Each image shows the ATAC-sequencing tracks and regions chosen as candidate enhancers for, A) MEOX2, B) FOXO1, C) UNCX, D) KLHL31 and E) ALX1. Above each peak is a unique identifying number for each candidate enhancer region. Alongside the ATAC-sequencing tracks are the gene expression levels for the corresponding gene, based on RNA sequencing, counted in log2 counts per million plus 1.

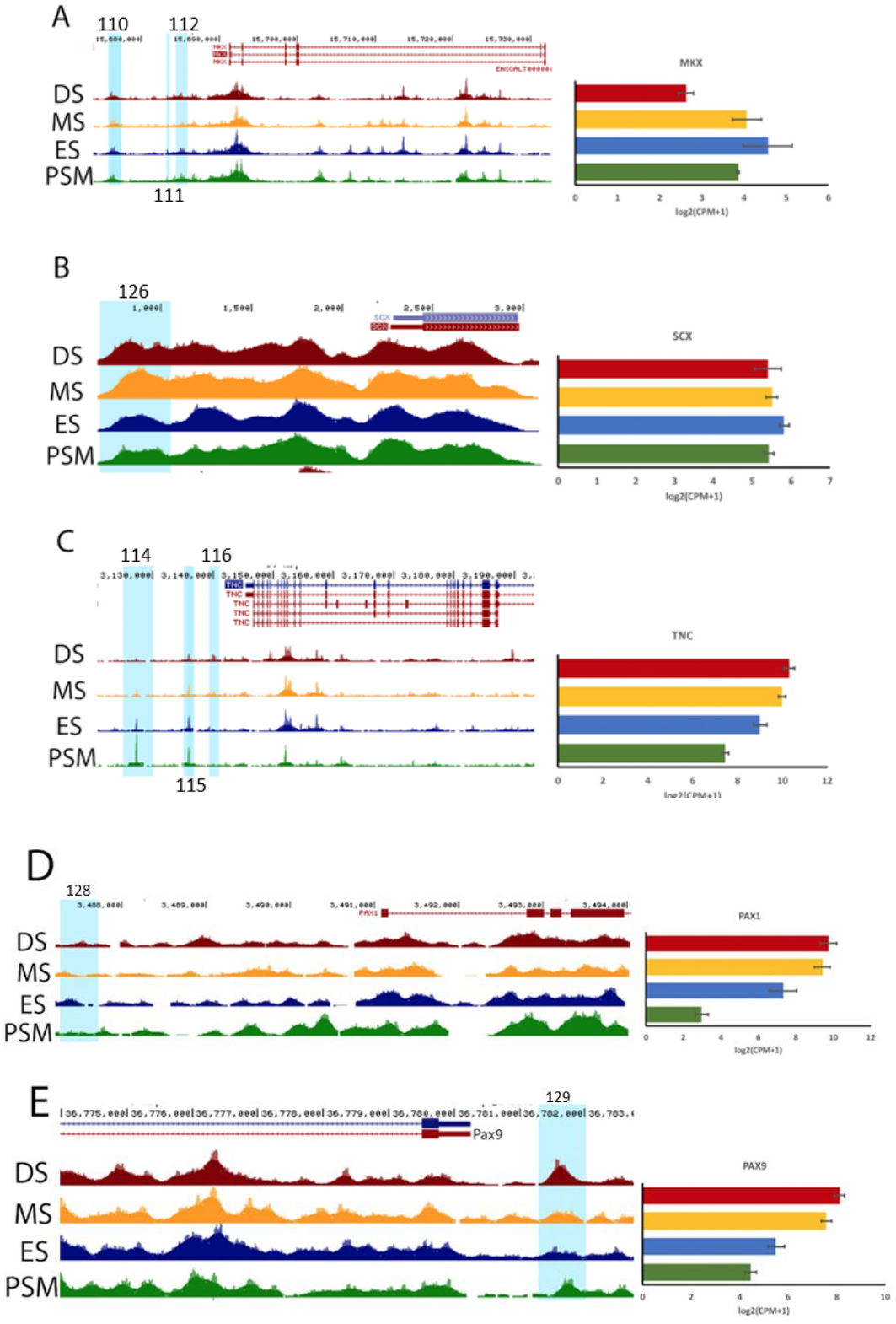

Figure 3.7*: The ATAC-sequencing dataset mapped to the UCSC chicken genome browser (2015) for genes involved with cartilage and tendon development. Peaks chosen for subsequent amplification and cloning are highlighted in blue. The different tracks represent different stages of somite de development; DS is the most differentiate somite, MS is a maturing somite, ES is an epithelial somite, and PSM is the presomitic mesoderm. Each image shows the ATAC-sequencing tracks and regions chosen as candidate enhancers for, A) Mxk, B) Scx, C) TNC D) Pax1 and E) Pax9. Above each peak is a unique identifying number for each candidate enhancer region. Alongside the ATAC-sequencing tracks are the gene expression levels for the corresponding gene, based on RNA sequencing, counted in log2 counts per million plus 1.*

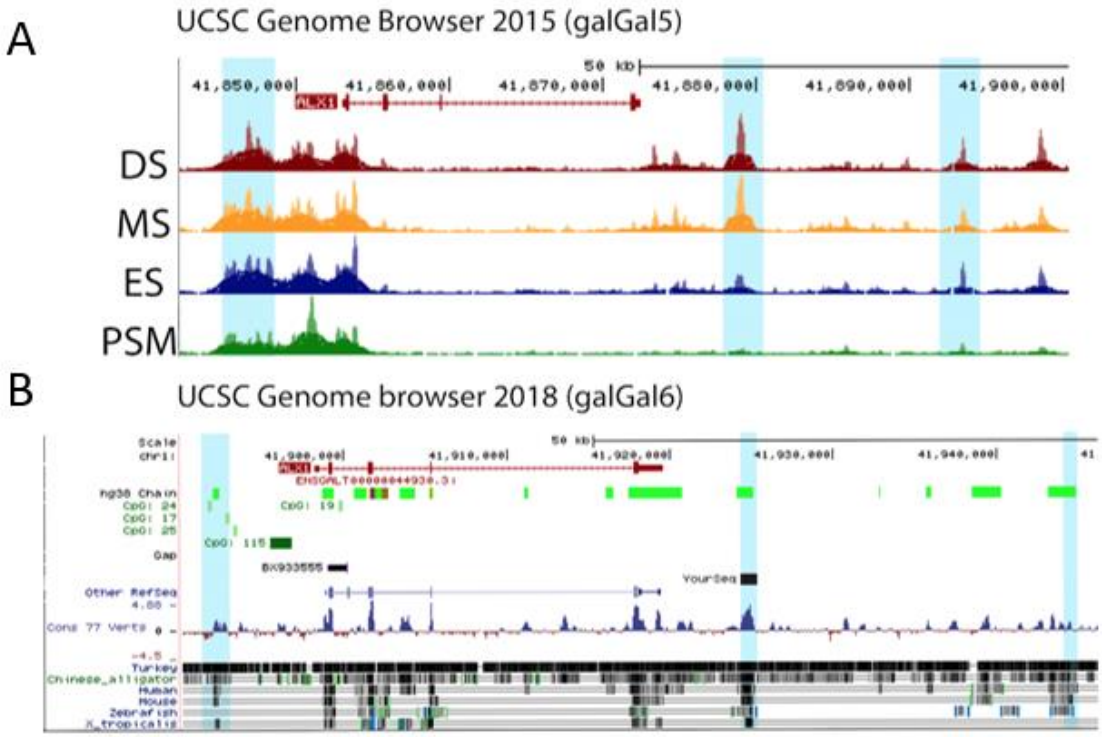

Figure 3.8: How putative enhancer regions were picked. A) Using the ATAC-sequencing data set which visualises open regions of chromatin as a peak and is mapped to the chicken UCSC genome browser galGal5 version. B) Using sequence conservation with other species, in the UCSC genome browser galGal6 version. there are tracks for conservation with turkey (Meleagris gallopavo), Chinese alligator (Alligator sinensis), human (Homo sapiens), mouse (Mus musculus), zebrafish (Danio Rerio), and Frog (Xenopus tropicalis). All candidate enhancer regions are highlighted in blue, using the ALX1 locus as an example.

3.3 Visualising the Fluorescent Reporter Constructs

HH3^{+} embryos were injected around the primitive streak and above the node with a GFP reporter construct that contained the putative enhancer region.

3.3.1 MEOX1 Enhancer

To test electroporation into HH3+ embryos positive controls were used. GFP and Cherry reporter constructs were injected and electroporated, to test for electroporation efficiency and targeting (see supplementary figure S.2 for examples). Another positive control was a MEOX1 enhancer reporter construct, which had been previously generated in the lab by the same pipeline discussed in the methods section. This had been shown to be active in somites (Mok *et al*., 2021). This construct was injected and electroporated and served as a control for electroporation efficiency. Figure 3.9 shows reporter enhancer activity in the somite of a HH10 embryo (N=15).

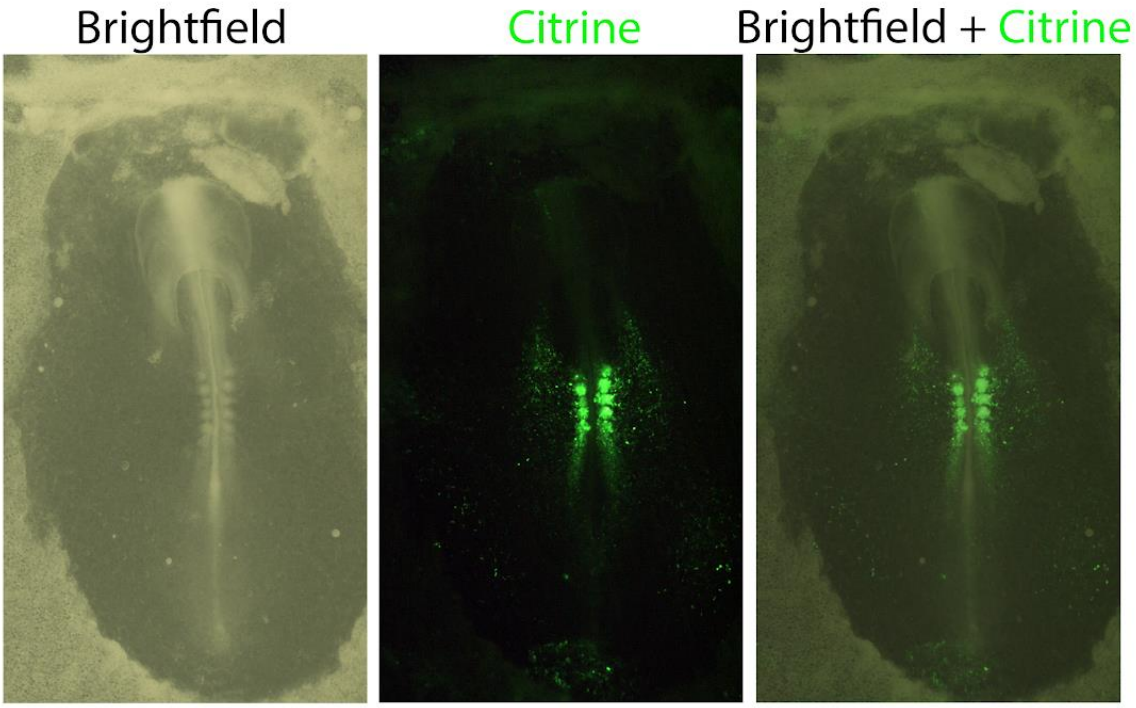

Figure 3.9: A HH10 embryo (ventral side) after injection and electroporation at HH4 with the Meox1 enhancer region previously described by the Münsterberg lab, showing reporter activity in the somites.

3.3.2 Foxo1 Enhancer

For the candidate regulatory elements associated with genes involved in somite patterning, only one of the putative enhancer regions cloned showed fluorescent reporter activity. This enhancer region was associated with the Foxo1 gene and is enh145 (see figure 3.6). Fluorescence was detected in the blood islands and pharyngeal mesoderm of a stage 10 embryo. This result was identified in three separate embryos form white eggs, and one from a brown egg. These embryos showed a different signal with the same fluorescent reporter, with signal in the inter-somitic blood vessels and at a later stage in the dorsal aorta (fig.3.10 and 3.11).

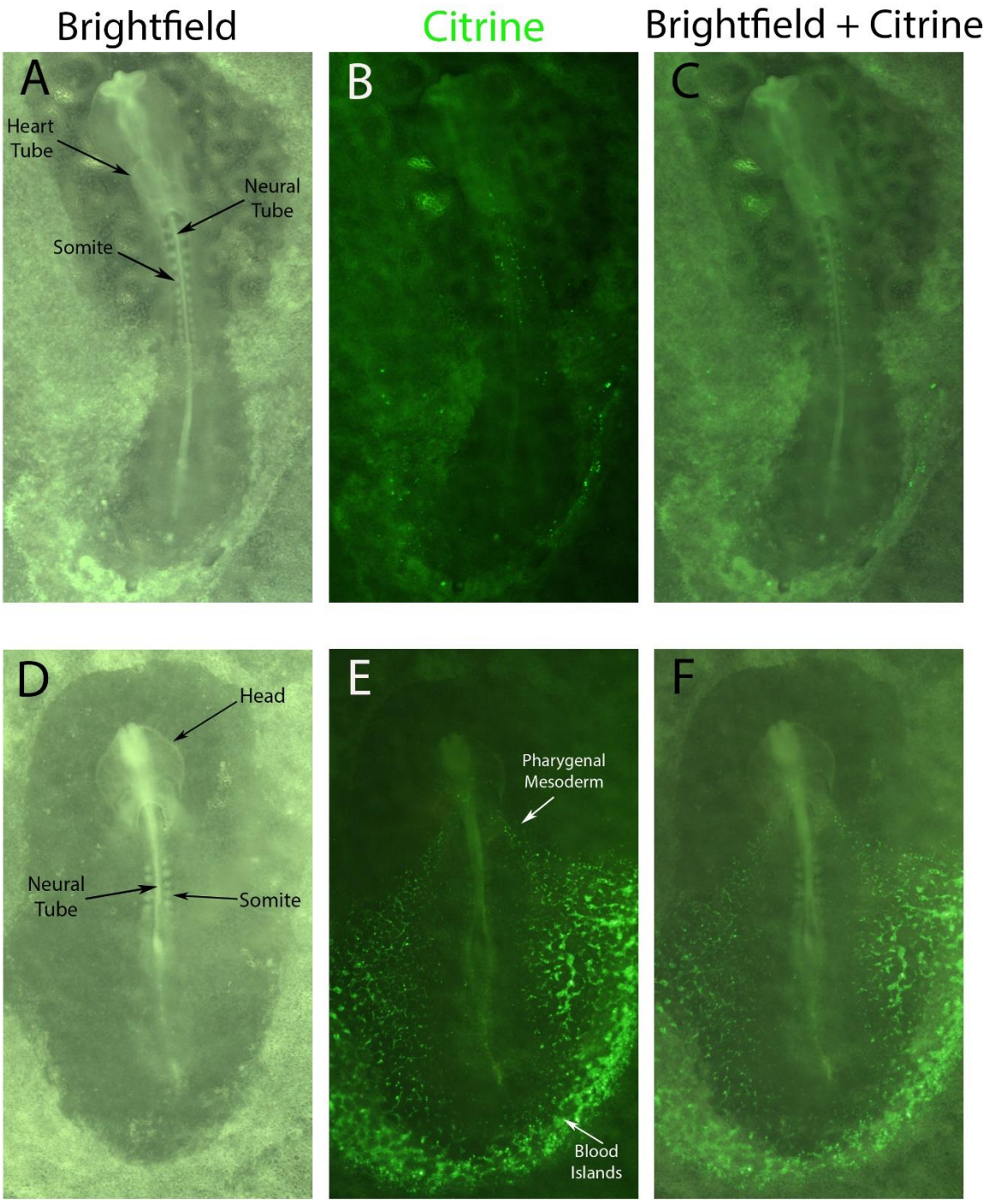

Figure 3.10: A HH10 embryo displaying fluorescent reporter activity with the putative Foxo1 enhancer reporter construct. Fluorescent reporter activity is shown in the blood islands and pharyngeal arches. Different reporter activity was visualised in brown eggs (ABC) and white eggs (DEF). A, D) Brightfield image B,E Citrine image and C,F) a combination of both brightfield and citrine images.

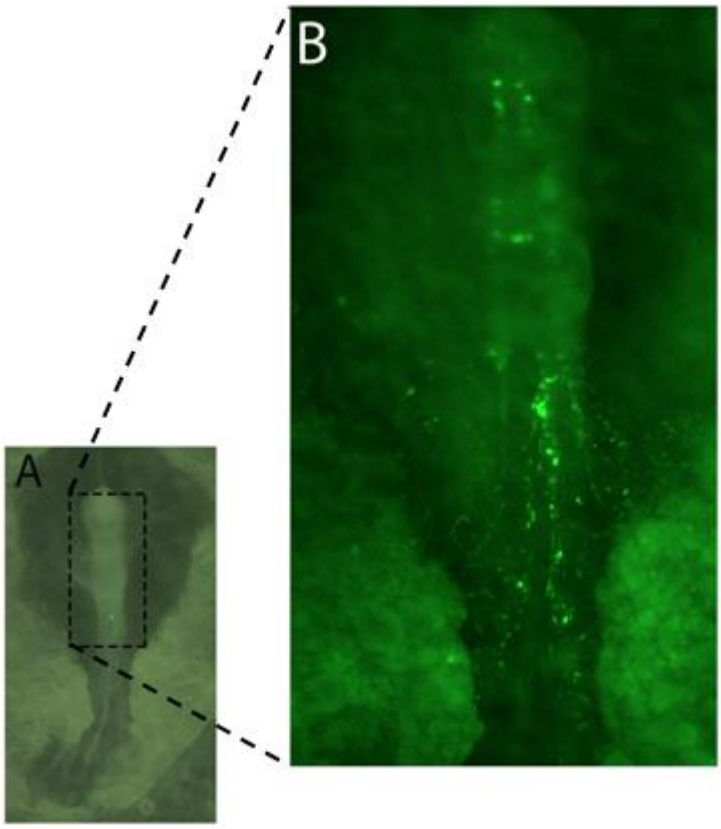

Figure 3.11: HH10 embryo ventral side cultivated from a brown egg and injected with the foxo-1 enhancer reporter construct. A) Citrine and brightfield image of the embryo at HH10. B) A close up of A, showing signal in the inter-somitic vessels, and the dorsal aorta.

3.4 Foxo-1 Sequence conservation

The sequence comparison of the Foxo1 chick cDNA indicated that most regions of the Foxo1 gene is conserved between human, mouse, and zebrafish (fig. 3.12). Phylogenetic tree analyses were then performed using MEGAX for the foxo1 gene in different species; chicken (Gallus gallus), duck (Anas Platyrhynchos), the green anole lizard (Anolis carolinesis), human (Homo sapiens), mouse (Mus musculus) Frog (Xenopus tropicalis), zebrafish (Danio rerio), Caenorhabditis elegans and Phallusia mammilata, a marine tunicate. Further analysis of the domains within the Foxo1 protein revealed the presence of a Forkhead DNA binding domain, a KIX binding domain and a TAD binding domain (fig. 3.14)

As the chick genome is poorly annotated, there was a possibility that the Foxo1 gene was incorrectly assigned. The cloned region of cDNA for Foxo1 DIG RNA probe could have been a segment of a different gene. To minimise this risk, the flanking regions around Foxo1 were examined and compared with other species. Genes such as *COG6*, *SLC25A1*, *MRPS31* and *DCLK1* were observed in the flanking regions of Foxo1 in the chicken genome. These genes were then searched for in other species with genomes that have been thoroughly sequenced, such as human and mouse. To obtain a more complete overview of blocks of synteny throughout evolutionary time, genomicus was used to produce a schematic (fig.3.15). Foxo1 is the central gene represented by a green block, and the other coloured blocks represent other genes that are near the Foxo1 locus in other species.

