



HAL
open science

Transcriptional regulatory network underlying connective tissue differentiation during limb development

Mickael Orgeur

► **To cite this version:**

Mickael Orgeur. Transcriptional regulatory network underlying connective tissue differentiation during limb development. *Development Biology*. Université Pierre et Marie Curie - Paris VI; Freie Universität (Berlin), 2016. English. NNT: 2016PA066325 . tel-01474874

HAL Id: tel-01474874

<https://theses.hal.science/tel-01474874>

Submitted on 23 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transcriptional regulatory network underlying connective tissue differentiation during limb development

Dissertation

submitted to obtain the academic degree of
Philosophiæ Doctor (PhD)

in agreement with the co-tutelle contract between
Université Pierre et Marie Curie, Paris
and
Freie Universität, Berlin

by
Mickael Orgeur

July 2016

Dissertation in agreement with the co-tutelle contract between

Université Pierre et Marie Curie, Paris

Ecole Doctorale Complexité du Vivant (ED 515)

Laboratoire de Biologie du Développement (IBPS)

and

Freie Universität, Berlin

Department of Biology, Chemistry and Pharmacy

Max Planck Institute for Molecular Genetics

**Transcriptional regulatory network underlying connective
tissue differentiation during limb development**

Mickael Orgeur

Supervised by Prof Sigmar Stricker and Dr Delphine Duprez

Publicly presented and defended on the 26th of September, 2016

in front of a dissertation committee composed of:

Dr Francesca Spagnoli (UPMC reviewer)

Prof Krzysztof Jagla (UPMC reviewer)

Prof Simone Spuler (FU reviewer)

Prof Sigmar Stricker (thesis co-supervisor, FU reviewer)

Dr Delphine Duprez (thesis co-supervisor)

Dr Gisèle Bonne (UPMC representative)

Dr Pascal de Santa Barbara (guest)

Dr Ezequiel Mendoza (FU post-doc)

I hereby declare that this thesis entitled “Transcriptional regulatory network underlying connective tissue differentiation during limb development” has been composed solely by myself, independently and with no other sources or aids than cited or acknowledged. I testify that this thesis has not been previously submitted, in whole or in part, for a degree or any other qualification.

Paris, 19th of July 2016

Mickael Orgeur

CONTENTS

ACKNOWLEDGMENTS	vi
ABSTRACT	ix
RÉSUMÉ	x
ZUSAMMENFASSUNG	xi
NOTATIONS AND ABBREVIATIONS	xii
INTRODUCTION	1
A. From a single unit to a multi-compartment system	1
Skeleton development	2
Skeletal muscle development	4
Tendon development	7
Muscle connective tissue development	9
Muscle interaction with other compartments of the musculoskeletal system	11
B. From a single gene to a multi-layered network	14
Regulation of gene transcription	15
DNA regulatory elements	18
Chromatin modelling	20
Promoter and enhancer domains	22
Transcription factor binding	24
Hierarchical cooperativity	26
C. PhD project outline	29
Selection of connective tissue-associated transcription factors	30
Deciphering the molecular aspects of connective tissue cell differentiation	31
MATERIALS AND METHODS	32
A. Experimental procedures	32
Fertilized chicken eggs	32
<i>In situ</i> hybridization probes	32
<i>In situ</i> hybridization on paraffin sections	32
Molecular cloning of the transcription factors	33
Grafts of virus-producing cells into chick embryo limbs	34
Whole-mount <i>in situ</i> hybridization	34
Skeletal preparation	35
Retroviral particle production	35
Chicken micromass cultures	36
Histological staining	37
Western blot	37
Gene expression profiling	38
Cell cross-link	38
Nuclear extraction	39
Chromatin sonication	39
Chromatin immunoprecipitation	40
Expression pattern of interesting targeted genes	41

B. Computational analysis	41
Transcript-discovery approach	41
Differential expression analysis	43
Estimation of transcript abundances	44
Prediction of cell type abundances	45
Estimation of overexpression levels	45
K-means gene clustering	46
Gene ontology analysis	46
ChIP-sequencing	46
Histone modification peak calling	47
Identification of regulatory domains	47
Quality control of transcription factor-binding profiles	48
Irreproducible discovery rate analysis	48
Determination of the transcription factor-binding sites	49
Binding motif analysis	49
Assessment of the similarity between the ChIP-seq profiles	50
Data visualization	50
RESULTS	51
A. Characterization of the transcription factors	51
The TFs are related to subcompartments of the musculoskeletal system	51
OSR1 and OSR2 overexpression induces musculoskeletal defects in forelimbs	53
The TFs influence cell differentiation processes in the chMM cultures	54
B. Gene expression profiling	57
Improving the gene expression quantification	57
Estimation of the levels of transcription factor overexpression	60
The gene expression profiles in culture recapitulate the limb expression patterns	62
The TFs share common regulatory patterns	63
The TFs share common regulatory functions	66
C. Chromatin landscape	69
Defining the regulatory domains	69
DE genes are associated with bivalent promoters	72
D. Transcription factor-binding sites	75
Quality control and peak calling	75
Functionality of TFBS	80
E. Characterization of candidate target genes	81
Selected candidate genes	81
Expression pattern of the TFs and their target genes in E5.5 chick limbs	83
The TFs and their target genes display overlapping expression domains in limb tissues at E8	85
DISCUSSION	87
A. Designing a three-level transcriptional network	87
Improving the gene annotation of the chicken genome	87
Distinguishing true ChIP-seq signal	89
Addressing the functionality of transcription factor-binding events	91

B. Effects of the transcription factors on cell differentiation	92
Odd-skipped related 1 and 2 (OSR1, OSR2)	93
Early growth response 1 (EGR1)	94
Krüppel-like factors 2 and 4 (KLF2, KLF4)	95
C. Regulatory profiles of the transcription factors	96
Biological processes and signalling pathways regulated by connective tissue cells	96
Characterization of genes directly regulated by the TFs	98
DE genes are enriched for bivalent promoter domains	99
Identification of misregulated non-coding RNAs	100
D. Future prospects	101
REFERENCES	103
SUPPLEMENTAL INFORMATION	123
Supplemental figures	123
Supplemental tables	130
SCIENTIFIC PRODUCTIONS	133
Publications	133
Scientific communications	133
LIST OF FIGURES AND TABLES	134
List of figures and supplemental figures	134
List of tables and supplemental tables	135

This PhD thesis is dedicated to everybody who has contributed, from near to far, to this 4-year adventure between Berlin and Paris.

ACKNOWLEDGMENTS

I have always heard that you complete a PhD thesis solely on your own. But actually, it is not entirely true, and this section is dedicated to everyone who supported me throughout this period.

First of all, I would like to express my gratitude to Prof Stefan Mundlos for the warm welcome in his group at the Max Planck Institute for Molecular Genetics in Berlin. Your contribution and assistance have been crucial for the initiation of this PhD project. For the Paris side, I am thankful to Dr Sylvie Schneider-Maunoury for welcoming me at the Developmental Biology Laboratory, which greatly enhanced the scope of my thesis.

I am deeply indebted to my both co-supervisors, Prof Sigmar Stricker and Dr Delphine Duprez, for offering me the exceptional opportunity to conduct this PhD project. Both of you have shown an outstanding diligence in supervising me from the beginning to the end of this thesis and beyond. I could write several paragraphs to express my gratefulness towards you two, but I will limit myself to only one reason, which is, to my eyes, the most important. Thank you very much for the confidence that you had in my work and in choosing me to conduct this PhD project, despite my limited knowledge in developmental biology when I applied.

I am grateful to Dr Francesca Spagnoli, Prof Krzysztof Jagla and Prof Simone Spuler for their consent to evaluate my PhD thesis, as well as Dr Pascal Santa de Barbara, Dr Gisèle Bonne and Dr Ezequiel Mendoza for accepting to be part of my dissertation jury.

I would like also to mention the members of my PhD committee, Prof Helge Amthor and Dr David Sassoon, for their comments and advice, essential to pursue my project in the right direction.

I am thankful to Dr Jochen Hecht, Dr Bernd Timmermann, Dr Stefan Börno and Ulrike Krüger for having taken care of the sequencing of my numerous and precious ChIP/RNA samples. Thank you also to Prof Peter Robinson, Dr Peter Hansen, Dr Christophe Antoniewski and Dr Marius Van Den Beek for maintaining the Galaxy servers necessary to the analysis of the sequencing data.