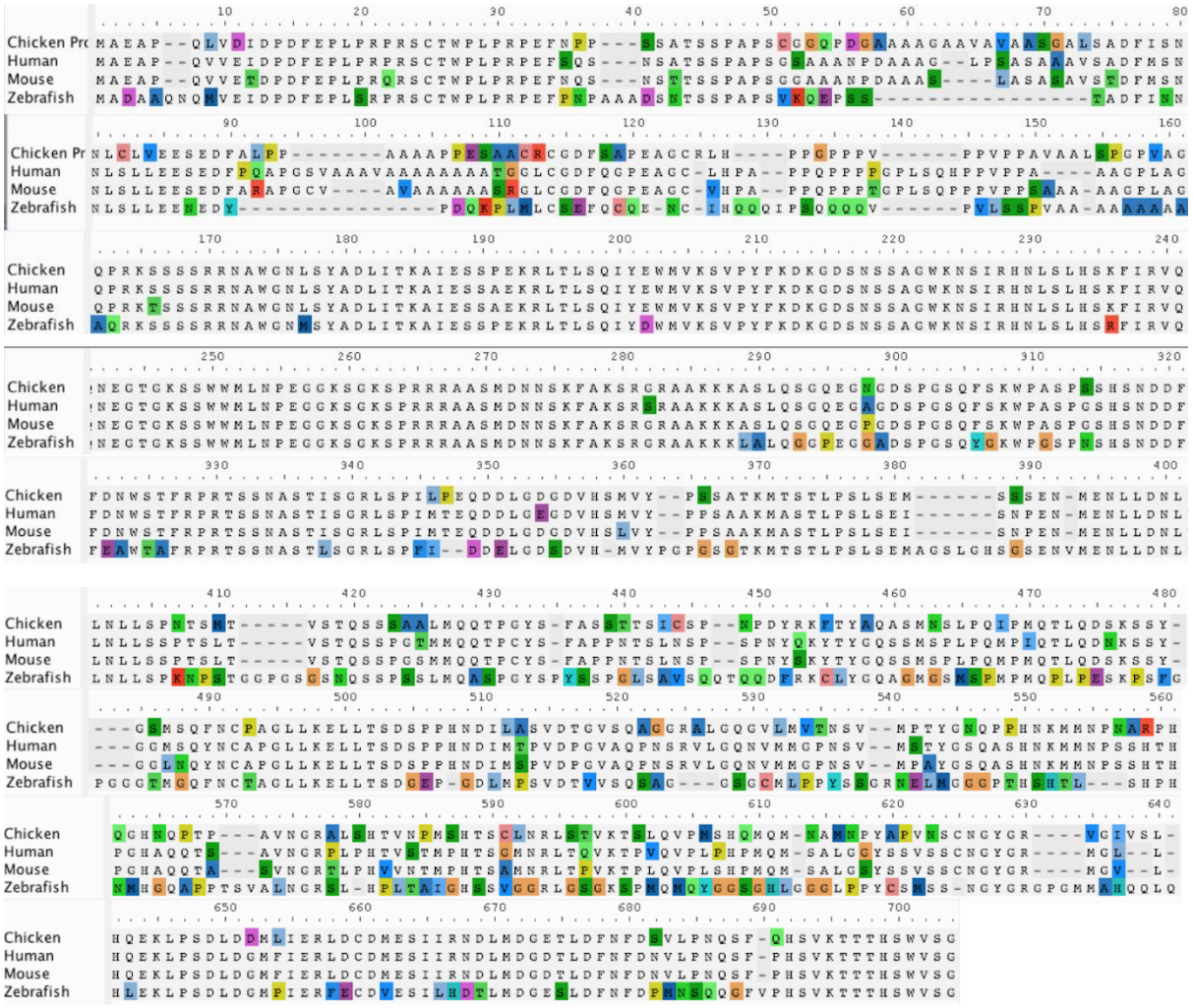

Figure 3.12: Protein alignment of the Foxo1 orthologues in the Chicken (Gallus Gallus), Human (Homo sapiens), Mouse (Mus musculus) and Zebrafish (Danio Rerio). The highlighted amino acids are different between species.

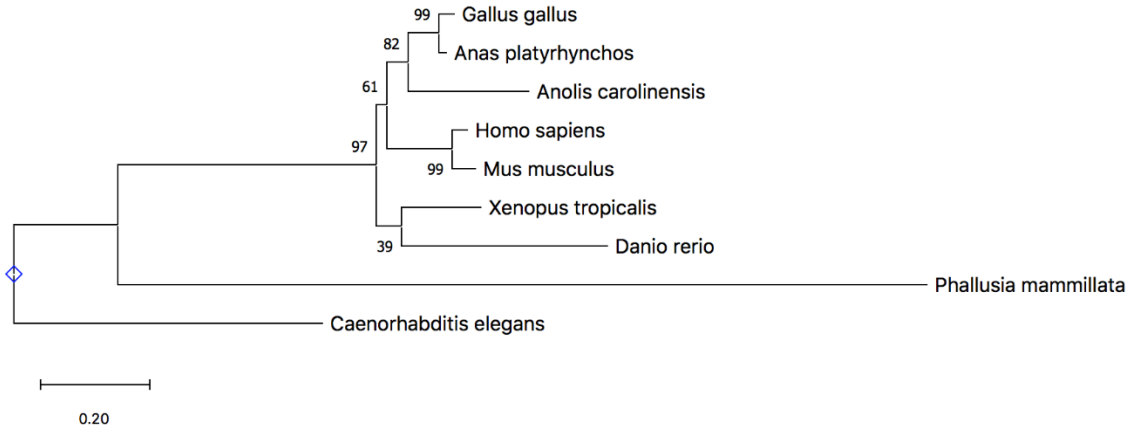

Figure 3.13: A phylogenetic tree for the Foxo1 gene in different vertebrae was inferred by using the Maximum Likelihood method and Jukes-Cantor models, displayed on a bifurcating tree. The bootstrap consensus tree inferred from 200 replicates, and the percentage of percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. Neighbour-Join and BioNJ algorithms were applied to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA X (Kumar.,2018).

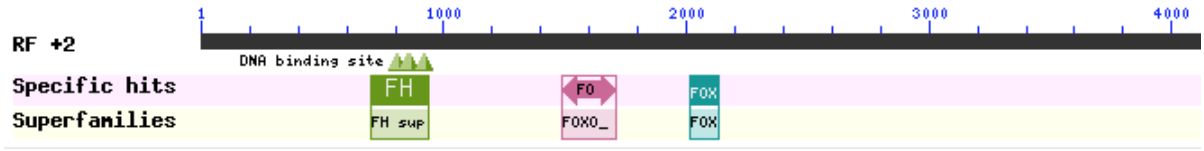

Figure 3.14: A screenshot of the conserved domains found within the FOXO1 gene, using the NCBI Conserved Domain tool. FOXO1 contains a Forkhead DNA binding domain (in green), a KIX binding domain (in pink) and a TAD binding domain (in blue).

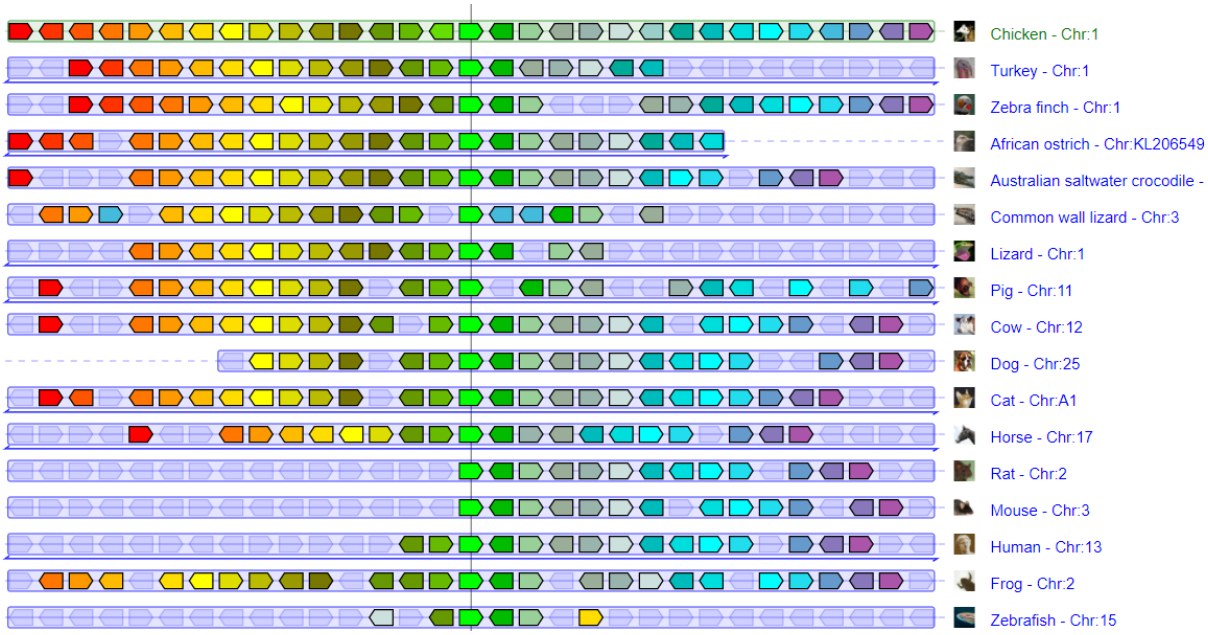

Figure 3.15: A schematic generated in Genomicus (https://www.genomicus.biologie.ens.fr/genomicus-98.01/cgi-bin/search.pl) depict regions of synteny around the Foxo1 gene within different species. Foxo1 is depicted by the light green box in the centre with a line. Other coloured blocks represent different genes. The other organisms analysed included other birds; the Turkey, Zebra finch and African ostrich. Reptiles included the Australian saltwater crocodile, the common wall lizard and lizard. Mammals included in the analysis included pug, cow, dog, car, horse, rat, mouse and human. The amphibian and fish included was frog (X. Tropicalis) and zebrafish.

3.5 NKX 2.5 enhancer

Whilst cloning constructs related to this project, one other putative enhancer region from an ongoing project was also cloned into their reporter vector, injected and electroporated using the same pipeline as discussed in the methods. This displayed a positive result; Nkx2.5 which is an early cardiac marker, displayed signal in the pharyngeal mesoderm and first heart field mesoderm in a HH10 embryo (Fig.3.16) (N=6).

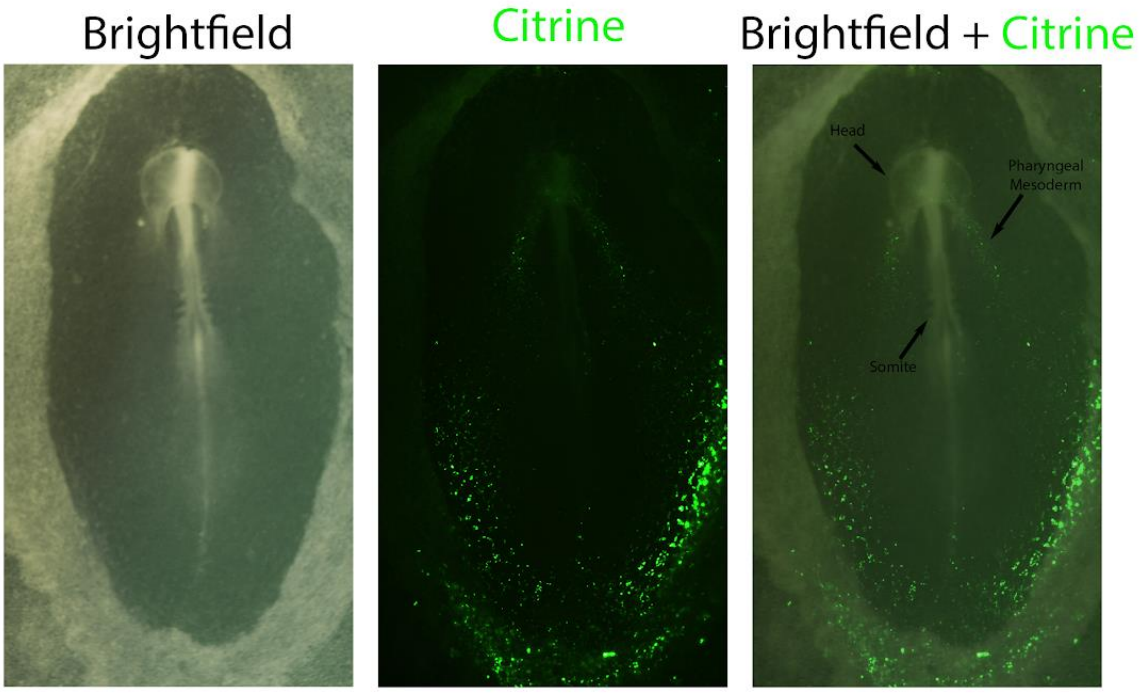

Figure 3.16: A HH15 embryo (ventral side) after injection and electroporation with a candidate Nkx2.5 enhancer reporter construct, showing fluorescent reporter activity in at the bottom of the primitive streak and in the pharyngeal mesoderm.

3.6 TOMO-sequencing whole mount in-situs

An ongoing project in the lab involves using the relatively new technique TOMO-sequencing to identify novel genes (Fruse *et al*., 2016). TOMO-seq provides genome-wide expression data with spatial information. Dr. G. Mok preformed the TOMO-sequencing protocol on a HH14 chick embryo, and sequencing was performed along the embryonic axis. The genes identified from the TOMO-sequencing dataset for further analysis by WISHs were TNFRSF14, ARIM1, AHNAK, Wnt5a, HBM, SZL, LRIG3 and NES. However, only the probes for Wnt5a, ARIM1, LRIG3 and HBM were successful in hybridising to mRNA transcript during WISH. A list of primer pairs that were used to generate the probes can be found in the supplementary information (Table 3). The Wnt5a expression pattern was specific to the tail bud of embryos, ARIM1 expression was restricted to the posterior neural tube, HMB expression was found in the inter-somitic blood vessels and dorsal aorta and LRIG3 expression was refined to the posterior epithelial somites and PSM (fig. 3.17).

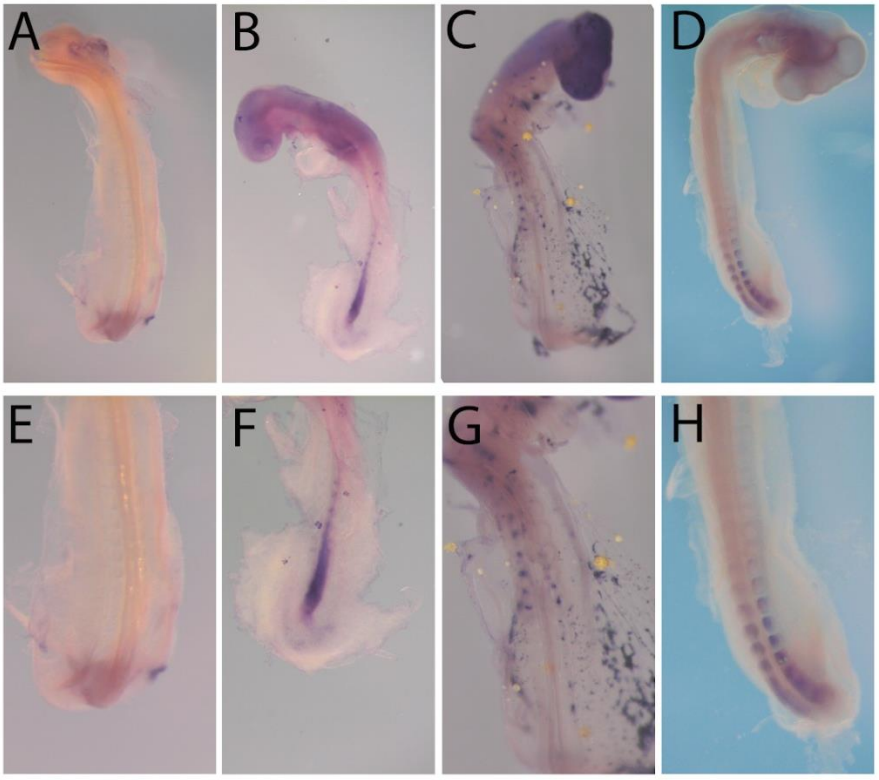

Figure 3.17: Whole mount in-situ hybridisations for genes chosen from the Tomo-sequencing dataset. A) Wnt5a, B) ARIM1, C) HBM, and D) LRIG3. E, F, G and H are all close ups of the regions of expression for Wnt51, ARIM1, HBM and LRIG3 respectively.

3.7 Single Cell RNA sequencing

To aid Dr. Gi Fay Mok on work with single cell RNA sequencing data taken from somites, other genes were chosen from a single-cell RNA-sequencing dataset, which showed differential expression in differentiated somites. The inclusive list of genes chosen, and primer pairs used to generate probes can be found in the supplementary information (Table 4). Whole-mount *in-situ* hybridisations were performed for these genes, with only RAPSN, UNC45b and LAMA2 showing any specific expression in the somites. (fig.3.18). In-situs were also performed for the gene *NPAS4l*, to aid Dr. Gi Fay Mok on work in the haematopoietic lineage (Fig.S1). The *NPAS4l* gene expression was found in the blood islands of HH4 embryos and in the dorsal aorta of HH10 embryos.

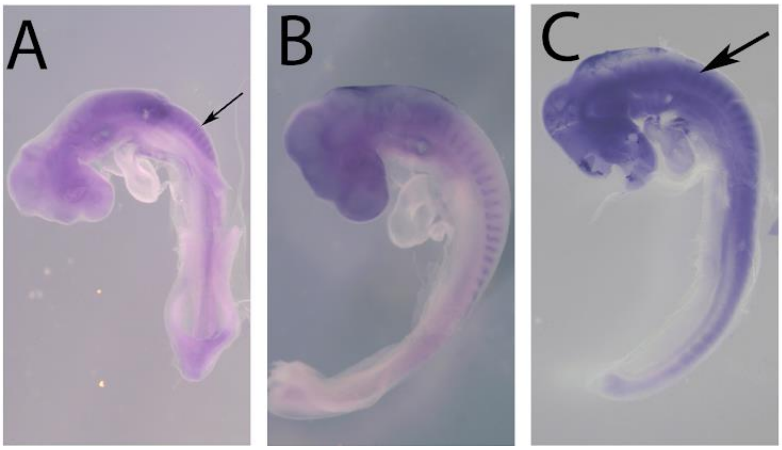

Figure 3.18: Whole mount in-situ hybridisations for genes identified by the Single Cell RNA-sequencing analysis. A) LAMA2, B) RAPSN and C) UNC45b.

4 Discussion

Previously, the Münsterberg lab has been successful in identifying enhancers associated with Meox1, Tal1, and Paraxis, using the ATAC-sequencing dataset, which was performed on somites of varying levels of differentiation (Mok *et al.*, 2021). In this study, the aim was to use this same pipeline to identify a range of enhancers associated with genes that are characteristic of most cell lineages within the somite. This study looked at 29 different putative enhancers regions, however, only one was identified as a potential enhancer, which was associated with the Foxo1 gene. *In-situ* hybridisations were also performed on genes associated with the chosen enhancer regions, and an expression profile for the Foxo-1 gene was generated.

4.1 RNA-sequencing

The RNA-sequencing of the genes stated in 1.6 indicated high levels of expression in somites. Alx1, Klhl31, UncX, and Meox2 showed similar expression profiles to genes such as Meox1 and MyoG, as they had lower levels of expression in the PSM and higher levels of expression in the fully differentiated somite. This suggests these genes are involved with muscle differentiation, as the myotome lineage is fully formed in the differentiated somite. Pax1 and Pax9, the key genetic markers of the sclerotome also showed the same expression pattern as the myotome genes. Pax1 and Pax9 are required for the proliferation of sclerotomal cells, so it is expected that these genes are highly expressed in later stage somites, as well as the expected epithelial and maturing somite stages. The genes associated with syndetome formation (Mkx, Scx and TnC) all showed ubiquitous expression throughout all stages of somite development. The exception to this is TnC, which showed slightly lower levels of expression in PSM and higher levels in the differentiated somite. However, it is interesting that these genes were present in the RNA sequencing of a HH14 embryo, as previous research has suggested that genes involved with tendon differentiation are not expressed until roughly HH18 (see section 4.11 for more details). This would suggest that cells of a syndetome lineage are perhaps specified earlier than previously thought, and further research into this could be interesting.

4.2 ATAC-sequencing

Physical access to DNA relies on chromatin accessibility, which is controlled by highly dynamic chromatin remodelling, and plays an essential role in establishing cellular identity. ATAC-sequencing was used as it identifies these regions of open chromatin. If a region of chromatin is accessible, it allows TFs to associate with their binding sites. ATAC-sequencing is therefore a good indication of if an enhancer is present in the chicken genome; however, not all open regions of DNA will be associated with an enhancer region. Of the ten regions successfully cloned into their reporter constructs, only one showed fluorescent reporter activity (see SI Table 2 for an overview of all

enhancer reporter constructs, and which were successfully generated). Whilst a region of open chromatin could be associated with an enhancer, they are also identified in other various roles, such as DNA replication (Rampakakis, *et al*., 2009), nuclear organisation (Biddie *et al*., 2010), and other regulatory regions such as insulators or silencers (Kolovos *et al*., 2012). For example, the origin recognition complex, a transcription factor that is essential for the initiation of DNA replication, has been shown to bind to regions depleted of nucleosomes (open regions of chromatin), that have no clear sequence motifs (MacAlpine *et al*., 2009). The chromatin landscape is highly dynamic, and often changes to accommodate the replication process. It is likely that the peaks selected as putative enhancers are open regions of chromatin associated with other regulatory elements or aid in chromatin dynamics, as most putative enhancer constructs were co-injected with a cherry reporter (SI Table 2), which demonstrates the efficiency of injecting and electroporating as targeted cells fluoresces red. This co-injecting shows that all the constructs were successfully injected and electroporated, however, only one enhancer construct showed green fluorescence in embryos.