I would like to thank Marvin Martens, Georgeta Leonte and Marie-Ange Bonnin for their technical assistance and valuable help in some of the experimental procedures included in this thesis. I am also thankful to Sophie Gournet for helping me with some of the illustrations of my thesis.

Thank you to the numerous and fantastic members of the research group Development and Disease. El Jüüürgen and El Pedro, Arriba Arriba, Sánchez! I could write several pages to relate all the diverting moments we spent together. You have been great and funny lab mates, inside and outside the lab, in Berlin and elsewhere. I really enjoyed working, talking and going out with you. What an amazing pocket map of Berlin!! The ChIP-seekers Daniel and Ivana, always up to troubleshoot background peaks and low-titre retroviruses, as well as to empathise these long days of chMM preparation/harvesting! My desk mates, el topo Dario (But why??? Pizza!) and

Stéphane, Christophe and Marc, always up for a scientific or non-scientific conversation (around a coffee or a few beers)!

If I am writing these lines, which conclude this 4-year thesis, it is thanks to several people that I would like to mention: Pierrot, Dominique, Alisson, Arne, Linnea, Itxaso, Rauan and Jean-Christophe. Thank you for giving me the wish to pursue in research. Vielen Dank Roland und Carmen! It has been a great pleasure to stay in touch over these last past years.

To my PhD fellows: Fabien (mon pote Claudy à qui tu parles autrement!) and Michaël (Bernie!). It has been a great pleasure to share this experience with you, for the worse but also for the many greatest moments of these last past years. Congratulations to both of you! Thank you also to Lucile and Maude, respectively, for taking care of them while I was absent!

Thank you to my numerous flatmates who supported me: Lucie, Elitza and Antoine. Special thanks to, not just Lucile Just, but also Solen and Nils. You have been wonderful. I honestly do not know how it would have ended without you following and holding me. Mein Genosse Faust, always up for a beer, a DnB session with High Rankin, a conversation, a meal, a festival, Dr Pong, a burger, a trip to Paris, Barcelona, Brussels or Berlin, and partying! Mr Roux, une bière du Nord!!

A mes compères des contrées obscures et pieds humides : Crestey, Bereskiki et LeHoux. Un énorme merci à Croust et Bouvier, là où vous savez ! Vous ne compreniez rien à ce que je faisais, mais cela ne vous a pas empêché d'avoir toujours été là pour me soutenir et me changer les idées pour toutes sortes de choses les plus folles les unes que les autres. Chailly-en-1664 !

A mes parents, mon frère et mon papy topette. Je vous remercie du fond du cœur pour votre soutien pendant ces 4 longues années, pour avoir toujours cru en moi et pour avoir accepté toutes les décisions que j'ai pu prendre. Je vous dédicace tout particulièrement cette thèse.

Finally, I would like to thank you Leonora. You have supported me the most and I could always rely on you to bring me happiness after tough days in the lab. I will never forget what you have done for me and I wish you all the best in your future. Tack så mycket!

ABSTRACT

The musculoskeletal system is composed of muscles, skeletal elements and connective tissues such as tendon and muscle connective tissue. Functional locomotion via this system relies on a precise coordination between its different components. Muscle connective tissue contributes to the elasticity and rigidity of muscles, while tendons transmit forces generated by muscles to the bone to allow body motion. Despite their distinct mesodermal origins, differentiation of muscle and connective tissue progenitor cells are closely related throughout limb development. In contrast to muscle and skeleton, connective tissue patterning and formation remain poorly investigated. In order to identify molecular mechanisms underlying connective tissue formation during limb development, five zinc-finger transcription factors were investigated: OSR1, OSR2, EGR1, KLF2 and KLF4. These transcription factors are expressed in distinct subcompartments of the musculoskeletal system and influence the differentiation of limb mesenchymal cells upon overexpression. To further investigate their roles at the molecular level, several genome-wide strategies were employed in limb mesenchymal explant cultures overexpressing each of the transcription factors. Whole-transcriptome sequencing revealed that the transcription factors share common regulatory functions and positively regulate biological processes related to signal transduction, cell communication and biological adhesion. ChIP-sequencing on histone tail post-translational modifications revealed that the differentially expressed genes were enriched for both active and repressive chromatin signatures at their promoters, suggesting that they are dynamically regulated and might therefore contribute to connective tissue differentiation. Occupancy of each transcription factor was finally investigated via ChIP-sequencing to distinguish between indirect and direct target genes. Altogether, the combination of *in vivo* and *in vitro* data with genome-wide profiling brings molecular insights underlying the differentiation of connective tissue cells. These results provide a framework for future investigations to better understand the interconnectivity between components of the musculoskeletal system.

RÉSUMÉ

Le système musculo-squelettique se compose des muscles, du squelette et du tissu conjonctif qui comprend, entre autres, les tendons et le tissu conjonctif musculaire. La locomotion est assurée par ce système via une coordination précise entre ses différents composants. Le tissu conjonctif musculaire contribue à l'élasticité et à la rigidité des muscles, alors que les tendons transmettent les forces musculaires à l'os nécessaires aux mouvements du corps. Malgré leurs origines mésodermiques distinctes, la différenciation des cellules progénitrices du muscle et du squelette est étroitement liée tout au long du développement des membres. Contrairement au muscle et au squelette, la mise en place et la formation du tissu conjonctif restent à ce jour peu étudiées. Afin d'identifier les mécanismes moléculaires sous-jacents à la formation du tissu conjonctif au cours du développement du membre, cinq facteurs de transcription à doigt de zinc ont été examinés : OSR1, OSR2, EGR1, KLF2 et KLF4. Ces facteurs de transcriptions sont exprimés dans différents sous-compartiments du système musculo-squelettique et leur surexpression influence la différenciation des cellules mésenchymateuses. Afin d'élucider leurs rôles au niveau de la régulation génique, plusieurs stratégies à haut-débit ont été mises en place dans des cultures de cellules mésenchymateuses de membres surexprimant chacun des facteurs de transcription. Le séquençage du transcriptome global a révélé que les facteurs de transcription partagent des fonctions régulatrices communes liées à la transduction du signal, à la communication cellulaire et à l'adhésion cellulaire. L'analyse par séquençage à haut-débit de modifications post-traductionnelles des queues d'histone a montré que les gènes différentiellement exprimés étaient enrichis pour des signatures d'activation et de répression chromatiniennes, suggérant qu'ils sont dynamiquement régulés et qu'ils pourraient contribuer à la différenciation du tissu conjonctif. Les sites de fixation à l'ADN ont été finalement étudiés par séquençage à grande échelle afin de distinguer les gènes cibles directs des cibles indirectes. En résumé, la combinaison de données *in vivo* et *in vitro* avec des méthodes de profilage à haut-débit a permis de mettre en évidence les mécanismes moléculaires sous-jacents à la différenciation des cellules du tissu conjonctif. Ces résultats fournissent une base pour des travaux futurs visant à mieux comprendre l'interconnectivité entre les différents composants de l'appareil locomoteur.

ZUSAMMENFASSUNG

Das muskuloskeletale System besteht aus den Muskeln, dem Skelett und verschiedenen Bindegeweben wie den Sehnen und dem Muskel-Bindegewebe. Die präzise Koordination zwischen den verschiedenen Komponenten dieses Systems ist eine Grundvoraussetzung für die Funktionalität des Bewegungsapparates. Das Bindegewebe der Muskeln trägt zu Elastizität und Rigidität des Muskels bei, die von den Muskeln generierte Kraft wird von den Sehnen auf das Skelett übertragen, um Bewegung zu ermöglichen. Obwohl Muskeln und das Muskel-Bindegewebe der Extremität einen unterschiedlichen mesodermalen Ursprung in der Entwicklung haben, ist ihre Entwicklung eng aneinander gekoppelt. Während die Entwicklung von Muskeln und Skelett intensiv beforscht wurden, ist über die Entwicklung des Bindegewebes sehr wenig bekannt. Um molekulare Mechanismen der Bindegewebsentwicklung zu identifizieren, wurden in dieser Arbeit fünf verschiedene Transkriptionsfaktoren untersucht: OSR1, OSR2, EGR1, KLF2 und KLF4. Diese Transkriptionsfaktoren werden in verschiedenen Bereichen des muskuloskeletalen Systems exprimiert und ihre Überexpression beeinflusst die Differenzierung von mesenchymalen Zellen. Um ihre molekulare Rolle zu charakterisieren, wurden verschiedene Genom-weite Analysestrategien durchgeführt, wofür Explantatkulturen von mesenchymalen Vorläuferzellen der Extremität verwendet wurden. Eine Transkriptomanalyse per RNA-Sequencing zeigte, dass alle fünf Transkriptionsfaktoren bestimmte gemeinsame regulatorische Funktionen miteinander teilen; so regeln sie biologische Prozesse, die mit der Signaltransduktion, der Zell-Zell Kommunikation und der biologischen Adhäsion assoziiert sind. ChIP-Sequencing auf verschiedene Histonmodifikationen zeigte, dass die Promotoren der differentiell regulierten Gene für spezifische aktivierte wie auch repressive Chromatinmodifikationen angereichert waren. Dies deutet auf dynamisch regulierte Gene hin, und somit wahrscheinlich auf Gene, die in die Differenzierung des Bindegewebes involviert sind. Die direkte Bindung der Transkriptionsfaktoren an die DNA wurde schließlich durch ChIP-Sequencing analysiert, um zwischen direkten und indirekten Zielgenen unterscheiden zu können. Zusammengefasst erlaubt die hier unternommene Kombination aus *in vivo* und *in vitro* Analysen in Kombination mit Genom-weiten Methoden einen ersten molekularen Einblick in die Komplexität der Differenzierung des Bindegewebes. Die hier vorgestellten Ergebnisse stellen damit einen Ausgangspunkt für weitere Arbeiten dar, um die Interaktion der verschiedenen Komponenten des muskuloskeletalen Systems besser zu verstehen.

NOTATIONS AND ABBREVIATIONS

3F	Triple-FLAG	lncRNA	Long non-coding RNA
AER	Apical ectodermal ridge	MCT	Muscle connective tissue
bHLH	Basic helix-loop-helix	MRF	Myogenic regulatory factor
CDS	Coding sequence	ncRNA	Non-coding RNA
ChIP	Chromatin immunoprecipitation	NSC	Normalized strand correlation
ChIP-seq	ChIP combined with massively parallel DNA sequencing	ORF	Open reading frame
chMM	Chicken micromass	padj	Benjamini-Hochberg adjusted p-value
DE	Differentially expressed	PCA	Principal components analysis
ECM	Extracellular matrix	PIC	Pre-initiation complex
eRNA	enhancer RNA	Pol II	RNA polymerase II
FAP	Fibro-adipogenic	PTM	Post-translational modification
FDR	False-discovery rate	PROMPT	Promoter upstream transcript
FRiP	Fraction of reads in peaks	RCAS	Replication-competent ASLV long terminal repeat with a splice acceptor
galGal4	Fourth version of the <i>Gallus gallus</i> reference genome	RNA-seq	Whole-transcriptome sequencing
GO	Gene ontology	RSC	Relative strand correlation
GTF	General transcription factor	TAD	Topologically associated domain
HH	Hamburger-Hamilton	TF	Transcription factor
HOT	High-occupancy target	TFBS	Transcription factor-binding sites
IDR	Irreproducible discovery rate	TPM	Transcripts per million
ISH	<i>In situ</i> hybridization	TSS	Transcriptional start site
		UTR	Untranslated region
		ZPA	Zone of polarizing activity

A. From a single unit to a multi-compartment system

In tetrapods, limb development initiates by expansion of lateral plate mesoderm cells at specified positions along the rostrocaudal axis of the embryo following an epithelial to mesenchymal transition (Gros and Tabin 2014; Figure 1A). Lateral plate progenitor cells will give rise to the skeleton, tendons and muscle connective tissue, whereas limb myogenic progenitor cells originate from another mesoderm compartment, the somites. Both cell populations accumulate under the ectodermal tissue to form a circular bulge, the limb bud, and then continue their progression distally while the limb grows and develops.

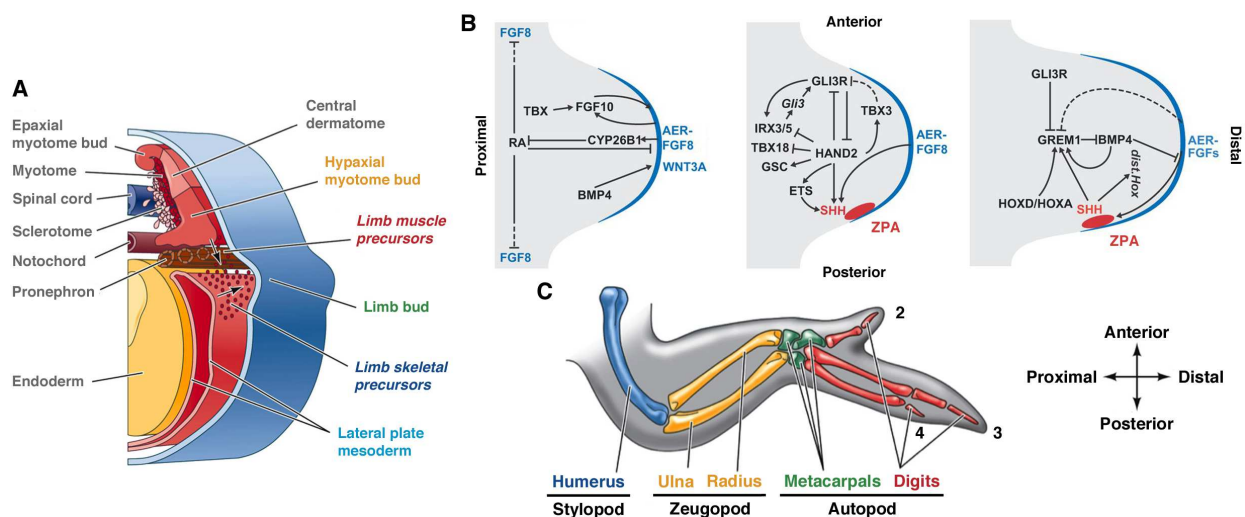


Figure 1. Development of the vertebrate limb. (A) Limb bud formation. Muscle progenitor cells originate from the hypaxial lip of the dermomyotome and migrate towards the limb field to reach lateral plate mesoderm-derived progenitor cells. Both cell populations accumulate and proliferate beneath the ectoderm to shape an emerging protrusion, the limb bud. Lateral plate mesodermal and somitic progenitor cells will constitute the distinct elements of the limb musculoskeletal system. (B) Gene regulatory networks mediating the early patterning of the mouse developing limb outgrowth. The proximodistal axis is established by a gradient between the expression of retinoic acid (RA) from the trunk and the distal apical ectodermal ridge (AER; blue). The anteroposterior patterning is controlled by the anterior zone of polarizing activity (ZPA; red) via a gradient of SHH signalling. Mesenchyme is depicted in grey. Dashed lines indicate unclear molecular mechanisms. (C) Skeletal elements of the chick forelimb. Chick wing skeleton is partitioned into the proximal stylopod (humerus; blue), intermediate zeugopod (radius, ulna; yellow) and distal autopod (metacarpals, green; digits, red). By convention, the chick forelimb digits are numbered 2, 3 and 4 along the anteroposterior axis. Adapted from Gilbert 2013; Zuniga 2015.

Limb development is coordinated in three dimensions along the proximal-distal (shoulder-finger; hip-toe), anterior-posterior (thumb-pinkie) and dorsal-ventral (knuckle-palm) axes (Figure 1B), although the time could be included as fourth dimension. The growth along the proximodistal axis is under the influence of the apical ectodermal ridge (AER), via the secretion of the fibroblast growth factors FGF4 and FGF8, which induces the proliferation of mesenchymal cells located beneath the AER (reviewed in Tabin and Wolpert 2007). The anteroposterior

specification is controlled by the secretion of Sonic hedgehog (Shh) from the zone of polarizing activity (ZPA) (Riddle et al. 1993), which confers the digit identity through a timely- and spatially-dependent concentration gradient (reviewed in Tabin and McMahon 2008). The dorsoventral polarity is determined by the ectoderm through the dorsal expression of *Wnt7a* (Dealy et al. 1993; Parr et al. 1993) and the ventral expression of *En1* (Engrailed-1) (Logan et al. 1997; Loomis et al. 1996). The development of the three limb axes is coordinated at the cellular and molecular levels, which influences the correct polarity and patterning of the different components of the musculoskeletal system and leads ultimately to a fully functional limb (reviewed in Rabinowitz and Vokes 2012; Zuniga 2015).