Other regions of open DNA can include smaller regions that contain binding sites for chromatin remodelling complexes or CTCF. These regions can act to open nearby chromatin which could contain a regulatory element (Clapier, *et al*., 2017). Chromatin is highly dynamic and can also undergo looping to bring a region of DNA into proximity with another (see section 1.53). While a region of open DNA may not be a cis-regulatory element, it could be playing an important role in the chromatin architecture and nuclear organisation (Biddie *et al*., 2010). These regions of non-coding DNA with binding sites for chromatin remodelling complexes such as ISWI and CHD or CTCF would be accessible and could have been selected as a putative enhancer region. If this were the case, and putative enhancer regions selected were binding sites for chromatin remodelling, they would show no reporter fluorescence.

However, it is also possible that the selected peaks were enhancer regions which were not activated until much later in development. This is most likely the case with putative enhancer regions associated with tendon and cartilage development, such as Tn-c and Mkx. These genes are associated with later stages of tendon development, and transcription of these genes may not occur before HH12, where survival rates for cultured manipulated embryos drops. The RNA-sequencing suggests these genes are present in HH14 embryos, and to fully test these putative enhancer regions, it would be necessary for injected embryos to survive past HH12 until roughly HH18. One way to overcome this issue would be to perform *in-ovo* electroporation and injections, as previously described by Scaal *et al*., 2004, as well as other culture methods such as the modified Cornish Pasty (MC) method (described below in 5.1).

There was a low efficiency of cloning the enhancers regions into their constructs. Of the twenty-nine peaks selected for cloning, only ten of these were successfully cloned into the reporter construct and electroporated (See SI table 2). The failed cloning attempts were unsuccessful for several reasons. Six of these showed incorrect sequencing after cloning into their constructs. This indicates that the primers are non-specific in their binding and are amplifying regions of genomic DNA that are the similar in size to the desired regions. Eleven of these regions were not amplified from gDNA with primers; these primer pairs may have formed hairpin loops and were unable to amplify gDNA. Two were unsuccessfully transformed into the citrine reporter plasmid with multiple attempts. These low rates of cloning could potentially be solved with more care in the design of primers, checking for binding to unintended targets.

4.3 Foxo1 Expression Pattern

The Foxo1 whole mount in-situs performed characterised a novel expression pattern in chicken embryos. This project shows for the first time the expression pattern of Foxo1 throughout different stages of chicken development, from HH4 to HH25. Embryos that were used in whole mount in-situ hybridisations with a Foxo1 anti-Dig mRNA probe showed signal above the node in HH4 and then a

restricted expression pattern in the somites throughout development. Sectioning of HH14 embryos revealed expression of Foxo1 throughout the epithelial somite. In more differentiated somites an interesting pattern emerged. Somite sections showed signal in the myotome, this alternated with sections that showed signal in the sclerotome. The expression pattern observed suggested that expression is restricted to the sclerotome in the anterior part of the somite, but in the posterior region expression is refined to the myotome, and the most anterior section showed no expression of Foxo1.

During the sectioning of HH14 embryos that had undergone a whole mount *in-situ* hybridisation with the Foxo1 probe, there were sections of somite without Foxo1 mRNA expression. This would indicate a rostral-caudal patterning of Foxo-1 expression, similar to the expression patterns seen in genes such as TBX18 and UNCX, which show signal restricted to either the rostral or caudal halves of the somite (Sánchez and Sánchez, 2013). This expression pattern could indicate a role for Foxo1 in the segmentation of somites. This specific pattern has not previously been recorded in other organisms. Studies that identify Foxo1 expression in mouse embryos depict it in the blood vessels, heart, and branchial arches of younger stage embryos. In older stage embryos, Foxo1 expression is restricted to somites and the expression in blood vessels is no longer detected (Villarejo-Balcells *et al*., 2011). To confirm the Foxo1 expression across the rostral-caudal extent of somites, more sectioning of both younger and older stage embryos is required.

Foxo1 expression in the myotome is not surprising as Foxo1 plays a role in muscle differentiation. Previous studies have shown the phosphatidylinositol 3-kinase (PI3K)/Akt and Rho/ROCK signalling pathways are involved in mediating Foxo1 transcriptional activity through phosphorylation of Foxo1 to exclude the protein from the nucleus (Hribal *et al*., 2003, Matsuzaki *et al*., 2003). Other studies support the role of Foxo1 as a negative regulator of early myoblast differentiation. For example, Wu *et al*., (2008) demonstrated a regulatory loop between Foxo1 and mTOR, an important regulator in the differentiation of C2C12 myoblasts, which upregulates the expression of IGF-II (Erbay *et al*., 2003). IGF-II is a critical regulator of skeletal muscle differentiation and adult muscle regeneration (Florini *et al*., 1991) and is an upstream signal of PI3K/Akt pathways (fig.4.1) (Stitt *et al*., 2004, Jiao *et al*., 2012). Foxo1 reduces the production of IGF-II and therefore initiates a negative feedback loop of myogenesis. The IGF-II-PI3K/Akt-Foxo1-mTOR regulatory loop seems to play a critical role in skeletal muscle differentiation, and deserves further study, to increase our understanding of skeletal muscle genetic regulation.

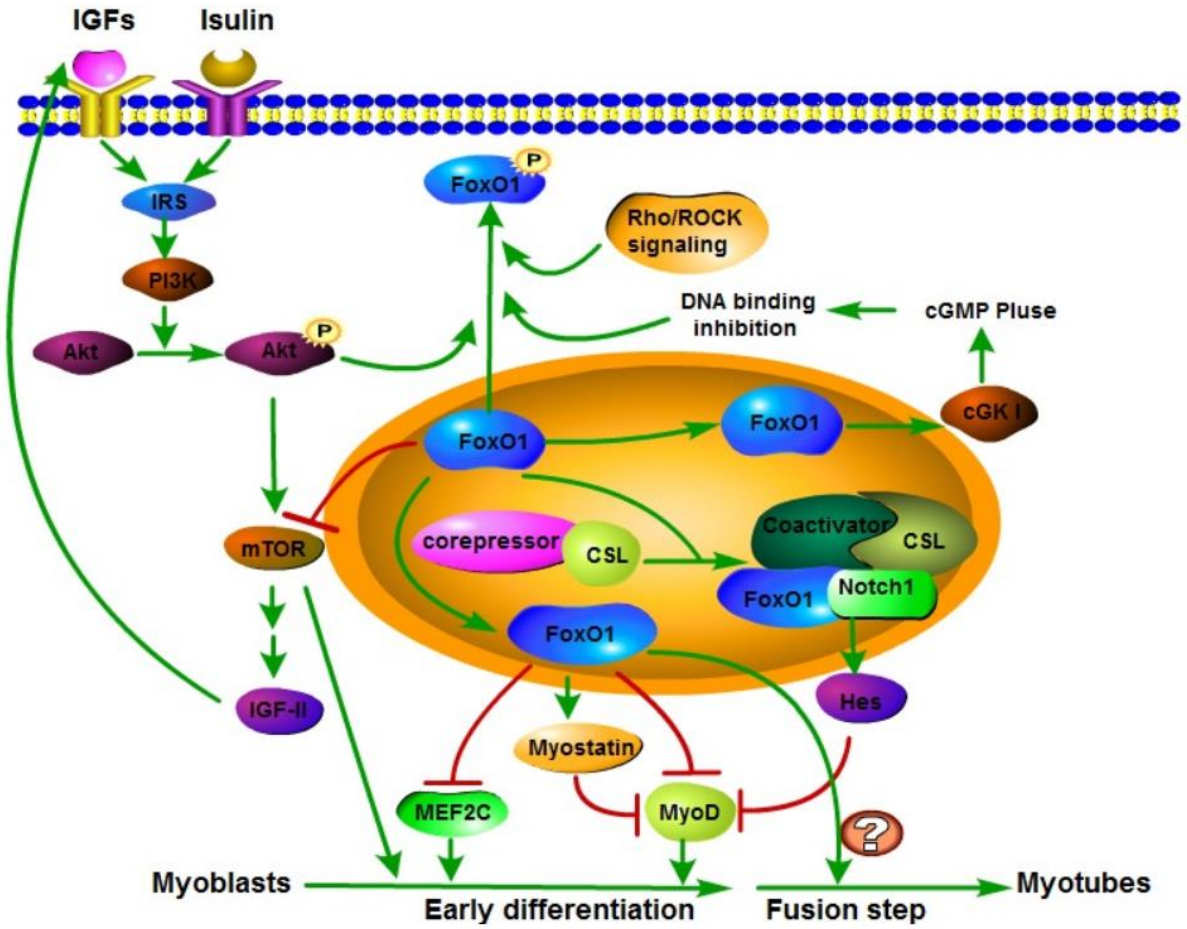

Figure 4.1: Foxo1 is regulated through retention in the cytoplasm by phosphorylation from upstream signals including IGFs and is mediated through the PI3K-Akt pathway. Downstream targets of Foxo1 include myostatin, MyoD, mTOR and MEF2C. Early myoblast differentiation is negatively regulated through inhibition of MyoD, MEF2C and mTOR, and upregulation of myostatin by Foxo1. Despite negatively regulating early myoblast differentiation, Foxo1 is required for myoblast terminal differentiation fusion into myotubes, however, the molecular mechanism regulating this is still unknown.

Foxo1 also represses myogenic differentiation through stimulation of myostatin and repression of MyoD (Allen *et al*., 2007, Liu *et al*., 2007). Myostatin has been reported as a potent negative regulator of myoblast proliferation and loss of function of myostatin leads to muscle overgrowth, as seen in Belgian Blue cattle and Texel sheep (McPherron, *et al*.,1997). Previous studies have demonstrated myostatin is a downstream signal of Foxo1 signalling, which can bind directly to the myostatin promoter region and increase myostatin mRNA expression (Allen *et al*., 2007). These studies suggest that Foxo1 could be repressing myogenic differentiation by stimulating expression of myostatin. The Foxo1 fusion proteins Pax3/Foxo1 and Pax7/Foxo1 can also interfere with *MyoD* transcriptional activity, which represses downstream targets of MyoD (Calhabeu *et al*., 2012). These two fusion proteins inhibit transcriptional activation of MyoD, and transgenic Pax3/Foxo1 in mice can disrupt normal myogenesis in somites (Finckenstein *et al*., 2006).

Foxo-1 expression in the myotome would suggest a role for Foxo1 in delaying myogenesis, preventing early myoblast specification. Whilst at these stages the myogenic differentiation programme would have already begun to produce committed cells committed to a skeletal muscle fate, the inhibitory action of Foxo1 would prevent too many cells from committing. This allows for the correct number of cells to adopt a muscle fate, without either the consequences of too many myogenic cells (hypertrophy, and potentially cancerous) or too few.

However, several other reports suggest that Foxo1 is necessary for myoblast terminal differentiation, and regulates muscle fibre type specification (Yuan *et al*., 2010). The expression pattern of Foxo1 in the stage 25 embryo could be in the primary myofibers in the somites, but further sectioning experiments would need to be performed to validate this theory.

This report also found expression of Foxo1 in the sclerotome (fig.3.5). This expression could be a way of maintaining the sclerotome as a region of cells not committed to the myogenic fate, with Foxo1 acting as a repressor of muscle differentiation. However, a recent study has described Foxo1 as a positive regulator of chondrogenic differentiation (Kurakazu *et al*., 2019), as it's upregulated by TGFβ, a growth factor that plays a key role in chondrogenic differentiation. Foxo1 expression increases with chondrogenic differentiation, while inhibition of Foxo1 results in supressed expression of collagen differentiation markers such as *Col2a1* and *Acan*, along with repressed chondrogenic differentiation (Kurakazu *et al*., 2019). Previous studies have demonstrated the binding of Foxo1 to the Sox9 promoter, increasing expression (van Gastel *et al*., 2020). As Sox9 initiates chondrogenesis (see section 1.24), Foxo1 as one of the transcription factors that initiates it, is clearly important in the sclerotome. Therefore, it is unsurprising that Foxo1 transcript is expressed in this somitic compartment at this stage.

4.4: A Novel Foxo1 Enhancer

Using the pipeline stated in section 3.7, this project was able to identify the presence of one regulatory element, associated with the Foxo1 gene (figures 3.10 and 3.11). This region is conserved in human, mouse and Chinese alligator (fig.4.2), and in human, the histone modification H327Ac has been detected nearby. When injected into HH4 embryos, the reporter construct that contains the putative regulatory region showed signal in the blood islands in HH10 chicken embryos cultured from white eggs (n=4), and in the inter-somitic vessels for chicken embryos cultured from brown eggs (n=1). In this embryo, that managed to develop to HH12, strong reporter activity can be seen in the dorsal aorta. (Fig.3.11)

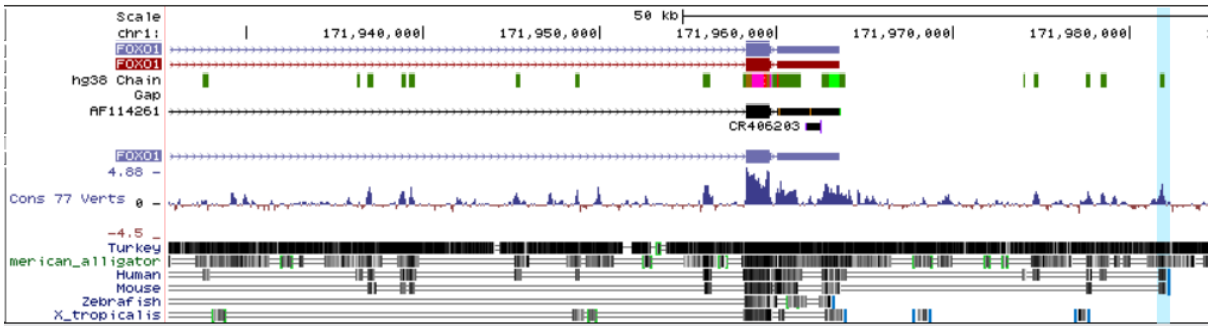

Figure 4.2: The 2018 genome browser for enhancer 145, highlighted in blue. Below are the tracks for species conservation. This putative enhancer region shows sequence conservation with turkey, American alligator, human and mouse. There is no sequence information for the zebrafish and X.tropicalis, so it is unknown if this region is conserved in these two species.

The reporter signal in the blood islands, dorsal aorta, and inter-somitic blood islands suggests that the cis regulatory element is active in endothelial cells and thus could indicate a role for Foxo1 in blood vessel development. Previous studies have indicated a crucial role of Foxo1 in blood vessel development (Furuyama *et al*., 2004, Park *et al*., 2009), with Foxo1 expression identified in the vascular system at early stages of development, in mouse whole mount in-situ hybridisation experiments. However, from E.11.5 vascular expression is no longer detected in mouse embryos. Instead, the gene is expressed in the somites; this could be showing Foxo1 playing a role in the fusion of myoblasts and fibre type specification (Villarejo-Balcells, *et al*., 2011). It has been shown that Foxo1 regulates the morphological response of endothelial cells to critical genes involved in angiogenesis, such as VEGF (Vascular endothelial growth factor) and angiopoietin-1 (Daly, 2004, Zhuang *et al*., 2013). Mice with Foxo1 knockout are embryonic lethal due to defects in the

development of brachial arches and malformation in major vessels of the embryo and yolk sac, potentially due to an improper response to VEGF, which is mediated by Foxo1. Younger embryos show severe disruption of blood vessel formation (Furuyama *et al*., 2004, Hosaka *et al*., 2004). Therefore, it is not surprising that an enhancer region associated with Foxo1 would be activating Foxo1 in cells of an endothelial lineage. What is surprising is the discrepancy between the enhancer reporter pattern and the in-situ pattern; this is discussed further below.

4.5 Differences in Foxo1 Expression Patterns and Putative Enhancer Patterns:

Foxo1 as a transcription factor plays a variety of roles, as discussed above. It has been shown to activate enhancers for separate genes, which gives it its diverse functionality. Here, Foxo1 was investigated due to its role in skeletal muscle development, as high expression levels were found in somites of developing chick embryos. Several candidate enhancers for the Foxo1 gene were cloned into fluorescent reporter constructs, one of which displayed reporter activity in the blood islands, inter-somitic vessels and the dorsal aorta. However, this expression profile was different to its corresponding whole mount *in-situ* hybridisation, which detected Foxo1 expression in developing somites (figures 3.1-3.4). There are several reasons why there is this unexpected discrepancy. It is possible that the anti-DIG RNA probe generated for Foxo-1 and used for *in-situ* hybridisation was complementary to a specific isoform of the gene, that is only produced in somites. It is also possible that the enhancer cloned was only specific to activating Foxo1 gene expression in blood vessels and endothelial cells. These are further discussed below.

4.6 Protein Isoforms

Protein isoforms are sets of similar proteins that originate from a single gene. While some isoforms are functionally redundant, several are reported to have unique biological functions (Breitbart, *et al*., 1987, Karlsson and Linnarsson, 2017). Isoforms can be formed from the process of alternative splicing, post translational modifications, or variable promoter usage (gene isoforms) (for a review see Hallegger *et al*., 2010 and Stamm *et al*., 2005). Isoforms explain the discrepancy between the large number of proteins present in an organism, and the surprisingly small number of protein coding genes; each gene has the potential to code for several isoforms, each with a functionally diverse role (Breitbart *et al*., 1987). Alternative RNA splicing is a post transcriptional modification, where the final mature mRNA transcript includes or excludes different exons. It was first believed that protein isoforms have similar functions to other isoforms from the same gene, whereby they work in the same regulatory systems, with slightly differing roles. However, recent work has shown that most protein isoforms share less than 50% of their interactions, behaving like separate proteins rather than slight variants of each other (Häfner, 2019). Some may exhibit dominant-negative effects over the other isoforms encoded by the same gene, up and down regulate other isoform family members. For example, there are two isoforms encoded by the *BCL2L1* gene. The longer isoform is involved in inhibiting apoptosis, whereas the shorter isoform actively promotes it (Schwerk and Schulze-Osthoff, 2005). Overall, there are several hundred human genes have been documented to produced functionally diverse transcript variants (Kelemen *et al*., 2013).

There are other transcriptional methods that produce different protein isoforms. Gene bodies can contain several different promoters within them, resulting in different transcriptional start sites throughout the gene producing different length mRNA transcripts (Wang *et al*., 2016). One recent study found that alternative transcription start sites and termination sites are the main cause of isoform diversity in human tissues and produced a large dataset of isoforms that are generated in this fashion (Reyes and Huber, 2017). The Foxo1 gene could contain other promoters within its gene loci, which could interact with different CREs and promote transcription of different parts of the Foxo1 gene, producing differing mRNA transcripts. These could have varying roles; one could play a part in the formation of blood vessels and the other could play a role in regulating somite patterning and differentiation, including myogenesis and chondrogenesis.

Programmes such as refseq and enembl genome browsers depict Foxo1 as having one isoform in the chick genome. However, the chick genome is poorly annotated, and this may not be correct, as other species have multiple transcript isoforms associated with the Foxo1 gene. The region of the Foxo1 gene cloned for probe synthesis is shown below in figure 4.3, depicted by arrow A. This is the second and final coding exon in the chick Foxo1 gene and shares many domains with the first coding exon. However, exon 2 contains some domains that are unique to it, such as the KIX binding domain and the transactivation domain. The transactivation domain is a binding domain that mediates the association of Foxo1 with the co-activator CBP/p300. The KIX binding domain binds to the CREB protein, and both domains simultaneously bind the KIX domain of CBP/p300 (Wang *et al*., 2012). It is feasible that this exon of the Foxo1 gene could be responsible for activity in only somites. Perhaps the function of Foxo1 in somites is related to the binding of these two proteins, which both play a role in regulating the transcription of genes.