Skeleton development

The skeleton of tetrapod limbs consists in the assembly of bone and cartilage elements via joints and is divided into three regions along the proximal-distal axis: the stylopod (humerus, femur), adjacent to the body wall; the zeugopod (radius, ulna; tibia, fibula), as intermediate region; and the autopod (carpal, fingers; tarsal, toes), at the distal extremity (Figure 1C). *In vivo* experiments in chick embryos have demonstrated that the AER is essential in the establishment of the limb skeleton. Indeed, AER removal at different time points induces the loss of skeleton elements along the proximodistal axis. Early AER ablation results in proximal truncations, while late AER removal affects only distal elements (Saunders 1948). In addition to their role during cell differentiation (reviewed in Cerdá-Esteban and Spagnoli 2014; Le Guen et al. 2015), *Hox* genes contribute also to the specification of the three skeleton elements along the proximodistal axis, especially *Hoxa* and *Hoxd* paralogous gene clusters. Similarly to the collinear expression within the main body axis, expression from *Hoxa9/Hoxd9* to *Hoxa13/Hoxd13* genes follows a proximal to distal pattern during early limb development that recapitulates the stylopod, zeugopod and autopod (reviewed in Zakany and Duboule 2007). Indeed, specification of the stylopod is under the control of *Hox9* and *Hox10* paralogous genes, whereas *Hoxa11* and *Hoxd11* genes are involved in the specification of the zeugopod, and *Hox12* and *Hox13* paralogous genes in the patterning of the autopod. At later stages of limb development, the production of Shh by the ZPA induces the expression pattern of *Hoxd10-13* genes following the anterior-posterior axis that contributes to the digits morphogenesis (Zakany et al. 2004). The Shh receptor Ptch1 appears to be an important mediator of Shh diffusion during digit patterning since its attenuated expression has been recently linked to the lost of digits in artiodactyls (Cooper et al. 2014; Lopez-Rios et al. 2014). Shh activates the bone morphogenetic protein (BMP) antagonist Gremlin (*Grem1*) to downregulate BMP signalling and to maintain the expression of *Fgf8* in the AER (Benazet et al. 2009). Although SHH signalling mediates the establishment of distal-posterior limb elements

(ulna, fibula; digits 2-5), proximal-anterior structures (stylopod; radius, tibia; digit 1) are *Shh* independent and require the expression of the Iroquois homeobox transcription factors (TFs) *Irx3* and *Irx5* (Harfe et al. 2004; Li et al. 2014).

Formation of limb skeleton along the three developmental axes requires the concomitant differentiation of lateral plate mesoderm-derived mesenchymal cells into cartilage. The endochondral ossification is the process consisting in the commitment of mesenchymal cells into chondrocytes to form cartilage that will be further replaced by bone tissue (Figure 2).

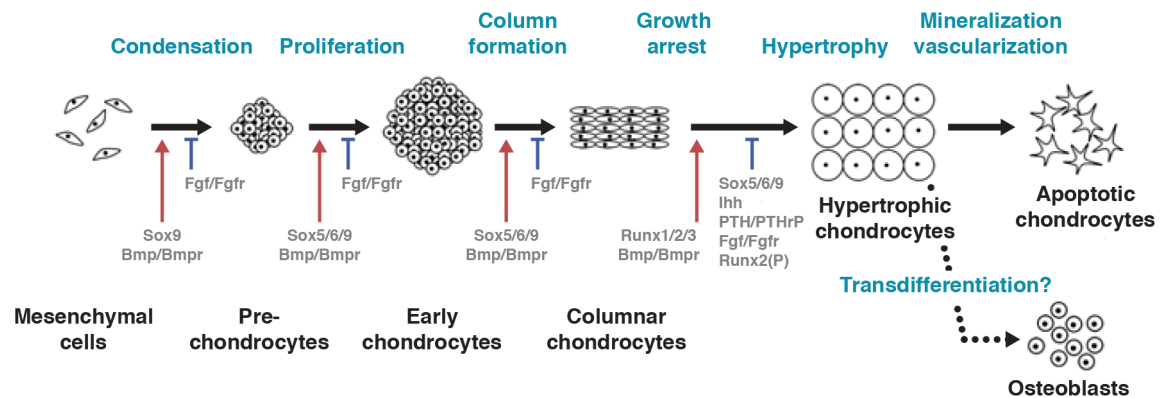


Figure 2. Schematic representation of the chondrogenesis process during limb development. Endochondral ossification during limb development is a multistep process requiring the compaction of mesenchymal progenitor cells into precartilaginous nodules, the proliferation of chondrocytes until reaching a hypertrophic state and the replacement of apoptotic hypertrophic chondrocytes by osteoblasts. Adapted from Shimizu et al. 2007.

First, mesenchymal progenitor cells aggregate in the middle of the limbs into compact nodules forming precartilaginous condensation regions that are characterized by a strong expression of the SRY-box TF *Sox9* (Zhao et al. 1997). Inner cells commit towards a cartilage lineage, whereas outer cells remain at first sight undifferentiated to form perichondrium, a layer of dense irregular connective tissue. Initiation of cartilage condensation regions requires the expression of N-cadherin and N-CAM adhesion molecules (Oberlender and Tuan 1994; Widelitz et al. 1993) and signal transduction via the BMP signalling receptor BMPRII (Zou et al. 1997). Inner cells undergoing chondrogenesis start proliferating and secrete a cartilaginous extracellular matrix (ECM) composed of type-II, -IX and -XI collagens and aggrecan. Later, chondrocytes stop dividing to reach a hypertrophic state while expressing Indian hedgehog (*Ihh*) and parathyroid hormone-related protein (*PTHrP*) to control their maturation (Vortkamp et al. 1996). The maturation of chondrocytes into hypertrophic chondrocytes is positively regulated by the Runt TF *Runx2* (Stricker et al. 2002). Hypertrophic chondrocytes elongate and produce type-X collagen that will contribute to the bones to reach their final size. In addition, they secrete the vascular growth factor VEGF that will be required for the formation of blood vessels (Gerber et al. 1999), as well as *Ihh* that will induce perichondrium cells to differentiate into osteoblasts via *Runx2* expression. Finally, hypertrophic chondrocytes undergo apoptosis and are gradually

replaced by osteoblasts recruited via blood vessels (Hatori et al. 1995). In parallel, perichondrial osteoblasts form a bone matrix and express the zinc-finger TF Osterix (also known as Sp7) via WNT signalling to differentiate into osteocytes (Hu et al. 2005; Nakashima et al. 2002).

Signalling pathways are critical during the chondrogenesis process. BMP signalling is crucial for the compaction of undifferentiated mesenchymal cells into precartilaginous nodules and throughout their commitment into chondrocytes (Yoon et al. 2005). Overexpression of *Noggin*, a BMP antagonist (Zimmerman et al. 1996), blocks chondrogenesis and prevents the formation of cartilage in chick limbs (Pizette and Niswander 2001). In addition, the proliferation pace of chondroblasts is tightly controlled by a crosstalk between BMP and FGF signalling pathways (Minina et al. 2002), whereas chondrocyte hypertrophy results from a synergetic balance between BMP and Ihh/PTHrP signalling cascades (Long et al. 2004; Zhang et al. 2003). WNT signalling has been associated with contradictory effects during chondrogenesis. For instance, Wnt3a accelerates BMP2-mediated chondrogenesis (Fischer et al. 2002), whereas Wnt1, Wnt5a and Wnt7a inhibit cartilage formation (Rudnicki and Brown 1997; Tufan and Tuan 2001). However, other observations pointed out an opposite effect of Wnt3a and Wnt5a on chondrogenesis. Wnt3a has been shown to inhibit chondrogenesis *in vitro* and *in vivo* and to downregulate *Sox9* expression (ten Berge et al. 2008), while overexpression of Wnt5a enhances cartilage nodule formation in micromass cultures (Church et al. 2002). In addition, activation of the canonical WNT/ β -catenin signalling pathway in chick chondrocytes accelerates their hypertrophic maturation (Kitagaki et al. 2003), whereas conditional deletion of β -*catenin* in mice depicted a delayed endochondral ossification (Akiyama et al. 2004).