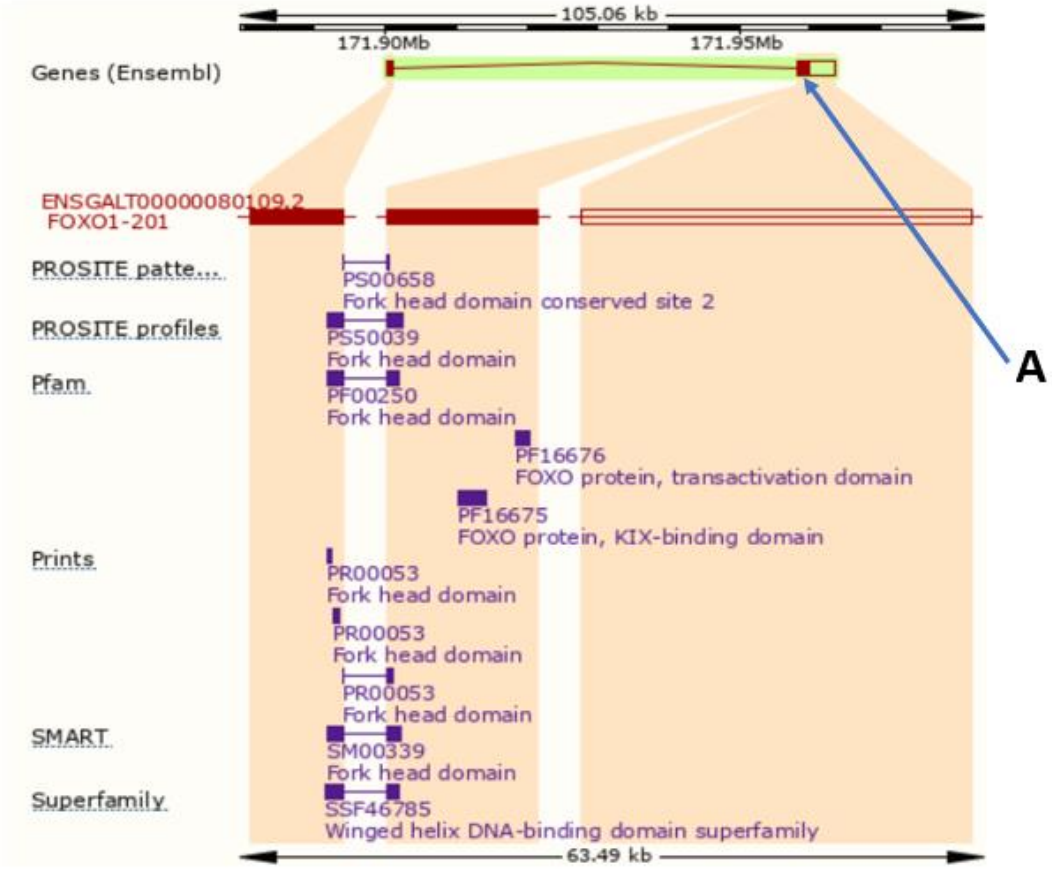

Figure 4.3: The protein domains found within the three exons of the gallus gallus Foxo1 gene. Arrow A depicts the second coding exon, which was the amplified and cloned region used to generate the Foxo1 probe. Adapted from ensembl. (https://www.ensembl.org/index.html

4.7 Enhancer Control

It is well documented that several enhancers can interact with one gene, regulating different spatial and temporal activation of a gene (Hong *et al*., 2008, and see section 1.53 for more information). Enhancer construct 145 could activate endogenous Foxo1 expression in cells of an endothelial lineage only. This enhancer could contain binding sites for transcription factors that are found only in cells committed to an endothelial fate, resulting in enhancer and gene activation specific to these cells. Perhaps this enhancer communicates with a secondary promoter (as discussed above) to only produce a certain splice variant of the Foxo1 gene, with an as yet to be identified separate enhancer associated with the Foxo1 gene that activates expression in the somites.

4.8 Gene Synteny

The results of the gene synteny showed regions of synteny especially with other birds and lizards. These species are quite closely related, so the genomic region of Foxo1 has not been subjected to much evolutionary change within these species. Even in mammals, that are not as closely related to birds, blocks of synteny can still be identified, again meaning the genomic region of Foxo1 is conserved between species. The only species where this is not true are the mouse and rat, where the 3' region to Foxo1 has no blocks of synteny. Potentially, this region of DNA was inserted at a new genomic location during evolution. There is also little gene synteny around the zebrafish Foxo1 gene, a species evolutionarily far from the chicken. However, this gene synteny analysis indicates that it is highly unlikely Foxo1 was mis-annotated in the chick genome. Therefore, from the gene synteny analysis, it is fairly certain that Foxo1 mRNA giving the strong expression pattern in the somites.

Another possibility is that the Foxo1 RNA probe is binding to mRNA transcripts of the structurally similar protein Foxo3. Foxo3 shows specific expression in the somites in *X.tropicalis* (Zheng *et al*., 2020). When sequenced, it was confirmed that Exon 2 of Foxo1 was cloned and amplified, and therefore the mRNA probe was Foxo-1 specific. However, the two mRNA sequences could be similar enough that this probe also bound to mRNA produced by Foxo3. The percentage similarity between the Foxo1 gene and the Foxo3 gene is only 45.5% (EMBOSS Water analysis). The actual region of Foxo1 cloned shows 51.7% identity with a region of the Foxo3 gene. The start of the Foxo1 probe region is very similar to a section of the Foxo3 cDNA, with only a few base pairs and missense regions (see 1.1 of appendix). It is possible that the Foxo3 mRNA produced by this region is similar enough to the RNA probe that it could bind and display signal. This possibility needs further investigation.

4.9 Foxo1 Sequence Conservation and Phylogeny

The sequence comparison of the Foxo1 gene between chick, human, mouse, and zebrafish indicated that most regions of the Foxo1 gene is conserved between (fig. 3.12). This suggests that Foxo1 is a functionally important gene, as there are few mutations within the gene between species, especially in the sequencing coding the protein domains. This suggests that any changes within the Foxo1 protein usually results in non-viable offspring. There are several other housekeeping genes that contain minimal mutations due to their importance in development, such as the HOX family of genes, which set the body plan of an organism in early embryonic development (Hrycaj and Wellik, 2016).

The phylogenetic gene tree analyses represents the evolutionary history of the Foxo1 gene and its orthologue (daf-16). This tree was rooted to C.elegans, as a member of the more ancient clade Protostomia. The rest of this gene tree is as expected, and similar to the standard phylogeny of all species (Sanderson., 2008). There is a divergence event, separating the truncate P.mammillata as a sister phylum to the rest of the Foxo-1 proteins from the other vertebrae species. This is expected as the deuterostome and protostome species are less closely related, and even functionally important housekeeping genes would be genetically more separate between these two subphyla. The phylogenetic tree then groups together all the vertebrates as a clade, with the zebrafish as the most ancient member of this group. The tree follows the expected format for vertebrates; the amniotes (birds, mammals, and lizards) all share the same common ancestor. The only slight difference is the amphibian X.tropicalis is a placed in a sister group to the zebra fish, suggesting they share a common ancestor, when normally amphibians share a more ancient common ancestor with amniotes.

Evolutionarily, Foxo1 is clearly an important gene, having orthologous (daf-16) in Turnicates and C.elegans, suggesting this gene is functionally ancient and is required throughout the animalia kingdom. This would suggest the Foxo1 protein is vital for a function that is required in less complex organisms, suggesting it has a fundamental role. This Foxo1 phylogenetic gene tree has a very similar format to the standard phylogenetic tree of life, with only the difference in the positioning of the frog species. This could be because Foxo1 is involved in the development of blood vessels, it may play a slightly different role in the amniotes. Fish and frogs have different circulatory systems than other

chordates (Heisler *et al*., 1999), and so during embryonic development, Foxo1 may contribute to the difference in the development of the vascular system. However, when looking at the sequence conservation, these differences are not major in any of the key domain regions of Foxo1, and the tree is probably diverging these species based on the non-protein domains of this gene, which does not have as important a role in the final protein produced.

4.10 NKX2.5 Enhancer

During this project, an enhancer reporter construct associated with the Nkx2.5 gene displayed fluorescent signal in cells of the pharyngeal mesoderm and arches, as well as cardiac progenitor cells. Nkx2.5 is homeobox transcription factor that is expressed in early cardiac mesoderm and cardiac progenitor cells (Jamali *et al*., 2001) and is used as a marker for cells involved with heart formation. Conditional knockout of Nkx2.5 in mice display hearts malformations, and heterozygous knockouts show morphological and functional abnormalities in adult mice (Furtado *et al.,* 2016). The potential enhancer associated with Nkx2.5 is found in the same cells as Nkx2.5, and it would be interesting to perform downstream analysis, including footprint analyses and CRISPR knockout experiments, to further validate and characterise this enhancer.

4.11 Tendon development and the role of Scleraxis

Scleraxis is a gene described as being at the top of the differentiation pathway for tendons. It was therefore of interest to us to find an enhancer for this gene to further understand differentiation of the syndetome. Many previous studies have detected *Scleraxis* mRNA in the limb bud stages of chicken embryos using whole mount *in-situ* hybridisation techniques (Bret and Tabin, 2004, Perez *et al*., 2003, Asou *et al*., 2002) The earliest stage of *Scleraxis* detection was in HH16, from previous research in the Münsterberg lab. However, the analysed RNA-sequencing data showed high levels of scleraxis expression in HH14 embryos (see figure 3.7). This research was able to capture expression of *scleraxis* in a HH14 embryo during whole-mount in-situ hybridisations (fig.3.5). It was previously thought that *Scleraxis* was not induced in embryos until later stages, with uncommitted tendon progenitors remaining in the syndetome until scleraxis induces the tendon differentiation pathway. Our finding demonstrates that tendon progenitors are specified earlier than previously thought. This may have implications for the rest of the tendon specification programme, with genes like *Mkx* and *Tendomoldin* also being expressed earlier, potentially producing tendon specific cells. However, *Scleraxis* and *Mkx* protein in HH14 embryos could be waiting in tendon progenitor cells for the right developmental time point, where it along with other TFs binds to downstream target enhancers or promoter regions, and then activate expression of genes which continue the tendon differentiation programme. The tendon differentiation regulatory pathways need to be revised to understand why scleraxis transcripts are in the embryo as early as HH14 when it has previously been reported that tenocytes do not differentiate until tendon develops at HH20 (Edom-Vovard *et al*., 2002).

One approach to better characterise spatially and temporally tendon progenitors arise would be to find the enhancer controlling this gene. With this reporter construct, it would be possible to perform a time-lapse microscopy experiment using GFP as a visualiser for cells containing transcribed scleraxis protein. This project did attempt to find an enhancer region associated with *Scleraxis*, however only one putative enhancer construct was found from the ATAC-sequencing data set. The Münsterberg lab have mapped the ATAC-sequencing data to the UCSC 2015 chick genome browser, where *Scleraxis* is poorly annotated. When mapped to the chick genome, it did not align with any chromosome, and when viewed in UCSC chicken genome browser 2015 there is only a small flanking region around the gene body. However, in the updated chicken genome (2018 version) scleraxis fully maps to chromosome 2. If the ATAC-sequencing could be properly mapped to the genome for this region, more peaks may be visible in the flanking regions, making a more successful selection of a putative enhancer region possible for this gene.

4.12 TOMO-sequencing and Single Cell Sequencing expression patterns

The TOMO-sequencing project utilised new sequencing approaches to identify novel genes found along the axis of the embryo. The genes *Wnt5a,* Nestin and HBM were known markers for dorsal mesoderm, neural crest cells and blood islands respectively, and both wnt5a and HBM in-situs show this expression pattern clearly. The genes TNFRSF14, ARIH1, AHNAK, LRIG3 and SZL were not associated with early embryonic development in the literature, with no expression data available in early embryos. Whilst the probes from TNFRSF14, AHNAK and SZL showed no specific activity, LRIG3 in-situs showed expression in the PSM and epithelial somites, and ARIM1 expression was restricted to the dorsal neural tube. The TOMO-sequencing identified high levels of these genes' expression in the corresponding regions, proving that the TOMO-sequencing does act as a novel method to identify spatial patterns of gene expression.

5. Future Work

5.1 Improvements to culture and embryo transfection methods

The chick model was ideal for this study, as the Easy Culture method of *ex-ovo* manipulation is relatively cheap and easy. Although the chicken embryo is useful for studying early embryonic development, embryos rarely survive past HH14 when using easy culture methods. (Chapmen *at al*., 2001). This issue could be tackled by implementing the modified Cornish pasty (MC) culture method, which enables *ex-ovo* growth of chicken embryos until HH18 (Nagai *et al*., 2011). If this could be implemented after injection and electroporation, it would potentially allow manipulated embryos to survive longer, thus allowing reporter construct fluorescence to be seen where previously embryos may not have survived long enough for activation of enhancer elements. This could be particularly important with putative enhancer elements associated with tendon development, as these genes are mainly expressed at the later stages of development.

Chicken embryos are highly sensitive when removed from their egg, and in more extreme weather conditions, with outside temperatures are either too cold or too hot, embryo quality seemed to decline. This is to be expected, however, removing embryos from their eggs would then further hamper their survivability. One way to maybe increase long-term survivability in the winter and summer months would be to implement the *in-ovo* electroporation technique (Scaal *et al*., 2004). Removal of an embryo from its environment using easy culture methods results in a drop in embryos survival, as this method cannot support the vascular development that surrounds the yolk and supplies the developing embryo with nutrients. However, *in-ovo* electroporation techniques allow the embryo to remain in the original environment, minimising the risk of the embryo dehydration when exposed to air upon removal for easy culture. This method also supplies the embryo with it all original supplements from the yolk, which would allow growth of an embryo to later stages, unlike easy culture methods where embryos are restricted for nutrient availability and space. If this method could be fully utilised in the lab, survival rates of embryos may increase, along with age of the surviving embryos. This could allow for identification of putative enhancer elements that are activated in later stages of development. A disadvantage is that there are no established protocols to electroporate gastrula stage embryos in ovo. Instead, somite-stage embryos from HH10 onwards would have to be injected and targeted, for example in the neural tube or epithelial somites. It may also be possible to use lipofection to deliver the enhancer-reporter construct into embryonic tissues *in-ovo*.

5.2 Foxo1 Isoforms

Usually, the same cells will express both the enhancer reporter activity and mRNA expression. This is because the enhancer activates expression of the gene, which then produces mRNA, which is detected by whole mount *in-situ* hybridisations. This study reported the unusual discordance between the Foxo1 expression pattern in the somites and the putative enhancer construct in the blood islands and dorsal aorta. This could be due to probe specificity to one isoform, as described in section 5.6. To

test this, probes should be designed capturing the other exon of Foxo1, to determine if this region displays similarities to the Foxo1 putative enhancer expression patterns.

5.3 Foxo1 Enhancers

As discussed above (5.6), the enhancer found close to the Foxo1 gene is responsible for activating expression in cells of an endothelial lineage. However, there could be another unknown enhancer acting upon the promoter region of Foxo1, which is responsible for activation in the somites. It would be worth revisiting the ATAC-sequencing dataset and identifying other regions of open chromatin that are near to Foxo1, with the aim of identifying a second enhancer that is involved with activating expression in the somites. However, during this project, six peaks were identified near the Foxo1 gene, with only one showing reporter activity (Fig.3.6). There are other peaks that could be investigated further, however, it is possible that this specific enhancer is located further away, and works over many base pairs, by-passing other genes.

5.4 Further Characterisation and ATAC-Foot printing analysis.

When transcription factors are bound to a region of DNA, it prevents the Tn5 enzyme binding and adding sequencing adaptors. This means these regions are not sequenced. These transcription factor bound regions leave a 'footprint' in the DNA; regions throughout the ATAC-sequencing peaks where the peak becomes a dip with two flanking shoulders. This has been exploited to identify candidate binding sites throughout the genome (Hesselberth *et al*., 2009) and has been advantageous in the systemic analysis of protein-DNA interactions, further characterising cis-regulatory elements (Sung *et al*., 2016). For this project, further characterisation of the Foxo1 enhancer using foot-printing methods would be interesting, especially to determine the presence of an endothelial-specific DNA-binding element that modulates the tissue specific activation of Foxo1 in endothelial cells. HINT-ATAC (Li *et al*., 2019) could be utilised in the lab to find differential TF footprints in regions of open chromatin as it significantly improves the recovery of footprints.

5.5 Enhancer Discovery Pipeline

More work needs to be done to streamline this pipeline. The ATAC-sequencing data set, whilst useful in depicting regions of open chromatin that potentially have functional significance, cannot alone determine the presence of an enhancer. Using more of the data available from the ENCODE data base, such as methylation and histone acetylation marks, may help to identify putative enhancer regions. Marks such as H3K27Ac are present on active enhancers, and while there are no available genome wide datasets available for Histone marks in the chick, they are readily available in the human. This coupled with searching for conserved regions should enable more accurate selection of candidate enhancer peaks. Previous papers have performed ATAC-sequencing on samples removed from the neural crest. It would be interesting to compare the ATAC sequencing from neural crest samples with the ATAC-sequencing that the Münsterberg lab generated from somites. This might reveal differential peaks between the two samples, indicating regions that are open in somites and closed in neural crest. This might help identify open regions of chromatin involved specifically in somite development, therefore narrowing down the selection of possible peaks that are associated with enhancer regions.

5.6 CRISPR Cas9 Experiments

The recent development of CRISPR-Cas9 (Clustered, regularly interspaced, short palindromic repeats (CRISPR) and the CRISPR- associated protein 9 (Cas9)) gives the opportunity for genome wide targeting (Cho *et al*., 2013). Until recently, its main use was targeted towards protein coding genes, but recently other studies have been successful in using CRISPR-Cas9 systems to identify enhancer elements in non-coding DNA (Korkmaz *et al*., 2016). This system has been utilised by the Sauka-Spengler lab (University of Oxford), who have recently developed a CRISPR-Cas9 'toolkit' for *in-vivo* modulation of cis-regulatory elements (Williams *et al.*,2018). This study was the first reported method of epigenome engineering in a developing chick embryo and was able to silence whole

enhancer regions. This toolkit has been used recently in the lab (Mok *et al.*, 2021), to functionally characterise the importance of the TCF15 and Meox1 cis-regulatory elements. There is potential that this approach to further characterise enhancers could be utilised in the lab.

Targeted deletions of transcription factor binding sites found using foot-printing data could be used to further characterise any enhancers discovered. Specifically targeting TF binding sites within a CRE with CRISPR-Cas9 would identify any key TFs that are required for activation of an enhancer element. This approach could be utilised here to further characterise the Foxo1 enhancer discovered, to determine if a specific TF is required for activation of this enhancer in cells of an endothelial lineage.

Previous studies have used a CRISPR-Cas9 genetic screen for enhancers within the human genome, by targeting TF binding sites in enhancer regions (Korkmaz *et al*., 2016). However, whilst this may work for some enhancer regions, and could be useful if incorporated, it is impossible to capture all enhancer regions with this method. This is because, as previously mentioned before, there is no underlying TF binding site that is mandatory for the function of all enhancer regions.

CRISPR-Cas9 could also be used here to determine the role of Foxo1 in rostral-caudal patterning. As this expression pattern has not been documented before, using CRISPR to knockout the Foxo1 gene would give an indication of its role. Whilst knockout experiments have already been performed for Foxo1 in mouse (Furuyama *et al*., 2004), the only aspect of development observed was the phenotypic result (abnormal angiogenesis). However, it would of interest to see if Foxo1 plays any role in the rostral-caudal patterning of somites. For this, Foxo1 expression could be silenced using CRISPR-cas9 systems and then visualising the expression of genes known to be involved in the rostral-caudal patterning of somites, such as *UNCX* and *TBX18*, using real-time qPCR, or whole mount *in-situ* hybridisations. If expression of these genes differs to wild type expression, then Foxo1 potentially plays a role in this patterning system. If not, then Foxo1 could be acting downstream of these genes, and it would be interesting to see if any cis-regulatory regions for Foxo1 contained binding sites for genes known to be involved with rostral-caudal patterning, many of which are transcription factors.

6. Concluding Remarks

Enhancers play a fundamental role in regulating the precise spatial and temporal gene expression patterns during embryonic development and differentiation. Many labs have focused their efforts on understanding how enhancers function, and in particular, how to identify which cis regulatory elements are involved in the control of a specific gene. There are many problems with attempting to identify specific enhancers, as there is no underlying sequence feature that discerns an enhancer from other non-coding regions of DNA. This project used ATAC-Sequencing data, which depicts regions of open chromatin to locate putative enhancer regions. The results discussed here show that even with the ATAC-sequencing tool, discovering functionally relevant CREs is difficult. The enhancer pipeline developed by the Münsterberg lab would benefit from both further bioinformatic and other functional data to find exact regions of open chromatin that correlate with an enhancer.

The continuation of this project will aim to characterise more cis-regulatory elements, and further our understanding of their involvement in gene regulatory networks involved with somite development throughout embryogenesis. It is critical that we understand how these non-coding regions of DNA work in biological systems, to understand diseases associated with them and to potentially develop therapies that target these regions.

7: Acknowledgements

I would like to thank Professor Münsterberg and Dr. Geoffrey (Gi Fay) Mok for all their mentorship and guidance with this project, and for constantly providing me with support and ideas. I would also like to thank Shannon Weldon for demonstrating nearly all the lab techniques I have used for this

project, and her encouragement throughout, and to everyone in the Wheeler and Grocott labs for helping me find my way round the lab and with troubleshooting and made my masters project enjoyable and entertaining.

8: Supplementary Information:

SI Table 1: Primer pairs for amplification of cDNA of genes picked from the RNA-sequencing performed on somites.

Gene Name	Forward Primer	Reverse Primer	Successfully made into a probe?
MKX	GTCGTTAAAAACAATTTGTTTCCCTCC	TTAAAACTGCTGGACTAATGGCA	No
Foxo1	GAAATGGTGACAGCCCTGGA	CTCCCATATGGTCGACCTGC	Yes
SCX	GGCCGTTACCTGTACCCC	TAATTATACGAGCTGCTCAGGCT	No
TN-C	TGTGACACACAGCAACCCTT	CTTGCCCACAGCTTGTTTCC	No
UncX	CGCCCACACCTCCTTGTATT	GTCGGGGTTTACCTGCACTC	No
ALX-1	CTCTGTTTAACCAGCTCACACTAACA	GATTGTTCTGAATCTGAGGATAGCTG	No
Pax9	GCAGACTTGGCAGTGTCCAT	GTGAGTGTCCAGGAAGACCC	Yes

SI Table 2: Enhancer reporter constructs. A table showing all the putative enhancer I.D. numbers, along with their associated gene chromosomal location and size. It records which candidate enhancer regions were successfully cloned into their reporter constructs, and for each of these if the injections and electroporation gave a positive result, and the number of times this was tested. It also shows if the reporter construct was electroporated alongside a control; pCIGcherry was injected and electroporated into each embryo at the same time as the fluorescent reporter construct. Enh88 was an enhancer region associated with MEOX1, that was injected into a separate embryo to the reporter construct, but under same conditions.

Enhancer I.D	Gene	Chromosomal Location	Size (bp)	Successfully made into a construct?	Fluorescence	Controls	Number of embryos injected and survived
110	Mkx	chr2:15,675,221-15,676,097	877	No	N/A	N/A	N/A
111	Mkx	chr2:15,683,191-15,683,507	317	Yes	No	Enh 88 and pCIGcherry	10

112	Mkx	chr2:15,684,450-15,685,904	1445	No	N/A	N/A	N/A
113	Tn-c	chr17:3,126,923-3,127,556	634	No	N/A	N/A	N/A
114	Tn-c	chr17:3,135,503-3,136,446	944	No	N/A	N/A	N/A
115	Tn-c	chr17:3,139,373-3,140,841	1453	Yes	No	Enh 88 and pCIG cherry	8
116	C-myc	chr2:139,861,206-139,861,744	644	No	N/A	N/A	N/A
117	C-myc	chr2:139,879,247-139,880,935	1689	No	N/A	N/A	N/A
118	ALX1	chr1:41,878,393-41,879,398	1006	Yes	No	Enh 88	15
119	ALX1	chr1:41,845,375-41,847,095	1721	No	N/A	N/A	N/A
120	ALX1	chr1:41,898,180-41,898,997	818	No	N/A	N/A	N/A
121	UNCX	chr14:2,403,710-2,404,244	664	No	N/A	N/A	N/A
122	UNCX	chr14:2,409,861-2,411,010	1106	Yes	No	Enh 88	12
123	UNCX	chr14:2,392,241-2,392,829	589	No	N/A	N/A	N/A
124	KLHL31	chr3:88,499,135-88,499,732	687	Yes	No	pCIGCherry	12
125	KLHL31	chr3:88,498,322-88,498,659	304	No	N/A	N/A	N/A
126	SCX	chrUn_NT_468793v1:69-333	265	No	N/A	N/A	N/A
127	Pax1	chr3:3,487,269-3,487,720	452	No	N/A	N/A	N/A
128	Pax1	chr3:3,513,348-3,513,973	644	Yes	No	Enh 88	9
129	Pax9	chr5:36,781,277-36,782,013	737	Yes	No	Enh 88	5
130	MEOX2	chr2:28,088,450-28,088,856407	407	No	N/A	N/A	N/A
131	MEOX2	chr2:28,078,535-28,079,316	728	No	N/A	N/A	N/A
132	Foxo1	chr1:170,566,684-170,567,320	637	Yes	No	pCIGCherry	8
133	Foxo1	chr1:170,723,431-170,724,139	709	Yes	No	pCIGCherry	7

141	Foxo1	chr1:170566788-170567431	644	No	N/A	N/A	N/A
142	Foxo1	chr1:170,576,501-170,577,708	1208	Yes	N/A	N/A	N/A
143	Foxo1	chr1:170,601,084-170,602,172	1089	Yes	No	pCIGCherry	6
144	Foxo1	chr1:170,619,863-170,621,075	1213	No	N/A	N/A	N/A
145	Foxo1	chr1:170,662,972-170,663,695	724	Yes	Yes	pCIGCherry	4
146	Foxo1	chr1:170,698,791-170,699,876	1068	No	N/A	N/A	N/A

SI Table 3: Successful PCR amplification primers. The forward and reverse primer pair for every fragment that was successfully amplified, inserted into a vector and made into a reporter construct. Each primer has the BSMBI sequence TTTTTTCGTCTCCAACAG at the 5' end for insertion during the BSMBI reaction.

Gene Name	Chromosomal location	Primer Pairs
MKX	chr2:15,683,191-15,683,507	Fw: GGGCCATCAAACTGAAGCG
		Rv: AATAATAACACTTCC
TN-C	chr17:3,139,373-3,140,841	Fw:AGCAGGTTAATAATAAAGCTACTGTACA
		Rv:TGAATTAGAACTCTAATCTTTACTTTCATCTCT
ALX1	chr1:41,878,393-41,879,398	Fw:AATATGCTAAACTGGCAGTCCAGA
		Rv:TTGACTCTCCTTTTTTAAGCCTGGAA
UNCX	chr14:2,409,861-2,411,010	FW: CCCGTAGAGGCTGTGC
		Rv:CTAGTTTGGAGTTAAATGGCTCATAAAAGC
KLHL31	chr3:88,499,135-88,499,732	Fw:ACATTGCCTTGCCAGCAGTA
		Rv:ATTCATGATAACACACTCTGGATTGTTTC
SCX	chrUn_NT_468793v1:69-333	Fw:GGGAGGGGGGAGTTGGGG

		Rv: AGGGTGCCGTGGCTCAG
PAX1	chr3:3,513,348-3,513,973	Fw:TTTGGGTTTGCCACACATCG
		Rv:ATACACAATCCAAATTAATGCAATGATCAACT
PAX9	chr5:36,781,277-36,782,013	Fw:TCAGAGATGGTTAAAACCAAGGAGAA
		Rv: TGGAAATAAACCAATTTCCTTTATGGTGG
FOXO1	chr1:170,566,684-170,567,320	Fw:AATTTGGCCTACTTACCAGAAGGG
		Rv:ATCTGCCTTGACAGTTTCTGGATAAAG
FOXO1	chr1:170,723,431-170,724,139	Fw:GAGGAACAGTTCTTTGCAAAACTGC
		Rv:TAGTACTGGCAGATGACATATCAAAGGA

SI Table 4: Primer sequences used to amplify regions of cDNA for generation of antisense RNA probes. This list encompasses the primers designed for the TOMO-Sequence project.

Gene name	Forward Primer	Reverse Primer	Successfully made into a probe?
TNFRSF14	GCAGACTTGGCAGTGTCCAT	GTGAGTGTCCAGGAAGACCC	No
ARIH1	ACCACATGGTCTGTCGGAAC	GCTGCAACAAAACCCTTCGT	Yes
NES	GGGGCGGAGGAAAGGAAATA	CCCTGATTCTGGAAGAACCAAC	No
Wnt5a	GCAGCACTGTGGACAACAAC	CACCGTCTTGAACTGGTCGT	Yes
AHNAK	CTGCTGACTGCAATCCCATTT	AACTGGTGAGGATCCCTGTG	Yes

HBM	CATCACACATTGCCACCAGC	GCAGCAATGGTGTCTTTATTGA	Yes
SZL	CAGCTGAAAACCAAGGGTTCG	TGTGCGTGAAGAGTTCCTCC	Yes
LRIG3	GTCCTGACGCCTGGGAATTT	AATCTGTGGGACAGGATGCC	Yes

SI Table 5: Primer sequences used to amplify regions of cDNA for generation of antisense probes for genes found to be of interest from the single cell RNA sequencing.

Gene name	Forward Primer	Reverse Primer	Successfully made into a probe?
OLFML3	GGGAGTTCACGCTCTTCTCG	GATGATCTGGTAGCCGTCGT	Yes
GPC4	CCACGAGATCAGCGAGTTCT	AAGTCGTACCTGCTCTTGGC	No
PCP4	GATAGGAGGAGTGGTGAAGCAG	GGACTGGGACCCCGCTTTTT	No
TBX22	GGATGTTCCCATCGGTCAGG	AGACTTAGCGCTCTTCAGGC	Yes
FABP5	TTTGGATCCATTCGCAGCGG	TACGGTCTTCTTGCCATCCC	No
NKX3-2	CTGACGCCTTTCTCCATCCA	GCTGGTACTGTCTCTGGTCG	No
DCN	TCTGGATTTGGGCCAGTGTG	TGGTGTTGTAGCCAAGAGGG	Yes
EBF3	CCGTGCATAAAGGCCATCAG	ACTGGTGCAAGATGGTGGAG	Yes
PRRX1	CGCCGGCCTGGACAATTTAC	GTGGCCATGGCACTGTAAGG	Yes
KRT18	CAGCACCACTTTCTCCACCT	GTTTGCGCAGAGCATCCAG	No
MSX2	AGCGAAGGAGGTTTTCTCCTC	CTCTGTGAGGTTGAGGGAGC	Yes
CDX4	CCCCTACTCAGACTACATGGG	TATGACCTGATGGATGTCAGCA	Yes
UNC45B	AAGCTCTACGATGACCTGCG	GAGGAACACCCTTGCGATGA	Yes
CHRNA1	GCACTGGGTTTACTACGCCT	ATTCAATAAGGCGACCGGCA	Yes
NEBL	AGGCCCTCACTGGATTTAGAC	GAGTCAAACAACTTGCCGAG	Yes
SGCG	CAACAACACCAGGCACTAGC	CGACTCCAGCCTTAGCTGTT	Yes
MUSTN1	CAGTCACCGCCAGGAATAAC	ACGTAGAAAGAACGCCCGTG	Yes
CTSK	AAAGCAGTACAACGGCAAGG	CACGTAGCCCTTATTGCCCC	Yes
LAMA2	CTACAGGCAGCCACTTTGGA	ACGTGTTGAGCGAAGTCCAT	Yes
B3GNT4	GTTTGTGCTGAAGGGCGATG	AGTGTTGTGAGCCTCCAGTG	Yes
CORO6	CAAAGTGTCACCGGCCTATT	GGGGTGACTCCCGTTGA	Yes
RAPSN	ATGCTGAAGTTCGCAGTGGT	GACTCTCCACACATGCCACA	Yes

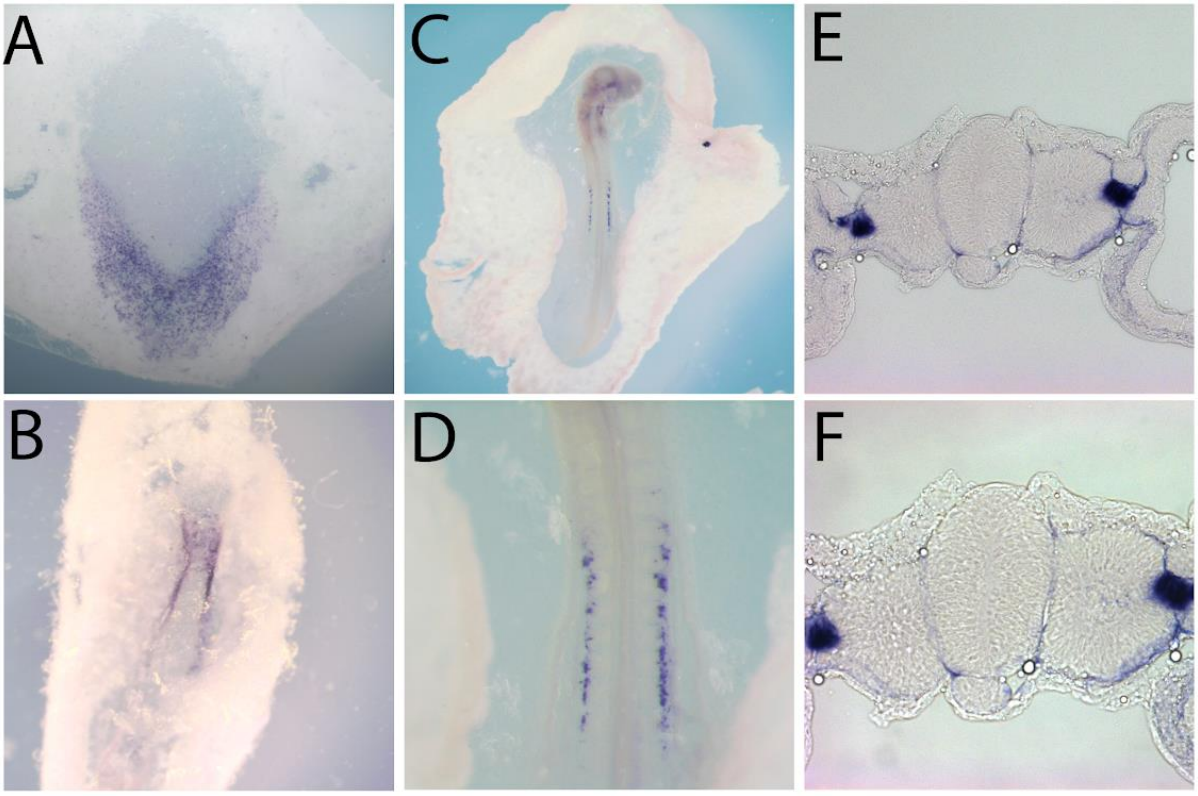

Figure S1: NPAS4 in-situ hybridisation and sectioning. A-D) In situ hybridisations to the NPAS4 RNA probe in A) HH4 embryo, B) HH8 embryo and C)HH14. D) Close up of the staining in C. E) 20µM section of the epithelial somite from the embryo shown in C, at 20X magnification. F) Section of somite from embryo in C at 40X magnification.

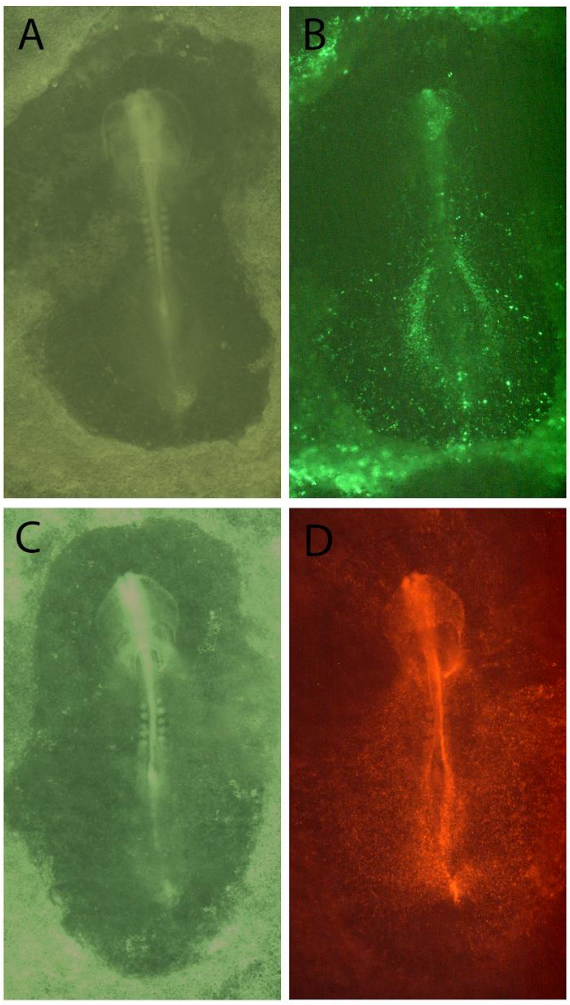

Figure S2: An example of the positive controls used for the electroporation technique. A and C show the brightfield images of a HH10 embryo. B) The GFP reporter construct injected alone shows signal in cells that were targeted. C) The pCIGcherry reporter construct also shows which cells were successfully injected and electroporated and was frequently used with tested putative candidate enhancer reporter constructs, as the signal is a separate colour.

Abbreviations

ATAC-Seq: Assay for Transposase Accessible Chromatin Sequencing
BMP: Bone morphogenic protein
ChIP-seq: Chromatin immunoprecipitation- sequencing
CREB: Cyclic AMP-responsive element-binding
CTCF: CCCTC-binding factor
DNA: deoxyribonucleic acid
EC: Early chick
EMT: Epithelial to mesenchymal transition
eQTL: Expression quantitative trait locus
FAIRE-Seq: formaldehyde-assisted identification of regulatory elements- sequencing
FGF: Fibroblast Growth Factor
GFP: Green fluorescent protein
HH: Hamburger-Hamilton stage
MET: Mesenchymal to epithelial transition
PAX: Paired box

PCR: Polymerase chain reaction
PSM: Pre-somitic Mesoderm
RA: Retinoic Acid
RT: Room temperature
Shh: Sonic Hedgehog
SIX: Sine oculis-related homeobox
TF: Transcription Factor
TGF: Transforming growth factor
TSS: Transcription start site
WISH: Whole mount *in-situ* hybridisation

For Images:
HF: Head fold
HT: Heat Tube
LB: Limb Bud
NT: Neural Tube
PS: Primitive Streak
So: Somite

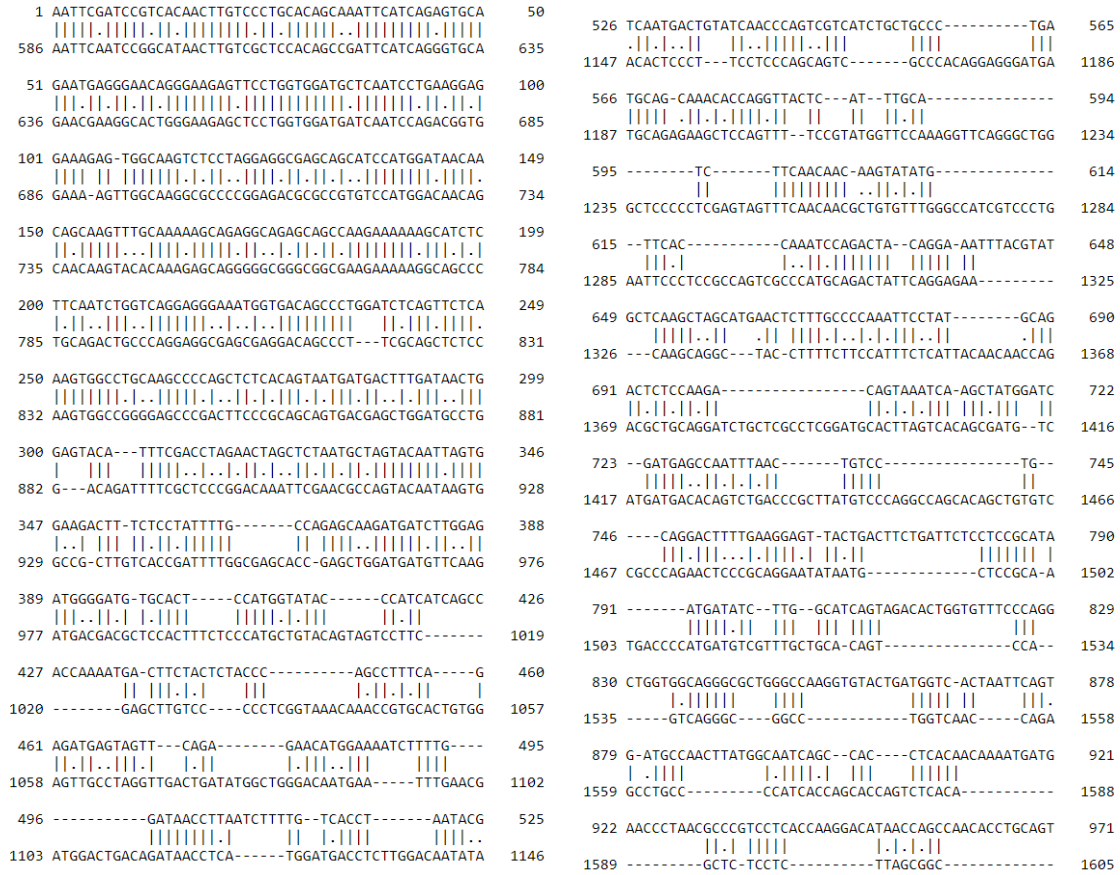

Figure S2: Pairwise sequence alignment of the Foxo1 sequence used for probe design (on top) and Foxo3 cDNA (on bottom). Analyses performed using the EMBOSS Water tool (https://www.ebi.ac.uk/Tools/psa/emboss_water/).

9: Bibliography:

Abou-Elhamd, A., Cooper, O. and Münsterberg, A. (2009). Klhl31 is associated with skeletal myogenesis and its expression is regulated by myogenic signals and Myf-5. *Mechanisms of Development*, 126(10), pp.852-862.

Abou-Elhamd, A., Alrefaei, A., Mok, G., Garcia-Morales, C., Abu-Elmagd, M., Wheeler, G. and Münsterberg, A. (2015). Klhl31 attenuates β-catenin dependent Wnt signaling and regulates embryo myogenesis. *Developmental Biology*, 402(1), pp.61-71.