Skeletal muscle development

In vertebrates, limb skeletal myogenic cells originate from the somites. Limb muscle development involves the successive steps of migration, organisation into dorsal and ventral muscle masses, differentiation and muscle splitting (Figure 3). Signals produced by the limb lateral plate mesoderm trigger muscle progenitor cells to delaminate from the lateral edge of the dermomyotome and migrate towards the limb buds (Chevallier et al. 1977; Christ et al. 1977; Hayashi and Ozawa 1995; Jacob et al. 1978; Solursh et al. 1987). Muscle progenitor cells express the paired box TF Pax3 and Pax7 (Relaix et al. 2005). Delamination and migration of myogenic progenitor cells are under the control of several molecules. Migration of muscle progenitor cells requires the Pax3-dependent expression of the ladybird homeobox TF Lbx1 (Brohmann et al. 2000; Gross et al. 2000; Mennerich et al. 1998; Schäfer and Braun 1999). The scatter factor (SF; also known as hepatocyte growth factor, HGF), induced by FGF signalling (Heymann et al. 1996), is expressed in limb mesenchyme and interacts with its receptor c-Met

expressed by myogenic progenitor cells to trigger their delamination and migration (Bladt et al. 1995; Daston et al. 1996). The migration is also under the control of the chemokines Cxcl12 (also known as Sdf1) and EphA5, which are both expressed in the limb mesenchyme, and that interact with their respective receptors Cxcr4 and EphA4 expressed in *Pax3*⁺ cells (Vasyutina et al. 2005; Swartz et al. 2001). Therefore, the migration of myogenic progenitor cells is also dependent of extrinsic signals produced by limb mesenchyme.

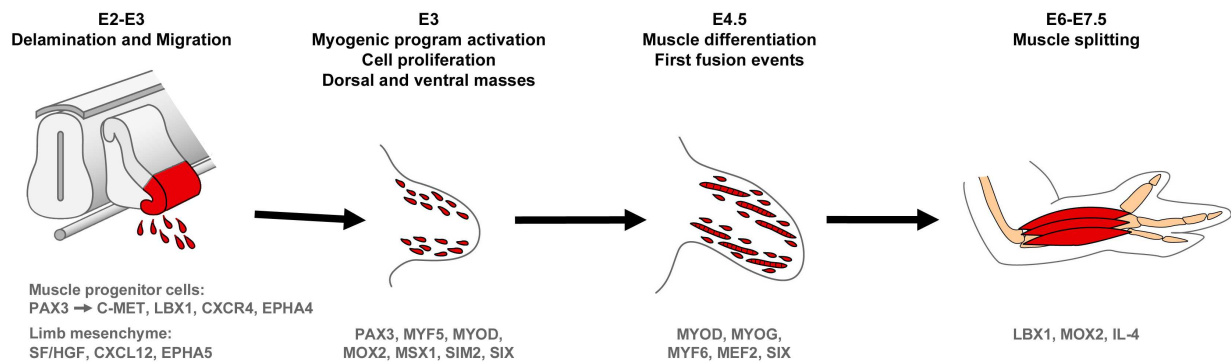


Figure 3. Schematic representation of the myogenesis process during chick forelimb development. Myogenesis initiates by delamination of muscle progenitor cells from the hypaxial lip of the somitic dermomyotome in response to signals from the adjacent lateral plate mesoderm. Myogenic progenitor cells migrate towards the limb field, enter within the limb bud and split into ventral and dorsal premuscular masses. Myogenesis depends on a fine tuning between proliferation and differentiation triggered by activation of the MRFs. The first multinucleated myotubes are visible at E4.5 in the chick embryo wing. Finally muscle masses undergo cleavage and splitting to give rise to the different limb muscles. Genes involved or potentially involved at each step are depicted. Adapted from Buckingham et al. 2003; Duprez 2002.

After migration, *Pax3*⁺ progenitor cells undergo the myogenic regulatory program under the influence of the myogenic regulatory factors (MRFs). The MRFs are the master genes of the skeletal muscle lineage, which include the basic helix-loop-helix (bHLH) TFs Myf5, MyoD, myogenin and Mrf4 (also known as Myf6). Overexpression of each of these factors commits a non-muscle cell towards a myogenic lineage, while preventing other cell fates (Auradé et al. 1994; Choi et al. 1990; Davis et al. 1987; Delfini and Duprez 2004). First signs of muscle differentiation are observed at E5 in chick limbs. Limb muscle growth relies on a fine balance between cell proliferation and differentiation. Pax3 and Pax7 are involved in muscle progenitor proliferation in addition to Myf5 and MyoD that control this process by a cell-cycle dependent mutually exclusive expression (Collins et al. 2009; Kitzmann et al. 1998). High level of *Myf5* expression is associated with proliferating primary mouse myoblasts, whereas high expression of *MyoD* induces cell cycle withdrawal and differentiation initiation *in vitro*. In chick embryos, *MYF5* expression is related to proliferating myoblasts, whereas *MYOD* is not detected (Delfini et al. 2000). In addition, Notch signalling pathway is involved in the maintenance of proliferating myogenic progenitor cells (Delfini et al. 2000; Zalc et al. 2014). Transcriptional repressors have been also associated with the regulation of the tuning between proliferation and differentiation of

myoblasts. *Msx1* is expressed in the lateral dermomyotome of the somites and in migrating *Pax3*⁺ cells and its repression is followed by *Myf5* activation (Houzelstein et al. 1999). *Msx1* overexpression in differentiated mouse myotubes induces their dedifferentiation and reversion to a proliferative state (Odelberg et al. 2000), while *Myf5* displays a reduced level of facultative heterochromatin in its vicinity and a higher expression level in *Msx1*^{-/-} mutant mice (Wang et al. 2011a). Another transcriptional repressor, Sim2 (single-minded 2), which is mainly expressed in ventral muscle masses of mouse and chick embryo limbs (Coumailleau and Duprez 2009), represses *MyoD* expression by direct binding to its enhancer region and prevent muscle differentiation to allow a preferential muscle growth in ventral limb regions (Havis et al. 2012).

The dorsal and ventral muscle masses are composed of proliferating Pax3/Pax7 progenitor cells and committed muscle cells, the myoblasts. Myoblast differentiation is marked by their exit from cell cycle followed by the expression of *MyoD* and the synthesis of muscle-specific proteins such as actin and myosin. Myogenin is also involved in muscle differentiation (Andrés and Walsh 1996; Bergstrom and Tapscott 2001). Committed myoblasts align together to form multinucleated myofibres by fusion of their membranes. The molecular mechanisms underlying myoblast fusion are not completely resolved in vertebrates. However, myoblast fusion is mediated by ECM components such as cell adhesion molecules (N-, V-CAM), fibronectin, integrins, cadherins and meltrins (reviewed in Hindi et al. 2013). Following the first fusion of myoblasts, myofibres keep growing by recruiting further myoblasts. Myoblast recruitment is mediated by the secretion of paracrine factors from newly formed myofibres, such as interleukin-4 (IL-4) (Horsley et al. 2003). Recently, the identification of myomaker (also known as TMEM8C) as a main actor triggering myoblast fusion opens new avenues to understand this biological process (Millay et al. 2013). Muscle masses undergo progressive splitting to give rise to the individual limb muscles. Although this process remains unclear, innervation and vascularization are thought to be involved in the cleavage process of muscle masses as it has been shown in chick embryos (Rong et al. 1992; Tozer et al. 2007).

To a similar extent with skeleton morphogenesis, signalling pathways contribute to muscle formation throughout limb development. Overexpression of SHH in chick limbs at the onset of the myogenic program induces muscle hypertrophy (Amthor et al. 1998; Duprez et al. 1998), whereas homozygous deletion of *Shh* in the mouse gives rise to a severe reduction of muscle formation (Krüger et al. 2001). Migration and activation of the myogenic program are independent of Shh activity as these both processes occur normally in *Shh*-mutant mice (Krüger et al. 2001). In addition to contribute to both delamination and migration of muscle progenitor cells via the scatter factor, FGF signalling have been associated with the activation of myoblast

proliferation while inhibiting their differentiation *in vitro* (reviewed in Olson 1992). Expression of FGF2, FGF4 and FGF8 in the limb AER prevents muscle differentiation by inhibiting *MYOD* transcription in order to maintain the distal migration of muscle progenitor cells (Robson and Hughes 1996). Similarly to skeleton development, WNT signalling has been characterized as a promoter and as an inhibitor of myogenesis. In mouse somitic explant cultures, *Wnt1* activates *Myf5* expression, whereas *Wnt7a* induces *MyoD* transcription (Tajbakhsh et al. 1998). *In vitro* experiments have highlighted that β -catenin negatively regulates myogenic differentiation (Gavard et al. 2004; Goichberg et al. 2001), whereas a recent *in vivo* study showed that β -catenin positively regulate the number of foetal muscle progenitor cells (Hutcheson et al. 2009). Although being mostly known for its role in bone and cartilage development, BMP signalling is also involved during myogenesis. In combination with FGF signalling and the scatter factor, BMP4 prevents the differentiation of myogenic progenitor cells during their migration (Pourquié et al. 1996; Amthor et al. 1998). At later stages of limb development, during foetal myogenesis, BMP signalling upregulates the expression of *PAX3* and *MYOD* (Amthor et al. 1999) and promotes foetal muscle growth (Wang et al. 2010).