Adams, J., Kelso, R. and Cooley, L., 2000. The kelch repeat superfamily of proteins: propellers of cell function. *Trends in Cell Biology*, 10(1), pp.17-24.

Adams, G., Chandru, A. and Cowley, S., 2018. Co-repressor, co-activator and general transcription factor: the many faces of the Sin3 histone deacetylase (HDAC) complex. *Biochemical Journal*, 475(24), pp.3921-3932.

Akiyama, H., 2002. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes & Development*, 16(21), pp.2813-2828.

Allen D.L., Unterman T.G., 2007. Regulation of myostatin expression and myoblast differentiation by FoxO and SMAD transcription factors. *Am J Physiol Cell Physiol*., 292 pp.188–199.

Alberton, P., Popov, C., Prägert, M., Kohler, J., Shukunami, C., Schieker, M. and Docheva, D., 2012. Conversion of Human Bone Marrow-Derived Mesenchymal Stem Cells into Tendon Progenitor Cells by Ectopic Expression of Scleraxis. *Stem Cells and Development*, 21(6), pp.846-858.

Almeida, R., Fernández-Justel, J., Santa-María, C., Cadoret, J., Cano-Aroca, L., Lombraña, R., Herranz, G., Agresti, A. and Gómez, M., 2018. Chromatin conformation regulates the coordination between DNA replication and transcription. *Nature Communications*, 9(1).

Alver, B., Kim, K., Lu, P., Wang, X., Manchester, H., Wang, W., Haswell, J., Park, P. and Roberts, C., 2017. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. Nature Communications, 8(1).

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H. and Shiroishi, T., 2009. Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Developmental Cell*, 16(1), pp.47-57.

Anderson, D., Arredondo, J., Hahn, K., Valente, G., Martin, J., Wilson-Rawls, J. and Rawls, A., 2006. Mohawkis a novel homeobox gene expressed in the developing mouse embryo. *Developmental Dynamics*, 235(3), pp.792-801.

Anderson, D. and Rawls, A., 2007. The homeobox gene Mohawk functions as a transcriptional repressor via three independent, evolutionarily conserved domains. *Developmental Biology*, 306(1), p.363.

Anderson, D., Beres, B., Wilson-Rawls, J. and Rawls, A., 2009. The homeobox gene Mohawk represses transcription by recruiting the sin3A/HDAC co-repressor complex. *Developmental Dynamics*, 238(3), pp.572-580.

Aoyama, H., Asamoto, K., 1998. Determination of somite cells: independence of cell differentiation and morphogenesis. *Development*, 104, pp.15-28

Aoyama, T., Okamoto, T., Nagayama, S., Nishijo, K., Ishibe, T., Yasura, K., Nakayama, T., Nakamura, T. and Toguchida, J., 2004. Methylation in the Core-promoter Region of the Chondromodulin-I Gene Determines the Cell-specific Expression by Regulating the Binding of Transcriptional Activator Sp3. Journal of Biological Chemistry, 279(27), pp.28789-28797.

Arnold, S. and Robertson, E., 2009. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, 10(2), pp.91-103.

Asou, Y., Nifuji, A., Tsuji, K., Shinomiya, K., Olson, E., Koopman, P. and Noda, M., 2002. Coordinated expression of scleraxis andSox9 genes during embryonic development of tendons and cartilage. *Journal of Orthopaedic Research*, 20(4), pp.827-833.

Baek, K., Packard, R., Hsu, J., Saffari, A., Ma, Z., Luu, A., Pietersen, A., Yen, H., Ren, B., Ding, Y., Sioutas, C., Li, R. and Hsiai, T., 2018. Ultrafine Particle Exposure Reveals the Importance of FOXO1/Notch Activation Complex for Vascular Regeneration. *Antioxidants & Redox Signaling*, 28(13), pp.1209-1223.

Baldock, R., Bard, J., Davidson, D., Morriss-Kay, G. and Kaufman, M., 2016. *Kaufman's Atlas Of Mouse Development Supplement*. Amsterdam: Elsevier/Academic Press, pp.165-175.

Bajard, L., 2006. A novel genetic hierarchy functions during hypaxial myogenesis: Pax3 directly activates Myf5 in muscle progenitor cells in the limb. *Genes & Development*, 20(17), pp.2450-2464.

Banerji, J., Rusconi, S. and Schaffner, W., 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2), pp.299-308.

Bannister, A. and Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell Research*, 21(3), pp.381-395.

Bannykh, S., Emery, S., Gerber, J., Jones, K., Benirschke, K. and Masliah, E. (2003). AberrantPax1 andPax9 expression in Jarcho-Levin syndrome: Report of two Caucasian siblings and literature review. *American Journal of Medical Genetics*, 120A(2), pp.241-246.

Barolo, S., 2011. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *BioEssays*, 34(2), pp.135-141.

Berkes, C. and Tapscott, S., 2005. MyoD and the transcriptional control of myogenesis. *Seminars in Cell & Developmental Biology*, 16(4-5), pp.585-595.

Beverdam, A. and Meijlink, F., 2001. Expression patterns of group-I aristaless-related genes during craniofacial and limb development. *Mechanisms of Development*, 107(1-2), pp.163-167.

Biddie, S., John, S., Sabo, P., Thurman, R., Johnson, T., Schiltz, R., Miranda, T., Sung, M., Trump, S., Lightman, S., Vinson, C., Stamatoyannopoulos, J. and Hager, G., 2011. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. *Molecular Cell*, 43(1), pp.145-155.

Biddie, S., 2010. Chromatin Architecture and the Regulation of Nuclear Receptor Inducible Transcription. *Journal of Neuroendocrinology*, 23(1), pp.94-106.

Bird, A., 1986. CpG-rich islands and the function of DNA methylation. *Nature*, 321(6067), pp.209-213.

Blow, M., McCulley, D., Li, Z., Zhang, T., Akiyama, J., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Bristow, J., Ren, B., Black, B., Rubin, E., Visel, A. and Pennacchio, L., 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nature Genetics*, 42(9), pp.806-810.

Borello, U., 2006. The Wnt/ -catenin pathway regulates Gli-mediated Myf5 expression during somitogenesis. *Development*, 133(18), pp.3723-3732.

Borycki, AG., Mendham, L., Emerson, C.P., 1998. Control of somite patterning by Sonic hedgehog and its downstream signal response genes. *Development*, 125, pp.77-790

Boutz, P., Chawla, G., Stoilov, P. and Black, D., 2007. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes & Development*, 21(1), pp.71-84.

Braun, T. and Gautel, M., 2011. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. *Nature Reviews Molecular Cell Biology*, 12(6), pp.349-361.

Brand-Saberi, B. and Christ, B., 2000. Evolution and development of distinct cell lineages derived from somites *Curr Top Dev Bio*l, 48, pp.1-42

Breitbart, R., Andreadis, A. and Nadal-Ginard, B., 1987. Alternative Splicing: A Ubiquitous Mechanism for the Generation of Multiple Protein Isoforms from Single Genes. *Annual Review of Biochemistry*, 56(1), pp.467-495.

Brent, A., Schweitzer, R. and Tabin, C., 2003. A Somitic Compartment of Tendon Progenitors. *Cell*, 113(2), pp.235-248.

Brent, A., 2004. FGF acts directly on the somitic tendon progenitors through the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. *Development*, 131(16), pp.3885-3896.

Brill, G., Kahane, N., Carmeli,C., von Schack, D., Barde, Y.A., Kalcheim, C., 1995 Epithelial-mesenchymal conversion of dermatome progenitors requires neural tube-derived signals: characterization of the role of Neurotrophin-3. *Development*, 121, pp.2583-2594

Buchtová, M., Kuo, W., Nimmagadda, S., Benson, S., Geetha-Loganathan, P., Logan, C., Au-Yeung, T., Chiang, E., Fu, K. and Richman, J., 2009. Whole genome microarray analysis of chicken embryo facial prominences. *Developmental Dynamics*, 239(2), pp.574-591.

Buono, L., Naranjo, S., Moreno-Marmol, T., de la Cerda, B., Polvillo, R., Díaz-Corrales, F., Bogdanovic, O., Bovolenta, P., Martínez-Morales, J., Analysis of gene network bifurcation during optic cup morphogenesis in zebrafish
bioRxiv 2020.05.28.121038; doi: https://doi.org/10.1101/2020.05.28.121038

Buenrostro, J., Giresi, P., Zaba, L., Chang, H. and Greenleaf, W., 2014. Transposition of Native Chromatin for Fast and Sensitive Mulitmodal Analysis of Chromatin Architecture. *Biophysical Journal*, 106(2), p.77a.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, *et al*. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015;523:486–90.

Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D. and Relaix, F., 2003. The formation of skeletal muscle: from somite to limb. Journal of Anatomy, 202(1), pp.59-68.

Butler, J., 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development*, 16(20), pp.2583-2592.

Cai, R., 1998. Human CART1, a Paired-Class Homeodomain Protein, Activates Transcription through Palindromic Binding Sites. *Biochemical and Biophysical Research Communications*, 250(2), pp.305-311.

Cai, S., Lee, C. and Kohwi-Shigematsu, T., 2006. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nature Genetics*, 38(11), pp.1278-1288.

Cairns, D., Sato, M., Lee, P., Lassar, A. and Zeng, L., 2008. A gradient of Shh establishes mutually repressing somitic cell fates induced by Nkx3.2 and Pax3. *Developmental Biology*, 323(2), pp.152-165.

Calhabeu, F., Hayashi, S., Morgan, J., Relaix, F. and Zammit, P., 2012. Alveolar rhabdomyosarcoma-associated proteins PAX3/FOXO1A and PAX7/FOXO1A suppress the transcriptional activity of MyoD-target genes in muscle stem cells. *Oncogene*, 32(5), pp.651-662.

Calo, E. and Wysocka, J., 2013. Modification of Enhancer Chromatin: What, How, and Why?. *Molecular Cell*, 49(5), pp.825-837.

Cambray, N. and Wilson, V., 2007. Two distinct sources for a population of maturing axial progenitors. *Development*, 134(15), pp.2829-2840.

Cech, T. and Steitz, J., 2014. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, 157(1), pp.77-94.

Chen, A., Chen, D. and Chen, Y., 2018. Advances of DNase-seq for mapping active gene regulatory elements across the genome in animals. *Gene*, 667, pp.83-94.

Chen, J., Mandel, E., Thomson, J., Wu, Q., Callis, T., Hammond, S., Conlon, F. and Wang, D., 2005. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genetics*, 38(2), pp.228-233.

Chen, T., Ueda, Y., Dodge, J., Wang, Z. and Li, E., 2003. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Molecular and Cellular Biology*, 23(16), pp.5594-5605.

Chen, Y., He, S., Yan, F., Zhou, P., Luo, K., Zhang, Y., Xiao, Y. and Lin, M., 2016. Dental pulp stem cells express tendon markers under mechanical loading and are a potential cell source for tissue engineering of tendon-like tissue. *International Journal of Oral Science*, 8(4), pp.213-222.

Cheung, K., Barter, M., Falk, J., Proctor, C., Reynard, L. and Young, D., 2020. Histone ChIP-Seq identifies differential enhancer usage during chondrogenesis as critical for defining cell-type specificity. *The FASEB Journal*,.

Cho, S., Kim, S., Kim, J. and Kim, J., 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature Biotechnology, 31(3), pp.230-232.

Christ, B. and Ordahl, C., 1995. Early stages of chick somite development. *Anatomy and Embryology*, 191(5), pp.381-396.

Christ, B., Huang, R. and Scaal, M., 2004. Formation and differentiation of the avian sclerotome. *Anatomy and Embryology*, 208(5).

Clapier, C., Iwasa, J., Cairns, B. and Peterson, C., 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, 18(7), pp.407-422.

Creyghton, M., Cheng, A., Welstead, G., Kooistra, T., Carey, B., Steine, E., Hanna, J., Lodato, M., Frampton, G., Sharp, P., Boyer, L., Young, R. and Jaenisch, R., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50), pp.21931-21936.

Cripps, R., Black, B., Zhao, B., Lien, C., Schulz, R. and Olson, E., 1998. The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during Drosophila myogenesis. *Genes & Development*, 12(3), pp.422-434.

Cripps, R., Lovato, T. and Olson, E., 2004. Positive autoregulation of the Myocyte enhancer factor-2 myogenic control gene during somatic muscle development in Drosophila. *Developmental Biology*, 267(2), pp.536-547.

Cserjesi, P., Brown, D., Ligon, KL., Lyons, GE., Copeland, NG., Gilbert, DJ. (1995) Scleraxis: a basic helix–loop–helix protein that prefigures skeletal formation during mouse embryogenesis *Development*, 121, pp. 1099-1110

Dale, J., Maroto, M., Dequeant, M., Malapert, P., McGrew, M. and Pourquie, O., 2003. Periodic Notch inhibition by Lunatic Fringe underlies the chick segmentation clock. *Nature*, 421(6920), pp.275-278.

Daly, C., 2004. Angiopoietin-1 modulates endothelial cell function and gene expression via the transcription factor FKHR (FOXO1). *Genes & Development*, 18(9), pp.1060-1071.

Damle, S. and Davidson, E., 2011. Precise cis-regulatory control of spatial and temporal expression of the alx-1 gene in the skeletogenic lineage of s. purpuratus. *Developmental Biology*, 357(2), pp.505-517.

Daubas, P., Duval, N., Bajard, L., Langa Vives, F., Robert, B., Mankoo, B. and Buckingham, M., 2015. Fine-tuning the onset of myogenesis by homeobox proteins that interact with the Myf5 limb enhancer. *Biology Open*, 4(12), pp.1614-1624.

Daubas, P. and Buckingham, M., 2013. Direct molecular regulation of the myogenic determination gene Myf5 by Pax3, with modulation by Six1/4 factors, is exemplified by the −111kb-Myf5 enhancer. *Developmental Biology*, 376(2), pp.236-244.

de Crombrugghe, B., Lefebvre, V., Behringer, R., Bi, W., Murakami, S. and Huang, W., 2000. Transcriptional mechanisms of chondrocyte differentiation. *Matrix Biology*, 19(5), pp.389-394.

de Wit, E. and de Laat, W., 2012. A decade of 3C technologies: insights into nuclear organization. *Genes & Development*, 26(1), pp.11-24.

Deaton, A. and Bird, A., 2011. CpG islands and the regulation of transcription. *Genes & Development*, 25(10), pp.1010-1022.

Dequéant, M. and Pourquié, O., 2008. Segmental patterning of the vertebrate embryonic axis. *Nature Reviews Genetics*, 9(5), pp.370-382.

Devbhandari, S., Jiang, J., Kumar, C., Whitehouse, I. and Remus, D., 2017. Chromatin Constrains the Initiation and Elongation of DNA Replication. *Molecular Cell*, 65(1), pp.131-141.

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A., Dixon, J., Maliskova, L., Guan, K., Shen, Y. and Ren, B., 2016. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Research*, 26(3), pp.397-405.

Dixon, J., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. and Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485(7398), pp.376-380.

Dixon, J., Gorkin, D. and Ren, B., 2016. Chromatin Domains: The Unit of Chromosome Organization. Molecular Cell, 62(5), pp.668-680.

Doane, A. and Elemento, O., 2017. Regulatory elements in molecular networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 9(3), p.e1374.

Edom-Vovard, F., Schuler, B., Bonnin, M., Teillet, M. and Duprez, D., 2002. Fgf4 Positively Regulates scleraxis and Tenascin Expression in Chick Limb Tendons. *Developmental Biology*, 247(2), pp.351-366.

Erbay, E., Park, I., Nuzzi, P., Schoenherr, C. and Chen, J., 2003. IGF-II transcription in skeletal myogenesis is controlled by mTOR and nutrients. *Journal of Cell Biology*, 163(5), pp.931-936.

Ettensohn, C., 2003. Alx1, a member of the Cart1/Alx3/Alx4 subfamily of Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo. *Development*, 130(13), pp.2917-2928.

Evrard, Y., Lun, Y., Aulehla, A., Gan, L. and Johnson, R. (1998). lunatic fringe is an essential mediator of somite segmentation and patterning. *Nature*, 394(6691), pp.377-381.

Ferjentsik, Z., Hayashi, S., Dale, J., Bessho, Y., Herreman, A., De Strooper, B., del Monte, G., de la Pompa, J. and Maroto, M., 2009. Notch Is a Critical Component of the Mouse Somitogenesis Oscillator and Is Essential for the Formation of the Somites. *PLoS Genetics*, 5(9), p.e1000662.

Finckenstein, F., Davicioni, E., Osborn, K., Cavenee, W., Arden, K. and Anderson, M., 2006. Transgenic mice expressing PAX3-FKHR have multiple defects in muscle development, including ectopic skeletal myogenesis in the developing neural tube. *Transgenic Research*, 15(5), pp.595-614.

Florini, J.R., K.A. Magri, D.Z. Ewton, P.L. James, K. Grindstaff, and P.S. Rotwein., 1991. "Spontaneous" differentiation of skeletal myoblasts is dependent upon autocrine secretion of insulin-like growth factor-II. J. Biol. Chem. 266, pp.15917–15923

Frankel, N., Davis, G., Vargas, D., Wang, S., Payre, F. and Stern, D., 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, 466(7305), pp.490-493.

Friedensohn, S. and Sawarkar, R., 2014. Cis-regulatory variation: significance in biomedicine and evolution. *Cell and Tissue Research*, 356(3), pp.495-505.

Furtado, M., Wilmanns, J., Chandran, A., Tonta, M., Biben, C., Eichenlaub, M., Coleman, H., Berger, S., Bouveret, R., Singh, R., Harvey, R., Ramialison, M., Pearson, J., Parkington, H., Rosenthal, N. and Costa, M., 2016. A novel conditional mouse model for Nkx2-5 reveals transcriptional regulation of cardiac ion channels. Differentiation, 91(1-3), pp.29-41.

Furuyama, T., Kitayama, K., Shimoda, Y., Ogawa, M., Sone, K., Yoshida-Araki, K., Hisatsune, H., Nishikawa, S., Nakayama, K., Nakayama, K., Ikeda, K., Motoyama, N. and Mori, N., 2004. Abnormal Angiogenesis inFoxo1(Fkhr)-deficient Mice. *Journal of Biological Chemistry*, 279(33), pp.34741-34749.

Fyodorov, D., Zhou, B., Skoultchi, A. and Bai, Y., 2017. Emerging roles of linker histones in regulating chromatin structure and function. Nature Reviews Molecular Cell Biology, 19(3), pp.192-206.

Gasperini, M., Tome, J. and Shendure, J., 2020. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature Reviews Genetics*

Gibb, S., Zagorska, A., Melton, K., Tenin, G., Vacca, I., Trainor, P., Maroto, M. and Dale, J., 2009. Interfering with Wnt signalling alters the periodicity of the segmentation clock. *Developmental Biology*, 330(1), pp.21-31.

Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M. and Maire, P., 2007. Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. *Proceedings of the National Academy of Sciences*, 104(27), pp.11310-11315.

Gluhak, J., Mais, A. and Mina, M., 1996. Tenascin-C is associated with early stages of chondrogenesis by chick mandibular ectomesenchymal cells in vivo and in vitro. *Developmental Dynamics*, 205(1), pp.24-40.

Goldstein, RS. and Kalchiem, C., 1992 Determination of epithelial half-somites in skeletal morphogenesis. *Development*, 116, pp.441-445

Gong, F., Sun, L., Wang, Z., Shi, J., Li, W., Wang, S., Han, X. and Sun, Y., 2011. The BCL2 gene is regulated by a special AT-rich sequence binding protein 1-mediated long range chromosomal interaction between the promoter and the distal element located within the 3′-UTR. *Nucleic Acids Research*, 39(11), pp.4640-4652.

Gossett, L., Kelvin, D., Sternberg, E. and Olson, E., 1989. A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Molecular and Cellular Biology*, 9(11), pp.5022-5033.

Grifone, R., 2005. Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. *Development*, 132(9), pp.2235-2249.

Gupta, V. and Beggs, A. (2014). Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skeletal Muscle*, 4(1), p.11.

Guss, K., 2001. Control of a Genetic Regulatory Network by a Selector Gene. *Science*, 292(5519), pp.1164-1167.

Hadjur, S., Williams, L., Ryan, N., Cobb, B., Sexton, T., Fraser, P., Fisher, A. and Merkenschlager, M., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*, 460(7253), pp.410-413.

Häfner, S., 2019. Puzzling mRNA: Alternative splicing fine-tunes specificity and function. *Biomedical Journal*, 42(3), pp.137-140.

Hallegger, M., Llorian, M. and Smith, C., 2010. Alternative splicing: global insights. *FEBS Journal*, 277(4), pp.856-866.

Heinzel, T., Lavinsky, R., Mullen, T., Söderström, M., Laherty, C., Torchia, J., Yang, W., Brard, G., Ngo, S., Davie, J., Seto, E., Eisenman, R., Rose, D., Glass, C. and Rosenfeld, M., 1997. A complex containing N-CoR, mSln3 and histone deacetylase mediates transcriptional repression. *Nature*, 387(6628), pp.43-48.