Tendon development

Tendon is a dense regular connective tissue composed of type-I collagen fibres extending laterally along the tendon axis and embedded in dense irregular connective tissue sheaths. Limb tendon cells originate from the lateral plate mesoderm (Kieny and Chevallier 1979; Figure 4).

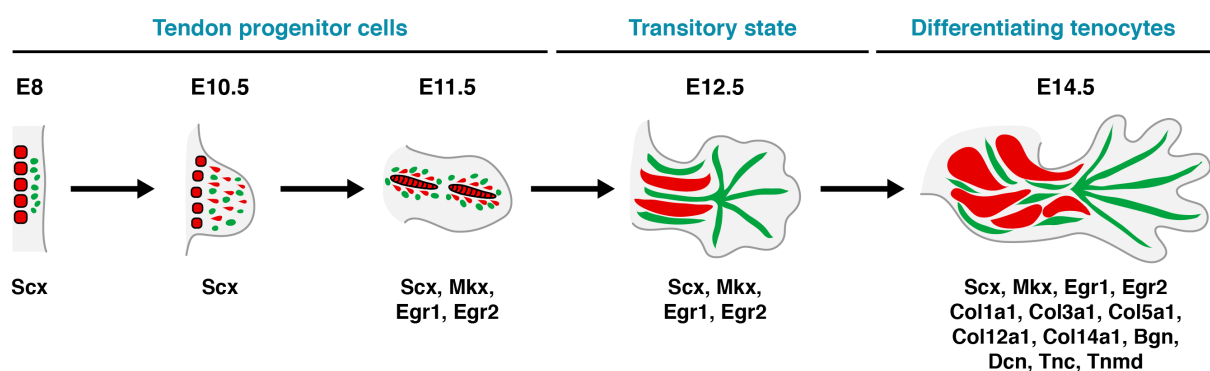


Figure 4. Schematic representation of the tenogenesis process during mouse limb development. *Scx*-expressing cells originate from the lateral plate mesoderm and subsequently subdivide into three tendon primordia. Each primordium gives rise to the distinct tendons of the knee/elbow, heel/wrist and hand/foot joints. Genes expressed, involved or potentially involved in tendon development are indicated. Muscles and tendons are depicted in red and green, respectively. Adapted from Gaut and Duprez 2016.

In contrast to myogenesis and chondrogenesis, tendon development, also termed tenogenesis, remains poorly described. The identification of an early marker of tendon progenitor cells, the bHLH TF *Scleraxis* (*Scx*), facilitated the understanding of tendon development (Schweitzer et al. 2001). Ectopic expression of *Scx* in human bone marrow-derived mesenchymal stem cells

induces the expression of tendon-associated genes and prevents their differentiation into chondrocytes and osteoblasts, although they retain their capacity to commit towards the adipogenic lineage (Alberton et al. 2012). Conversely, *Scx*-deficient mice present severe defects in tail and limb force-transmitting tendons, ranging from their complete loss to the disorganization of their structure, while anchoring tendons of the back remain weakly affected (Murchison et al. 2007).

In the developing limb, *Scx*-expressing primordium is marked by two distinct cell populations: *Scx*⁺/*Sox9*⁻ progenitor cells and *Scx*⁺/*Sox9*⁺ progenitor cells (Sugimoto et al. 2013b). The latter cells commit into either tenocytes by maintaining *Scx* and repressing *Sox9*, or chondrocytes by downregulating *Scx* and sparing *Sox9*, at places where tendon-bone junctions will take place (Blitz et al. 2013; Sugimoto et al. 2013a). In the chick, tendon progenitor cells are organized into three primordia along the proximal-distal axis of the hindlimb (Kardon 1998). Proximal primordium gives rise to the future tendons of the knee (thigh-shank), intermediate primordium to the future tendons of the intertarsal joint (shank-foot), and distal primordium to the future tendons of the metatarsal-phalangeal and interphalangeal joints. *Scx* positively regulates the expression of type-I collagen, the main functional component of tendon (Léjard et al. 2007; Murchison et al. 2007). It is noteworthy that tendon specificity is given by the spatial organisation of type-I collagen fibrils and not by its expression itself, since this collagen is found in many other tissues and organs. Besides *Scx*, three additional TFs have been identified as being expressed in developing tendons: *Mkx*, *Egr1* and *Egr2* (Anderson et al. 2006; Lejard et al. 2011). Although these TFs are not tendon-specific, ectopic expression of *Mkx* or *Egr1* in mesenchymal stem cells promotes the expression of tendon-associated genes and collagens as well as impairs their commitment towards other cell fates (Guerquin et al. 2013; Liu et al. 2015b; Otabe et al. 2015). *Mkx*-null mice exhibit tendon differentiation defects visible by a reduction of tendon collagen fibrils and smaller tendons (Ito et al. 2010; Kimura et al. 2011; Liu et al. 2010). Homozygous deletion of *Egr1* in the mouse induces a disorganization of tendon collagen fibrils, reduced tendon mechanical properties and an impaired healing following tendon injury (Guerquin et al. 2013). A recognized tendon marker for tenocytes is tenomodulin (*Tnmd*) (reviewed in Dex et al. 2016). *Tnmd* encodes a glycoprotein that is present in differentiated tendon cells and necessary for their proliferation and maturation (Docheva et al. 2005). Recently, two global gene expression profiles of *Scx*⁺ cells isolated from mouse limbs between E11.5 and E14.5 have been generated (Havis et al. 2014; Liu et al. 2015a). Both studies identified a valuable resource of genes regulated during limb tendon development that will necessarily contribute to extend the comprehension of the tenogenesis process.

In addition to TFs, two main signalling pathways, FGF/MAPK/ERK and TGF- β /Smad2/Smad3, have been associated with tendon specification and differentiation based on their influence on *Scx* expression in chick and mouse embryos. Ectopic expression of *FGF4* in chick forelimbs upregulates *SCX* and *TNC* expression (Edom-Vovard et al. 2002). *TNC* is a gene that encodes tenascin, an ECM protein that is associated, among others, with developing tendons (Hurle et al. 1990). However, application of FGF4 in mouse limb explants or murine mesenchymal stem cells downregulates *Scx* expression, suggesting a different effect of FGF signalling in chick and mouse (Havis et al. 2014). Additionally to this finding, the authors observed an increased *Scx* expression upon inactivation of MAPK/ERK signalling by chemical reagent in both culture models. TGF- β signalling has been identified as the most enriched pathway in a gene expression profiling of mouse *Scx*⁺ cells isolated from limbs by FACS sorting (Havis et al. 2014). Treatment by TGF- β 2 of mouse limb explants or murine mesenchymal stem cells increases the level of *Scx* expression and tendon-associated genes (Havis et al. 2014; Pryce et al. 2009). By contrast, chemical blocking of Smad2/Smad3 intracellular signalling pathway in mouse limb explants displays the opposite pattern with a strong downregulation of *Scx* transcription (Havis et al. 2014). Consistently, application of TGF- β 2 in chicken micromass cultures made of limb mesenchymal cells promotes the expression of *SCX* and *TNMD*, while inhibiting chondrogenesis, whereas chemical inactivation of SMAD2 and SMAD3 reverts the cell fate commitment (Lorda-Diez et al. 2009). In addition, disruption of TGF- β signalling in the mouse by deleting either *Tgfb2* alone, both *Tgfb2* and *Tgfb3* genes, or by inactivating the type-II TGF- β receptor TGFBR2, leads to severe limb tendon defects from E12.5 (Pryce et al. 2009). BMP signalling seems to act in an opposite manner to TGF- β signalling by restricting *Scx* expression, while inactivation of BMP signalling via the BMP antagonist Noggin upregulates *Scx* expression (Schweitzer et al. 2001).