Heisler, N., Ishimatsu, A. and Iwama, G., 1999. Physiological role of the secondary circulatory system in fish. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 124, p.S40.

Herrgen, L., Ares, S., Morelli, L., Schröter, C., Jülicher, F. and Oates, A., 2010. Intercellular Coupling Regulates the Period of the Segmentation Clock. *Current Biology*, 20(14), pp.1244-1253.

Hesselberth, J., Chen, X., Zhang, Z., Sabo, P., Sandstrom, R., Reynolds, A., Thurman, R., Neph, S., Kuehn, M., Noble, W., Fields, S. and Stamatoyannopoulos, J., 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods, 6(4), pp.283-289.

Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* 187, 226–232 (1975).

Hong, J., Hendrix, D. and Levine, M., 2008. Shadow Enhancers as a Source of Evolutionary Novelty. *Science*, 321(5894), pp.1314-1314.

Hosaka, T., Biggs, W., Tieu, D., Boyer, A., Varki, N., Cavenee, W. and Arden, K., 2004. Disruption of forkhead transcription factor (FOXO) family members in mice reveals their functional diversification. *Proceedings of the National Academy of Sciences*, 101(9), pp.2975-2980.

Hribal, M., Nakae, J., Kitamura, T., Shutter, J. and Accili, D., 2003. Regulation of insulin-like growth factor–dependent myoblast differentiation by Foxo forkhead transcription factors. *Journal of Cell Biology*, 162(4), pp.535-541.

Hrycaj, S. and Wellik, D., 2016. Hox genes and evolution. F1000Research, 5, p.859.

Hu, J., Yuan, W., Tang, M., Wang, Y., Fan, X., Mo, X., Li, Y., Ying, Z., Wan, Y., Ocorr, K., Bodmer, R., Deng, Y. and Wu, X. (2010). KBTBD7, a novel human BTB-kelch protein, activates transcriptional activities of SRE and AP-1. *BMB Reports*, 43(1), pp.17-22.

Ikeya M, Takada S 1998. Wnt signaling from the dorsal neural tube is required for the formation of the medial dermomyotome. Development 125: 4969–4976

Ishii, H., Kadonaga, J. and Ren, B., 2015. MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proceedings of the National Academy of Sciences*, 112(27), pp.E3457-E3465.

Ito, Y., Toriuchi, N., Yoshitaka, T., Ueno-Kudoh, H., Sato, T., Yokoyama, S., Nishida, K., Akimoto, T., Takahashi, M., Miyaki, S. and Asahara, H., 2010. The Mohawk homeobox gene is a critical regulator of tendon differentiation. *Proceedings of the National Academy of Sciences*, 107(23), pp.10538-10542.

Jabbari, K. and Bernardi, G., 2004. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene*, 333, pp.143-149.

Jamali, M., Rogerson, P., Wilton, S. and Skerjanc, I., 2001. Nkx2–5 Activity Is Essential for Cardiomyogenesis. Journal of Biological Chemistry, 276(45), pp.42252-42258.

Jambhekar, A., Dhall, A. and Shi, Y., 2019. Author Correction: Roles and regulation of histone methylation in animal development. *Nature Reviews Molecular Cell Biology*, 21(1), pp.59-59.

Jin, J. and Ding, J., 2006. Analysis of Meox-2 mutant mice reveals a novel postfusion-based cleft palate. *Developmental Dynamics*, 235(2), pp.539-546.

Jacob, F. and Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), pp.318-356.

Jiao, S., Ren, H., Li, Y., Zhou, J., Duan, C. and Lu, L., 2012. Differential regulation of IGF-I and IGF-II gene expression in skeletal muscle cells. *Molecular and Cellular Biochemistry*, 373(1-2), pp.107-113.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G., 2009. H3.3/H2A.Z double variant–containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature Genetics*, 41(8), pp.941-945.

Jones., S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biology*, 226

Kadonaga, J., 2011. Perspectives on the RNA polymerase II core promoter. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(1), pp.40-51.

Kagey, M., Newman, J., Bilodeau, S., Zhan, Y., Orlando, D., van Berkum, N., Ebmeier, C., Goossens, J., Rahl, P., Levine, S., Taatjes, D., Dekker, J. and Young, R., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), pp.430-435.

Kahane N., Cinnamon Y., Bachelet I. and Kalcheim C. (2001) The third wave of myotome colonization by mitotically competent progenitors.*Development.* 128, pp.2187-2198

Kanhere, A., Viiri, K., Araújo, C., Rasaiyaah, J., Bouwman, R., Whyte, W., Pereira, C., Brookes, E., Walker, K., Bell, G., Pombo, A., Fisher, A., Young, R. and Jenner, R., 2010. Short RNAs Are Transcribed from Repressed Polycomb Target Genes and Interact with Polycomb Repressive Complex-2. *Molecular Cell*, 38(5), pp.675-688.

Kamei, Y., Miura, S., Suzuki, M., Kai, Y., Mizukami, J., Taniguchi, T., Mochida, K., Hata, T., Matsuda, J., Aburatani, H., Nishino, I. and Ezaki, O., 2004. Skeletal Muscle FOXO1 (FKHR) Transgenic Mice Have Less Skeletal Muscle Mass, Down-regulated Type I (Slow Twitch/Red Muscle) Fiber Genes, and Impaired Glycemic Control. *Journal of Biological Chemistry*, 279(39), pp.41114-41123.

Kalcheim, C. and Ben-Yair, R., 2005. Cell rearrangements during development of the somite and its derivatives. *Current Opinion in Genetics & Development*, 15(4), pp.371-380.

Karnuta, J. and Scacheri, P., 2018. Enhancers: bridging the gap between gene control and human disease. *Human Molecular Genetics*, 27(R2), pp.R219-R227.

Kassar-Duchossoy, L., Gayraud-Morel, B., Gomès, D., Rocancourt, D., Buckingham, M., Shinin, V. and Tajbakhsh, S., 2004. Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature*, 431(7007), pp.466-471.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, S., Falaleeva, M., Stamm, S. 2013 Function of alternative splicing *Gene*, 514, pp. 1-30.

Kiecker, C., Bates, T. and Bell, E., 2015. Molecular specification of germ layers in vertebrate embryos. *Cellular and Molecular Life Sciences*, 73(5), pp.923-947.

Kim, M., Lesoon-Wood, L., Weintraub, B. and Chung, J., 1996. A soluble transcription factor, Oct-1, is also found in the insoluble nuclear matrix and possesses silencing activity in its alanine-rich domain. *Molecular and Cellular Biology*, 16(8), pp.4366-4377.

Kishore, S., 2006. The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C. *Science*, 311(5758), pp.230-232.

Knoll, M., Lodish, H. and Sun, L., 2015. Long non-coding RNAs as regulators of the endocrine system. *Nature Reviews Endocrinology*, 11(3), pp.151-160.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A., Heidemann, M., Hintermair, C., Eick, D., Gut, I., Ferrier, P. and Andrau, J., 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology*, 18(8), pp.956-963.

Kolesky, S., Ouhammouch, M. and Peter Geiduschek, E., 2020. *The Mechanism Of Transcriptional Activation By The Topologically DNA-Linked Sliding Clamp Of Bacteriophage T4*.

Kolovos, P., Knoch, T., Grosveld, F., Cook, P. and Papantonis, A., 2012. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin*, 5(1), p.1.

Korkmaz, G., Lopes, R., Ugalde, A., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R. and Agami, R., 2016. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nature Biotechnology*, 34(2), pp.192-198.

Krivega, I. and Dean, A., 2012. Enhancer and promoter interactions—long distance calls. *Current Opinion in Genetics & Development*, 22(2), pp.79-85.

Kurakazu, I., Akasaki, Y., Hayashida, M., Tsushima, H., Goto, N., Sueishi, T., Toya, M., Kuwahara, M., Okazaki, K., Duffy, T., Lotz, M. and Nakashima, Y., 2019. FOXO1 transcription factor regulates chondrogenic differentiation through transforming growth factor β1 signaling. *Journal of Biological Chemistry*, 294(46), pp.17555-17569.

Kurat, C., Yeeles, J., Patel, H., Early, A. and Diffley, J., 2017. Chromatin Controls DNA Replication Origin Selection, Lagging-Strand Synthesis, and Replication Fork Rates. *Molecular Cell*, 65(1), pp.117-130.

Kwasnieski, J., Fiore, C., Chaudhari, H. and Cohen, B., 2014. High-throughput functional testing of ENCODE segmentation predictions. *Genome Research*, 24(10), pp.1595-1602.
Laclef, C., 2003. Altered myogenesis in Six1-deficient mice. *Development*, 130(10), pp.2239-2252.

Lagha, M., Kormish, J., Rocancourt, D., Manceau, M., Epstein, J., Zaret, K., Relaix, F. and Buckingham, M., 2008. Pax3 regulation of FGF signaling affects the progression of embryonic progenitor cells into the myogenic program. *Genes & Development*, 22(13), pp.1828-1837.

Lamichhaney, S., Berglund, J., Almén, M., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C., Wang, C., Zamani, N., Grant, B., Grant, P., Webster, M. and Andersson, L., 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), pp.371-375.

Lecerf, L., Kavo, A., Ruiz-Ferrer, M., Baral, V., Watanabe, Y., Chaoui, A., Pingault, V., Borrego, S. and Bondurand, N., 2014. An Impairment of Long Distance SOX10 Regulatory Elements Underlies Isolated Hirschsprung Disease. *Human Mutation*, 35(3), pp.303-307.

Lee, C.S., Buttitta, L.A., May, N.R., Kripert, A., and Fan. C.M., 2000. SHH-N upregulates Sfrp2 to mediate its competitive interaction with WNT1 and WNT4 in the somitic mesoderm., *Development*, 127, pp.109-118

Lee CS, Buttitta L, Fan CM (2001) Evidence that the WNT-inducible growth arrest-specific gene 1 encodes an antagonist of sonic hedgehog signaling in the somite. Proc Natl Acad Sci USA 98(20):11347–11352

Lee, H., 2005. Nuclear integration of positive Dpp signals, antagonistic Wg inputs and mesodermal competence factors during Drosophila visceral mesoderm induction. *Development*, 132(6), pp.1429-1442.

Lee, T. and Young, R., 2013. Transcriptional Regulation and Its Misregulation in Disease. *Cell*, 152(6), pp.1237-1251.

Leitges M, Neidhardt L, Haenig B, Herrmann BG & Kispert A 2000 The paired homeobox gene Uncx41 specifies pedicles, transverse processes, and proximal ribs of the vertebral column. Development 127 2259–2267.

Lefebvre, V. and Dvir-Ginzberg, M., 2016. SOX9 and the many facets of its regulation in the chondrocyte lineage. *Connective Tissue Research*, 58(1), pp.2-14.

Léjard, V., Brideau, G., Blais, F., Salingcarnboriboon, R., Wagner, G., Roehrl, M., Noda, M., Duprez, D., Houillier, P. and Rossert, J., 2007. Scleraxis and NFATc Regulate the Expression of the Pro-α1(I) Collagen Gene in Tendon Fibroblasts. *Journal of Biological Chemistry*, 282(24), pp.17665-17675.

Liu, N., Nelson, B., Bezprozvannaya, S., Shelton, J., Richardson, J., Bassel-Duby, R. and Olson, E., 2014. Requirement of MEF2A, C, and D for skeletal muscle regeneration. *Proceedings of the National Academy of Sciences*, 111(11), pp.4109-4114.

Liu, C., Yang, Z., Liu, C., Wang, R., Tien, P., Dale, R. and Sun, L., 2007. Effect of RNA oligonucleotide targeting Foxo-1 on muscle growth in normal and cancer cachexia mice. *Cancer Gene Therapy*, 14(12), pp.945-952.

Lührig, S., Kolb, S., Mellies, N. and Nolte, J. (2013). The novel BTB-kelch protein, KBTBD8, is located in the Golgi apparatus and translocates to the spindle apparatus during mitosis. *Cell Division*, 8(1), p.3.

Lupiáñez, D., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A. and Mundlos, S., 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell, 161(5), pp.1012-1025.

Lupien, M., Eeckhoute, J., Meyer, C., Wang, Q., Zhang, Y., Li, W., Carroll, J., Liu, X. and Brown, M., 2008. FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell*, 132(6), pp.958-970.

MacAlpine, H., Gordan, R., Powell, S., Hartemink, A. and MacAlpine, D., 2009. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Research*, 20(2), pp.201-211.

Mackie, E., Thesleff, I. and Chiquet-Ehrismann, R., 1987. Tenascin is associated with chondrogenic and osteogenic differentiation in vivo and promotes chondrogenesis in vitro. *The Journal of Cell Biology*, 105(6), pp.2569-2579.

Mankoo, B., Collins, N., Ashby, P., Grigorieva, E., Pevny, L., Candia, A., Wright, C., Rigby, P. and Pachnis, V., 1999. Mox2 is a component of the genetic hierarchy controlling limb muscle development. *Nature*, 400(6739), pp.69-73.

Mansouri, A., Yokota, Y., Wehr, R., Copeland, N., Jenkins, N. and Gruss, P., 1997. Paired-related murine homeobox gene expressed in the developing sclerotome, kidney, and nervous system. *Developmental Dynamics*, 210(1), pp.53-65.
Mariño-Ramírez, L., Kann, M., Shoemaker, B. and Landsman, D., 2005. Histone structure and nucleosome stability. *Expert Review of Proteomics*, 2(5), pp.719-729.

Marcelle C, Stark MR, Bronner-Fraser M (1997) Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. Development 124(20):3955–3963

Maroto, M., Reshef, R., Münsterberg, A., Koester, S., Goulding, M. and Lassar, A., 1997. Ectopic Pax-3 Activates MyoD and Myf-5 Expression in Embryonic Mesoderm and Neural Tissue. *Cell*, 89(1), pp.139-148.

Maston, G., Evans, S. and Green, M., 2006. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, 7(1), pp.29-59.

Matsuzaki, H., Daitoku, H., Hatta, M., Tanaka, K. and Fukamizu, A., 2003. Insulin-induced phosphorylation of FKHR (Foxo1) targets to proteasomal degradation. *Proceedings of the National Academy of Sciences*, 100(20), pp.11285-11290.

Mayuranathan, T., Rayabaram, J., Nath, A. and Velayudhan, S., 2015. Identification of Erythroid Specific Enhancers By Analyzing Genome Wide Association of Transcriptional Co-Activators CBP and P300. *Blood*, 126(23), pp.3330-3330.

Merkenschlager, M. and Odom, D., 2013. CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets. *Cell*, 152(6), pp.1285-1297.

Mezger A, Klemm S, Mann I, Brower K, Mir A, Bostick M, *et al*. High-throughput chromatin accessibility profiling at single-cell resolution. Nat Commun. 2018;9:3647

McGaughran, J., Oates, A., Donnai, D., Read, A. and Tassabehji, M. (2003). Mutations in PAX1 may be associated with Klippel–Feil syndrome. *European Journal of Human Genetics*, 11(6), pp.468-474.

Mok, G., Lozano-Velasco, E. and Münsterberg, A., 2017. microRNAs in skeletal muscle development. *Seminars in Cell & Developmental Biology*, 72, pp.67-76.

Mok, G., Folkes, L., Weldon, S., Maniou, E., Martinez-Heredia, V., Godden, A., Williams, R., Wheeler., G.N., Moxon, S., Münsterberg, A., 'Characterising open chromatin identifies novel cis-regulatory elements important for paraxial mesoderm formation and axis extension' bioRxiv 2020.01.20.912337; doi: https://doi.org/10.1101/2020.01.20.912337

Monsoro-Burq, A., 2005. Sclerotome development and morphogenesis: when experimental embryology meets genetics. *The International Journal of Developmental Biology*, 49(2-3), pp.301-308.

Moreno, T. and Kintner, C., 2004. Regulation of Segmental Patterning by Retinoic Acid Signaling during Xenopus Somitogenesis. *Developmental Cell*, 6(2), pp.205-218.

Muerdter, F., Boryń, Ł. and Arnold, C., 2015. STARR-seq — Principles and applications. *Genomics*, 106(3), pp.145-150.

Munsterberg, A., Kitajewski, J., Bumcrot, D., McMahon, A. and Lassar, A., 1995. Combinatorial signaling by Sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. *Genes & Development*, 9(23), pp.2911-2922.

Murchison, N., Price, B., Conner, D., Keene, D., Olson, E., Tabin, C. and Schweitzer, R., 2007. Regulation of tendon differentiation by scleraxis distinguishes force-transmitting tendons from muscle-anchoring tendons. *Development*, 134(14), pp.2697-2708.

Nakae, J., Kitamura, T., Kitamura, Y., Biggs, W., Arden, K. and Accili, D., 2003. The Forkhead Transcription Factor Foxo1 Regulates Adipocyte Differentiation. *Developmental Cell*, 4(1), pp.119-129.

Nakamura, H. and Funahashi, J., 2001. Introduction of DNA into Chick Embryos by in Ovo Electroporation. *Methods*, 24(1), pp.43-48.

Nakamura, H., Katahira, T., Sato, T., Watanabe, Y. and Funahashi, J., 2004. Gain- and loss-of-function in chick embryos by electroporation. *Mechanisms of Development*, 121(9), pp.1137-1143.

Nan, X., Ng, H., Johnson, C., Laherty, C., Turner, B., Eisenman, R. and Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), pp.386-389.

Nature, 2012. An integrated encyclopedia of DNA elements in the human genome. 489(7414), pp.57-74.

Neidhardt, L., Kispert, A. and Herrmann, B., 1997. A mouse gene of the paired-related homeobox class expressed in the caudal somite compartment and in the developing vertebral column, kidney and nervous system. *Development Genes and Evolution*, 207(5), pp.330-339.

Nieto, NA., and Wilkinson., D.G., 1996. In situ hybridization analysis of chick embryos in whole mount and in tissue sections *Methods Cell Biol*., 51, pp. 219-235

Nittoli, V., Fortunato, A., Fasano, G., Coppola, U., Gentile, A., Maiella, S., Langellotto, F., Porreca, I., De Paolo, R., Marino, R., Fiengo, M., Donizetti, A., Aniello, F., Kondo, T., Ristoratore, F., Canzoniero, L., Duboule, D., Wilson, S. and Sordino, P., 2019. Characterization of paralogous uncx transcription factor encoding genes in zebrafish. *Gene: X*, 2, p.100011.

Ng, S., Yue, W., Oppermann, U. and Klose, R., 2008. Dynamic protein methylation in chromatin biology. *Cellular and Molecular Life Sciences*, 66(3), pp.407-422.

Nguyen, H. and Xu, X., 1998. Drosophila mef2Expression during Mesoderm Development Is Controlled by a Complex Array ofcis-Acting Regulatory Modules. *Developmental Biology*, 204(2), pp.550-566.

Ogbourne, S. and Antalis, T., 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochemical Journal*, 331(1), pp.1-14.

Ong, C. and Corces, V., 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12(4), pp.283-293.

Ortega, N., Behonick, D. and Werb, Z., 2004. Matrix remodeling during endochondral ossification. Trends in Cell Biology, 14(2), pp.86-93.

Papizan, J., Garry, G., Brezprozvannaya, S., McAnally, J., Bassel-Duby, R., Liu, N. and Olson, E., 2017. Deficiency in Kelch protein Klhl31 causes congenital myopathy in mice. *Journal of Clinical Investigation*, 127(10), pp.3730-3740.

Park, S., Sakamoto, H., Tsuji-Tamura, K., Furuyama, T. and Ogawa, M., 2009. Foxo1 is essential for in vitro vascular formation from embryonic stem cells. *Biochemical and Biophysical Research Communications*, 390(3), pp.861-866.

Peters, H., Will, B., Sakai, N., Mass, N., and Balling, R. 1999. Pax1 and Pax9 synergistically regulate vertebral column development. *Development* 126(23), pp. 5399-408

Perez, A., Perrine, M., Brainard, N. and Vogel, K., 2003. Scleraxis (Scx) directs lacZ expression in tendon of transgenic mice. *Mechanisms of Development*, 120(10), pp.1153-1163.

Perry, M., Boettiger, A. and Levine, M., 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proceedings of the National Academy of Sciences*, 108(33), pp.13570-13575.

Peng, Y. and Zhang, Y., 2018. Enhancer and super-enhancer: Positive regulators in gene transcription. *Animal Models and Experimental Medicine*, 1(3), pp.169-179.

Potthoff, M. and Olson, E., 2007. MEF2: a central regulator of diverse developmental programs. *Development*, 134(23), pp.4131-4140.

Pourquie, O., Coltey, M., Breant, C. and Le Douarin, N., 1995. Control of somite patterning by signals from the lateral plate. *Proceedings of the National Academy of Sciences*, 92(8), pp.3219-3223.

Pourquié, O., 2011. Vertebrate Segmentation: From Cyclic Gene Networks to Scoliosis. *Cell*, 145(5), pp.650-663.