Muscle connective tissue development

Muscle connective tissue (MCT) belongs to the category of the connective tissue proper. It is a loose irregular connective tissue that is arranged as an interwoven meshwork of irregular fibres. MCT was described as being able to drive non-muscle cells to depict muscle-like structures (Grim and Wachtler 1991). As cells forming the skeleton and tendons, limb MCT cells are derived from the limb lateral plate mesoderm (Chevallier et al. 1977; Christ et al. 1977). However, in contrast to cartilage, bone and skeletal muscle, but similar to tendon, no master regulator for MCT formation has been identified yet. Although several TFs have been characterized as being related to MCT during limb development, they all seem to be expressed in distinct or partly overlapping MCT cell populations. The component of WNT signalling *TCF4*

was the first described MCT marker gene (Kardon et al. 2003). Overexpression of a dominant negative form of TCF4 in chick limbs induces both muscle patterning and myofibre organization defects. TCF4 is among the most downstream effectors of the canonical WNT/ β -catenin signalling pathway (Korinek et al. 1997). Ectopic activation of the WNT/ β -catenin signalling in chick legs leads to the appearance of additional muscles as compared to contralateral limbs, indicating that *TCF4*-expressing cells from the MCT may contribute to create a resident pattern by extrinsic signals for muscle development (Kardon et al. 2003; Figure 5). The role of the WNT/ β -catenin pathway was also evidenced in the axolotl, where overexpression of active β -catenin induces ectopic muscle formation in the regenerating limb (Nacu et al. 2013). *Tcf4* exhibits a similar expression pattern during mouse limb development, as well as in neonatal and postnatal mice (Mathew et al. 2011). Conditional homozygous deletion of *Tcf4* induces severe limb muscular defects and co-cultures of *Tcf4*-expressing fibroblasts with myoblasts enhances myotube formation as compared to myoblasts cultured alone. In addition, *Tcf4*⁺ cells contribute to muscle regeneration upon injury by dynamic and reciprocal interactions with myogenic satellite cells since ablation of either satellite cells or *Tcf4*⁺ fibroblasts results in impaired muscle healing (Murphy et al. 2011). Nonetheless, *Tcf4* is not only restricted to MCT cells as it is also detected in tenocytes, chondrocytes and at low levels in myogenic cells.

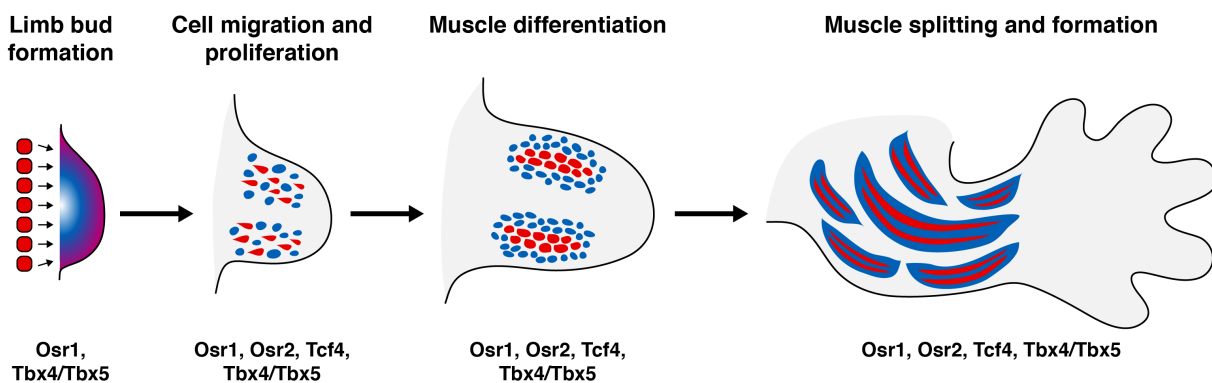


Figure 5. Schematic representation of muscle connective tissue formation during limb development. Muscle progenitor cells delaminate and migrate towards the limb field in response to signals coming from the adjacent lateral plate mesoderm. Muscle connective tissue progenitor cells migrate in the limb bud in close association to muscle progenitor cells and lie adjacent to proliferating and differentiating muscle cells. Genes expressed or involved or potentially involved are indicated. *Tbx4* and *Tbx5* are expressed in the hindlimb and forelimb, respectively. Muscles are depicted in red; MCT is represented in blue. Adapted from Kardon et al. 2003.

More recently, two T-box TFs, *Tbx5* and *Tbx4*, have been associated with forelimb and hindlimb MCT, respectively. Both factors are broadly expressed in the mesenchyme prior to limb bud initiation and throughout chick and mouse limb development (Hasson et al. 2007; Logan et al. 1998; Naiche and Papaioannou 2007; Rallis et al. 2003). Conditional deletion of either *Tbx5* or *Tbx4* in limb mesenchyme leads to severe muscle and tendon defects (Hasson et al. 2010). Although muscle and tendon specification and differentiation remains unaffected, mutant mice display muscle splitting and tendon defects. Interestingly, forelimb and hindlimb

skeleton elements remain unaffected upon deletion of either *Tbx5* or *Tbx4*, respectively, suggesting that skeleton patterning occur independently of muscle/tendon patterning. Muscle and tendon patterning seem to be dependent on WNT signalling and biological adhesion since β -catenin and N-cadherin were strongly reduced in *Tbx5* and *Tbx4* mutant embryos. Two additional zinc-finger TFs, *Osr1* and *Osr2*, have been identified as being related to limb MCT. Prior to chick limb bud initiation, *OSR1* expression is detected in the lateral plate mesoderm suggesting that it is expressed in mesenchymal progenitor cells (Stricker et al. 2006). Later during limb development, *Osr1* and *Osr2* are widely expressed and partly overlapping in limbs of mouse and chick embryos. Interestingly, their expression patterns are closely associated with myogenic cells during limb development although both *OSR* factors are not expressed in muscle cells (Stricker et al. 2012). Ectopic expression of *Osr1* and *Osr2* in undifferentiated myogenic cells or bone marrow stromal cells reduces their capacity to differentiate into myotubes or osteoblasts, respectively. Retroviral overexpression of *Osr1* in chick hindlimb induces a reduction of tendons and skeleton elements. In addition, double mutant mice for *Osr1* and *Osr2* display increased cartilage differentiation in synovial joints and abnormal joint fusion (Gao et al. 2011). *Osr1*-deficient mice exhibit mispatterning of individual muscles and myotendinous junctions as well as increased expression levels of tendon- and cartilage-associated genes (Vallecillo García et al., in revision). Altogether, *Osr1* and *Osr2* seem to control the differentiation of mesenchymal cells into irregular connective tissue.

Muscle interaction with other compartments of the musculoskeletal system

The musculoskeletal system gives the ability for an organism to move by a precise coordination between its components. Forces generated by muscle contraction are transmitted to the skeleton via connective tissue to allow body motion. Connective tissue is one of the main constituents of the body that supports, connects and separates tissues and organs from each other. The term of connective tissue gathers together an ensemble of tissues such as tendon and MCT, which are intimately linked to muscles. In the developing limb, MCT cells are positioned surrounding muscles but also interstitial to the muscle fibres and play a role of binder providing elasticity and rigidity for muscle function. Throughout limb development, lateral plate-derived connective tissue cells are temporally and spatially associated with myogenic cells. Each individual muscle is characterized by its own organization, fibre orientation and composition, while expressing a similar gene program. Resident tissues at places where muscles will form are then good candidates to contribute to the correct muscle patterning. Indeed, emerging evidence have highlighted that interactions between muscle and connective tissue are crucial for proper limb patterning and morphogenesis.