Pownall, M. and Isaacs, H., 2010. *FGF Signalling In Vertebrate Development*. San Rafael, Calif.: Morgan & Claypool.

Provot, S., 2006. Nkx3.2/Bapx1 acts as a negative regulator of chondrocyte maturation. *Development*, 133(4), pp.651-662.

Pryce, B., Watson, S., Murchison, N., Staverosky, J., Dunker, N. and Schweitzer, R., 2009. Recruitment and maintenance of tendon progenitors by TGF signaling are essential for tendon formation. Development, 136(8), pp.1351-1361.

Petrascheck, M., 2005. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Research*, 33(12), pp.3743-3750.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S., Flynn, R. and Wysocka, J., 2010. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), pp.279-283.

Rainbow, R., Kwon, H. and Zeng, L., 2014. The role of Nkx3.2 in chondrogenesis. *Frontiers in Biology*, 9(5), pp.376-381.

Rampakakis, E., Arvanitis, D., Di Paola, D. and Zannis-Hadjopoulos, M., 2009. Metazoan origins of DNA replication: Regulation through dynamic chromatin structure. *Journal of Cellular Biochemistry*, 106(4), pp.512-520.

Ranganayakulu, G., Zhao, B., Dokidis, A., Molkentin, J., Olson, E. and Schulz, R., 1995. A Series of Mutations in the D-MEF2 Transcription Factor Reveal Multiple Functions in Larval and Adult Myogenesis in Drosophila. *Developmental Biology*, 171(1), pp.169-181.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B., Sun, Z., Schmid, M., Opravil, S., Mechtler, K., Ponting, C., Allis, C. and Jenuwein, T., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), pp.593-599.

Rao, P., Kumar, R., Farkhondeh, M., Baskerville, S. and Lodish, H., 2007. Myogenic factors that regulate expression of myogenic microRNAs. *The FASEB Journal*, 21(5).

Relaix, F., 2004. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes & Development*, 18(9), pp.1088-1105.

Relaix, F., Rocancourt, D., Mansouri, A. and Buckingham, M., 2005. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature*, 435(7044), pp.948-953.

Ren, X., Siegel, R., Kim, U. and Roeder, R., 2011. Direct Interactions of OCA-B and TFII-I Regulate Immunoglobulin Heavy-Chain Gene Transcription by Facilitating Enhancer-Promoter Communication. *Molecular Cell*, 42(3), pp.342-355.

Reyes, A. and Huber, W., 2017. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Research*, 46(2), pp.582-592.

Rinn, J. and Chang, H., 2012. Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, 81(1), pp.145-166.

Rhee, H. and Pugh, B., 2011. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell*, 147(6), pp.1408-1419.

Rodrigo, I., Hill, R.E., Balling, R., Münsterberg, A., Imai, K. (2003). Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development*, 130(3), pp.473-482.

Roth, S., Denu, H.M., and Allis., D. 2001. Histone Acetyltransferases. *Annual review of Biochemistry*., 70, pp.81-120

Rudnicki MA, Schnegelsberg PN, Stead RH, Braun T, Arnold HH, Jaenisch R 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. Cell 75: 1351–1359

Sammeta, N., Hardin, D. and McClintock, T. (2010). Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium. *Molecular and Cellular Neuroscience*, 45(4), pp.398-407.

Sanchez, A., 2015. FoxO transcription factors and endurance training: a role for FoxO1 and FoxO3 in exercise-induced angiogenesis. *The Journal of Physiology*, 593(2), pp.363-364.

Sánchez, R. and Sánchez, S., 2013. Characterization of pax1, pax9, and uncx sclerotomal genes during Xenopus laevis embryogenesis. *Developmental Dynamics*, 242(5), pp.572-579.

Sanderson, M., 2008. Phylogenetic Signal in the Eukaryotic Tree of Life. Science, 321(5885), pp.121-123.

Sato, T., Rocancourt, D., Marques, L., Thorsteinsdóttir, S. and Buckingham, M., 2010. A Pax3/Dmrt2/Myf5 Regulatory Cascade Functions at the Onset of Myogenesis. *PLoS Genetics*, 6(4), p.e1000897.

Saunders, C., van der Merwe, L., Posthumus, M., Cook, J., Handley, C., Collins, M. and September, A., 2012. Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations. *Journal of Orthopaedic Research*, 31(4), pp.632-637.

Scaal, M. and Christ, B., 2004. Formation and differentiation of the avian dermomyotome. *Anatomy and Embryology*, 208(6).

Scaal, M., Gros, J., Lesbros, C. and Marcelle, C., 2004. In ovo electroporation of avian somites. *Developmental Dynamics*, 229(3), pp.643-650.

Schmidt, D., Schwalie, P., Ross-Innes, C., Hurtado, A., Brown, G., Carroll, J., Flicek, P. and Odom, D., 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Research*, 20(5), pp.578-588.

Schupbach T, Wieschaus E. 1991. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* 129(4):1119-1136.

Schwerk, C. and Schulze-Osthoff, K., 2005. Regulation of Apoptosis by Alternative Pre-mRNA Splicing. *Molecular Cell*, 19(1), pp.1-13.

Sekimata, M., Pérez-Melgosa, M., Miller, S., Weinmann, A., Sabo, P., Sandstrom, R., Dorschner, M., Stamatoyannopoulos, J. and Wilson, C., 2009. CCCTC-Binding Factor and the Transcription Factor T-bet Orchestrate T Helper 1 Cell-Specific Structure and Function at the Interferon-γ Locus. *Immunity*, 31(4), pp.551-564.

Selleck, M.A., Stern, C.D. (1991) Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development*, 112, pp.615-626

Simon, M., Campos-Xavier, A., Mittaz-Crettol, L., Valadares, E., Carvalho, D., Speck-Martins, C., Nampoothiri, S., Alanay, Y., Mihci, E., van Bever, Y., Garcia-Segarra, N., Cavalcanti, D., Mortier, G., Bonafé, L. and Superti-Furga, A., 2012. Severe neurologic manifestations from cervical spine instability in spondylo-megaepiphyseal-metaphyseal dysplasia. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 160C(3), pp.230-237.

Simon, J., Giresi, P., Davis, I. and Lieb, J., 2012. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature Protocols*, 7(2), pp.256-267.

Sheng, G., dos Reis, M. and Stern, C., 2003. Churchill, a Zinc Finger Transcriptional Activator, Regulates the Transition between Gastrulation and Neurulation. *Cell*, 115(5), pp.603-613.

Shibata, S., Zhang, J., Puthumana, J., Stone, K. and Lifton, R. (2013). Kelch-like 3 and Cullin 3 regulate electrolyte homeostasis via ubiquitination and degradation of WNK4. *Proceedings of the National Academy of Sciences*, 110(19), pp.7838-7843.

Smith, T., 2005. Feedback interactions between MKP3 and ERK MAP kinase control scleraxis expression and the specification of rib progenitors in the developing chick somite. *Development*, 132(6), pp.1305-1314.

Shukunami, C., Takimoto, A., Oro, M. and Hiraki, Y., 2006. Scleraxis positively regulates the expression of tenomodulin, a differentiation marker of tenocytes. *Developmental Biology*, 298(1), pp.234-247.

Spilianakis, C., Lalioti, M., Town, T., Lee, G. and Flavell, R., 2005. Interchromosomal associations between alternatively expressed loci. *Nature*, 435(7042), pp.637-645.

Stamm,S., Ben-Ari, S., Rafalska,I., Tang,Y., Zhang, Z., Toiber, D., Thanarajc, T.A., Soreq, H. 2005. Function of alternative splicing. *Gene*, 334, pp.1-20

Steensel, B. and Henikoff, S., 2000. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nature Biotechnology*, 18(4), pp.424-428.

Stern, C. and Piatkowska, A., 2015. Multiple roles of timing in somite formation. *Seminars in Cell & Developmental Biology*, 42, pp.134-139.

Stitt, T., Drujan, D., Clarke, B., Panaro, F., Timofeyva, Y., Kline, W., Gonzalez, M., Yancopoulos, G. and Glass, D., 2004. The IGF-1/PI3K/Akt Pathway Prevents Expression of Muscle Atrophy-Induced Ubiquitin Ligases by Inhibiting FOXO Transcription Factors. *Molecular Cell*, 14(3), pp.395-403.

Strober, B.J., Elorbany, R., Rhodes, K., Krishnan, N., Tayeb, K., Battle, A. and Gilad, Y., 2019. Dynamic genetic regulation of gene expression during cellular differentiation. *Science*, 364, pp.1287-1290

Sung, M.-H., Baek, S. and Hager, G. L., 2016. Genome-wide footprinting: ready for prime time? *Nat. Methods* 13, 222–228.

Sweetman, D., Goljanek, K., Rathjen, T., Oustanina, S., Braun, T., Dalmay, T. and Münsterberg, A., 2008. Specific requirements of MRFs for the expression of muscle specific microRNAs, miR-1, miR-206 and miR-133. *Developmental Biology*, 321(2), pp.491-499.

Suzuki, H., Ito, Y., Shinohara, M., Yamashita, S., Ichinose, S., Kishida, A., Oyaizu, T., Kayama, T., Nakamichi, R., Koda, N., Yagishita, K., Lotz, M., Okawa, A. and Asahara, H., 2016. Gene targeting of the transcription factor Mohawk in rats causes heterotopic ossification of Achilles tendon via failed tenogenesis. *Proceedings of the National Academy of Sciences*, 113(28), pp.7840-7845.

Symmons, O., Uslu, V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L. and Spitz, F., 2014. Functional and topological characteristics of mammalian regulatory domains. Genome Research, 24(3), pp.390-400.

Takahashi, J., Ohbayashi, A., Oginuma, M., Saito, D., Mochizuki, A., Saga, Y. and Takada, S., 2010. Analysis of Ripply1/2-deficient mouse embryos reveals a mechanism underlying the rostro-caudal patterning within a somite. *Developmental Biology*, 342(2), pp.134-145.

Takahashi, Y., Yasuhiko, Y., Takahashi, J., Takada, S., Johnson, R., Saga, Y. and Kanno, J., 2013. Metameric pattern of intervertebral disc/vertebral body is generated independently of Mesp2/Ripply-mediated rostro-caudal patterning of somites in the mouse embryo. *Developmental Biology*, 380(2), pp.172-184.

Tajbakhsh S, Borello U, Vivarelli E, Kelly R, Papkoff J, Duprez D, Buckingham M, Cossu G 1998. Differential activation of *Myf5* and *MyoD* by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of *Myf5*. Development 125: 4155–4162

Takimoto, A., Mohri, H., Kokubu, C., Hiraki, Y. and Shukunami, C., 2013. Pax1 acts as a negative regulator of chondrocyte maturation. *Experimental Cell Research*, 319(20), pp.3128-3139.

Tan, G., Pryce, B., Stabio, A., Brigande, J., Wang, C., Xia, Z., Tufa, S., Keene, D. and Schweitzer, R., 2020. Tgfβ signaling is critical for maintenance of the tendon cell fate. *eLife*, 9.

Tanaka, M. and Tickle, C., 2004. Tbx18 and boundary formation in chick somite and wing development. *Developmental Biology*, 268(2), pp.470-480.

Tenin, G., Wright, D., Ferjentsik, Z., Bone, R., McGrew, M. and Maroto, M., 2010. The chick somitogenesis oscillator is arrested before all paraxial mesoderm is segmented into somites. BMC Developmental Biology, 10(1), p.24.

Thurman, R., Rynes, E., Humbert, R. *et al.*, 2012. The accessible chromatin landscape of the human genome. *Nature* (489), pp.75–82.

Towers, M. and Tickle, C., 2008. Growing models of vertebrate limb development. Development, 136(2), pp.179-190.

Uz, E., Alanay, Y., Aktas, D., Vargel, I., Gucer, S., Tuncbilek, G., von Eggeling, F., Yilmaz, E., Deren, O., Posorski, N., Ozdag, H., Liehr, T., Balci, S., Alikasifoglu, M., Wollnik, B. and Akarsu, N., 2010. Disruption of ALX1 Causes Extreme Microphthalmia and Severe Facial Clefting: Expanding the Spectrum of Autosomal-Recessive ALX-Related Frontonasal Dysplasia. *The American Journal of Human Genetics*, 86(5), pp.789-796.

Valcourt, U., Thuault, S., Pardali, K., Heldin, C. and Moustakas, A., 2007. Functional role of Meox2 during the epithelial cytostatic response to TGF-β. *Molecular Oncology*, 1(1), pp.55-71.

Vakoc, C., Letting, D., Gheldof, N., Sawado, T., Bender, M., Groudine, M., Weiss, M., Dekker, J. and Blobel, G., 2005. Proximity among Distant Regulatory Elements at the β-Globin Locus Requires GATA-1 and FOG-1. *Molecular Cell*, 17(3), pp.453-462.

van Gastel, N., Stegen, S., Eelen, G., Schoors, S., Carlier, A., Daniëls, V., Baryawno, N., Przybylski, D., Depypere, M., Stiers, P., Lambrechts, D., Van Looveren, R., Torrekens, S., Sharda, A., Agostinis, P., Lambrechts, D., Maes, F., Swinnen, J., Geris, L., Van Oosterwyck, H., Thienpont, B., Carmeliet, P., Scadden, D. and Carmeliet, G., 2020. Lipid availability determines fate of skeletal progenitor cells via SOX9. *Nature*, 579(7797), pp.111-117.

van der Wijst, M., Brugge, H., de Vries, D., Deelen, P., Swertz, M. and Franke, L., 2018. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nature Genetics*, 50(4), pp.493-497.

Villarejo-Balcells, B., Guichard, S., Rigby, P. and Carvajal, J., 2011. Expression pattern of the FoxO1 gene during mouse embryonic development. *Gene Expression Patterns*, 11(5-6), pp.299-308.

Visel, A., Minovitsky, S., Dubchak, I. and Pennacchio, L., 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Research*, 35(Database), pp.D88-D92.

Visel, A., Blow, M., Li, Z., Zhang, T., Akiyama, J., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. and Pennacchio, L., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), pp.854-858.

WADDINGTON, C., 1942. CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. *Nature*, 150(3811), pp.563-565.

Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B., and Balling, R., 1994. The role of Pax-1 in axial skeleton development. *Development*, 120, pp.1109-1121.

Wang, X., Hou, J., Quedenau, C. and Chen, W., 2016. Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Molecular Systems Biology*, 12(7), p.875.

Wang, H., 2001. Methylation of Histone H4 at Arginine 3 Facilitating Transcriptional Activation by Nuclear Hormone Receptor. *Science*, 293(5531), pp.853-857.

Weinhold, N., Jacobsen, A., Schultz, N., Sander, C. and Lee, W., 2014. Genome-wide analysis of noncoding regulatory mutations in cancer. *Nature Genetics*, 46(11), pp.1160-1165.

Williams, R. M., Senanayake, U., Artibani, M., Taylor, G., Wells, D., Ahmed, A. A. and 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. *Dev. Camb. Engl*. 145,.

White, P., 2003. Defective somite patterning in mouse embryos with reduced levels of Tbx6. *Development*, 130(8), pp.1681-1690.

Wittkopp, P. and Kalay, G., 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1), pp.59-69.

Weber, M., Hellmann, I., Stadler, M., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genetics*, 39(4), pp.457-466.

Wolfman, N., Hattersley, G., Cox, K., Celeste, A., Nelson, R., Yamaji, N., Dube, J., DiBlasio-Smith, E., Nove, J., Song, J., Wozney, J. and Rosen, V., 1997. Ectopic induction of tendon and ligament in rats by growth and differentiation factors 5, 6, and 7, members of the TGF-beta gene family. *Journal of Clinical Investigation*, 100(2), pp.321-330.

Wu, A., Kim, J., Zhang, C., Unterman, T. and Chen, J., 2007. Forkhead Box Protein O1 Negatively Regulates Skeletal Myocyte Differentiation through Degradation of Mammalian Target of Rapamycin Pathway Components. *Endocrinology*, 149(3), pp.1407-1414.

Wu, Y. and Gong, Z., 2004. A novel zebrafish kelchlike gene klhl and its human ortholog KLHL display conserved expression patterns in skeletal and cardiac muscles. *Gene*, 338(1), pp.75-83.

Vermot, J. and Pourquié, O., 2005. Retinoic acid coordinates somitogenesis and left–right patterning in vertebrate embryos. *Nature*, 435(7039), pp.215-220.

Yaneza, M., Gilthorpe, J., Lumsden, A. and Tucker, A., 2002. No evidence for ventrally migrating neural tube cells from the mid- and hindbrain. *Developmental Dynamics*, 223(1), pp.163-167.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1).

Yusuf, F. and Brand-Saberi, B., 2006. The eventful somite: patterning, fate determination and cell division in the somite. *Brain Structure and Function*, 211(S1), pp.21-30.

Zaret, K. and Carroll, J., 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, 25(21), pp.2227-2241.

Zheng, L., Mao, C., Bi, Y.G., Zhou, Y., Zhang, Z., Zhao, H., Park, K., Huang, R., Cai, D., Qi, X. 2020. Differential expression of foxo genes during embryonic development and in adult tissues of Xenopus tropicalis. *Gene expression patterns*, 35, 119091.

Zeng, L., 2002. Shh establishes an Nkx3.2/Sox9 autoregulatory loop that is maintained by BMP signals to induce somitic chondrogenesis. *Genes & Development*, 16(15), pp.1990-2005.

Zentner, G., Tesar, P. and Scacheri, P., 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, 21(8), pp.1273-1283.

Zeng, Y. and Chen, T., 2019. DNA Methylation Reprogramming during Mammalian Development. *Genes*, 10(4), p.257

Zhuang, G., Yu, K., Jiang, Z., Chung, A., Yao, J., Ha, C., Toy, K., Soriano, R., Haley, B., Blackwood, E., Sampath, D., Bais, C., Lill, J. and Ferrara, N., 2013. Phosphoproteomic Analysis Implicates the mTORC2-FoxO1 Axis in VEGF Signaling and Feedback Activation of Receptor Tyrosine Kinases. *Science Signaling*, 6(271), pp.ra25-ra25.

Zhang, X., Ramalho-Santos, M. and McMahon, A., 2001. Smoothened Mutants Reveal Redundant Roles for Shh and Ihh Signaling Including Regulation of L/R Asymmetry by the Mouse Node. *Cell*, 105(6), pp.781-792.

Ziller, C. and Smith, J., 1982. Migration and differentiation of neural crest cells and their derivatives : in vivo and in vitro studies on the early development of the avian peripheral nervous system. *Reproduction Nutrition Développement*, 22(1B), pp.153-162..
Zimmerman LB, De Jesús-Escobar JM, Harland RM. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. Cell. 1996;86:599–606.

10: Post-Script Comment

After submission of the thesis and completion of the year in the lab, it was discovered that a mistake had been made in the lab regarding the labelling of probes. It transpired that the Foxo1 probe had been mislabelled and was a probe for the TBX22 gene. The produced in-situ images (Figures 3.1-3.4) were actually in-situ images for the gene TBX22. This means there is no need to further investigate the disparity between the in-situ images and the enhancer images. However, the finding of an enhancer associated with Foxo1 is still important for understanding the genetic regulation in the haematopoietic lineage, and a corresponding correct in-situ image for Foxo1 mRNA transcripts would show where Foxo1 is expressed in the chick embryo.

TBX22 is a gene heavily involved in the rostral caudal patterning of somites in early development (Haneg *et al*., 2002). The in-situ images produced by this work also demonstrates this patterning (fig. 3.1-3.4). In later development, TBX22 is crucial for the formation of the beak in chickens, and missense mutations in the coding region of the gene result in syndromes like clip left palate in humans

(Braybrook *et al*., 2002, Pauws *et al.*, 2009 and Higashihori *et al*., 2010). However, it's role in early development not been extensively studied in somitogenesis. As this in-situ has such strong specificity in early somites and pre-somitic mesoderm, it would be interesting to further study this gene and its genetic interactions at this earlier stage.

Braybrook, C., 2002. Craniofacial expression of human and murine TBX22 correlates with the cleft palate and ankyloglossia phenotype observed in CPX patients. *Human Molecular Genetics*, 11(22), pp.2793-2804.

Haenig, B., Schmidt, C., Kraus, F., Pfordt, M. and Kispert, A., 2002. Cloning and expression analysis of the chick ortholog of TBX22, the gene mutated in X-linked cleft palate and ankyloglossia. *Mechanisms of Development*, 117(1-2), pp.321-325.

Higashihori, N., Buchtová, M. and Richman, J., 2009. The function and regulation of TBX22 in avian frontonasal morphogenesis. *Developmental Dynamics*, 239(2), pp.458-473.

Pauws, E., Hoshino, A., Bentley, L., Prajapati, S., Keller, C., Hammond, P., Martinez-Barbera, J., Moore, G. and Stanier, P., 2010. Tbx22null mice have a submucous cleft palate due to reduced palatal bone formation and also display ankyloglossia and choanal atresia phenotypes. *Human Molecular Genetics*, 19(15), pp.3103-3103.