Muscle-bone interaction. Muscle development requires bone formation and modifications of skeleton induce changes in muscle formation. Limb skeleton develops autonomously and is uncoupled of soft tissue patterning (Li et al. 2010). However, the arrangement of muscles, orientation of muscle fibres and integration of tendons must be related to some extent to the skeletal elements for proper organization of the limb musculoskeletal system. Indeed, manipulations of chick limb buds clearly evidenced that musculature and skeletal structures are closely associated. For instances, ZPA grafts to the anterior margin of limb buds lead to duplicated skeletal structures and muscles (Duprez et al. 1999; Robson et al. 1994). Likewise, reversion of the dorsal-ventral axis in chick limb bud by rotation of the ectoderm results in the dorsoventral respecification of both skeleton and muscles (Akita 1996). In addition, IHH signalling, which is important during chondrogenesis, have been lately associated with muscle patterning. In chick embryos, downregulation of *IHH* results in muscle and skeletal defects in hindlimbs, similarly to those observed in *Ihh*-deficient mouse embryos, while restoration of SHH signalling rescues muscle masses (Bren-Mattison et al. 2011). As previously mentioned, *Hox11* paralogous genes function in the establishment of the zeugopod during limb development. *Hoxa11*; *Hoxd11* double mutant mice exhibit skeleton, muscle and tendon mispatterning in the zeugopod (Swinehart et al. 2013). However, when only one of both paralogous genes is deleted, either *Hoxa11*^{-/-} or *Hoxd11*^{-/-}, patterning defects remain exclusively visible for muscles and tendons, while zeugopod skeleton appear normal. Whether muscle and skeleton development occurs autonomously, proper patterning of both tissues seem related since mechanical forces driven by muscles contribute to bone shape, growth and fracture healing as well as joint formation (Hall and Herring 1990; Hosseini and Hogg 1991; Kahn et al. 2009; Rot et al. 2014; Sharir et al. 2011).

Muscle-tendon interaction. The interface between muscle and tendon is named the myotendinous junction and consists in interdigitations of the basement membranes of tenogenic and myogenic cells through interactions between tendon collagen fibrils and muscle laminins or integrins (Bökel and Brown 2002; Tidball and Lin 1989). Reciprocal interplay between muscles and tendons that are necessary for the correct patterning of both tissues have been highlighted (Soler et al. 2004; Wan et al. 2012; Figure 6). In chick and mouse limbs, late tendon development requires the presence of muscles (Bonnin et al. 2005; Edom-Vovard et al. 2002; Eloy-Trinquet et al. 2009; Huang et al. 2015; Kardon 1998). The absence of muscles induces defects in the segregation of tendon primordia into individual tendons leading to their degeneration (Huang et al. 2015; Kardon 1998). Reciprocally, surgical removal of tendon primordia at the early onset of hindlimb development causes the appearance of ectopic muscle in the knee of chick embryos (Kardon 1998). FGF signalling seems to be involved in tendon-

muscle interplay (Edom-Vovard et al. 2001). In chick embryos, *FGF4* is expressed in muscles, close to myotendinous junctions. Overexpression of *FGF4* induces an upregulation of *SCX* and *TNC* expression, and vice versa, downregulation of *FGF4* is followed by downregulation of *SCX* expression. FGF signalling modulators *Pea3*, *Spry1* and *Spry2* are expressed at myotendinous junctions in both mouse and chick embryo limbs, and their expression is dependent on muscles and *FGF4* activation (Eloy-Trinquet et al. 2009). However, if late tendon development is muscle-dependent, early tendon development does not require the presence of muscles. Induction of *Scx* expression in the developing limb occurs normally in *Myod*^{-/-}; *Myf5*^{-/-} and *Pax3* mutant mice (Brent et al. 2005; Schweitzer et al. 2001), as well as in muscleless chick limbs (Edom-Vovard et al. 2002; Kardon 1998). Therefore, although initiation of muscle and tendon patterning occurs independently of each other, maintenance and establishment of muscles and tendons are closely related and seem to require spatial, temporal and mechanical interactions between both tissues. Indeed, it has been recently characterized that correct positioning of the flexor digitorum superficialis muscle in the mouse forelimb arises after differentiation and is dependent on both muscle contraction and tendons (Huang et al. 2013). In *Drosophila*, similar muscle-tendon interactions exist for flight motility. In contrast to the mesodermal origin in vertebrates, tendons (and exoskeleton) originate from the ectoderm and contribute to the connection between the exoskeleton and flight muscles. Tendons initiate their development independently of muscles but require muscle attachment to pursue their patterning at later stages (reviewed in Schweitzer et al. 2010; Soler et al. 2016).

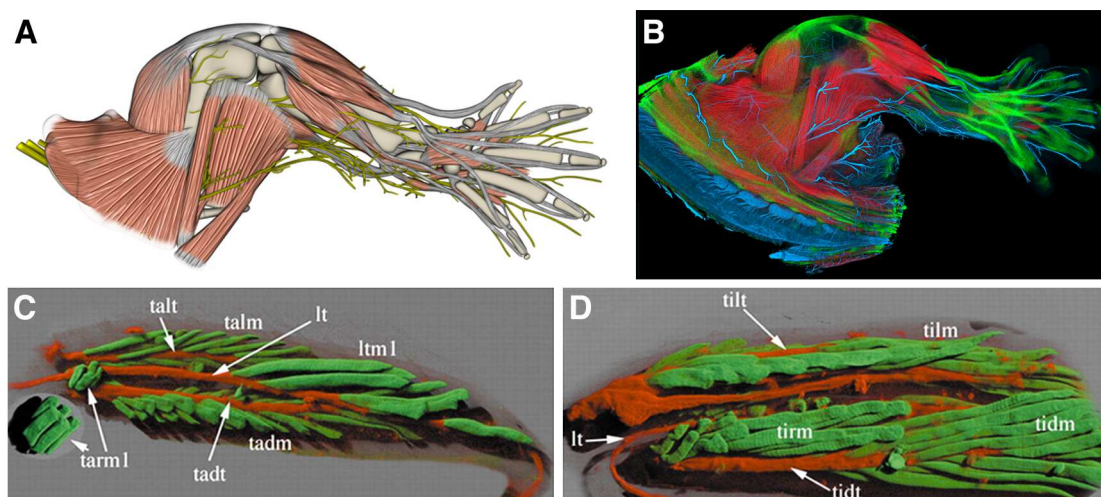


Figure 6. Representation of the limb musculoskeletal system in the mouse and the fruit fly. (A,B) Lateral views of the musculoskeletal system of the mouse embryonic hindlimb at E14.5. Muscles are labelled in red; tendons in green; nerves in blue. (C,D) Three-dimensional reconstruction of appendicular muscles and tendons of the tibia (C) and the femur (D) in *Drosophila* 55h after pupae formation. Muscles are depicted in green; tendons in red. In contrast to vertebrates, muscles are anchored in the exoskeleton in the fly. Extracted from Soler et al. 2004; Wan et al. 2012.

Muscle-MCT interaction. Despite their distinct embryonic origin, spatial organisation and differentiation of MCT cells occur in close proximity to myogenic cells. Although MCT develops normally in muscleless limbs, emerging evidences have highlighted that muscle patterning is under the influence of MCT (Grim and Wachtler 1991; Kardon et al. 2003). Ectopic expression or mutation of the MCT markers *Tcf4*, *Tbx4*, *Tbx5*, *Osr1* or *Osr2* leads to patterning defects of limb skeletal muscles (Hasson et al. 2010; Kardon et al. 2003; Mathew et al. 2011; Murphy et al. 2011; Stricker et al. 2012; Vallecillo García et al., in revision). This highlights the requirement of MCT for skeletal muscle formation. Reciprocally, MCT function requires signals from muscles. For instance, type-VI collagen is an ECM protein found in muscle endomysium and is a recognized marker of MCT. Mutations in the genes *COL6A1*, *COL6A2* or *COL6A3* are associated with myopathies and muscular dystrophy (reviewed in Lampe and Bushby 2005). Activation of the enhancer required for *Col6a1* expression in MCT cells is dependent on muscle cells, since its expression is absent in muscleless limbs (Braghetta et al. 2008). In addition, lysyl oxidase (Lox), which is an extracellular copper enzyme contributing to the formation of collagens and elastin (Mäki 2009), is expressed in myofibres in the mouse developing limb (Kutchuk et al. 2015). *Lox*-deficient mice exhibit shorter and smaller skeletal muscles, as well as disorganized MCT depicting type-I collagen fibres with increased diameter and a higher concentration of *Tcf4* (Kutchuk et al. 2015). Although the molecular interplay between MCT and muscle remains poorly characterized, reciprocal interactions between both tissues seem to be required for proper limb morphogenesis. A recent transcriptomic approach performed on isolated mouse *Osr1*-expressing cells revealed that *Osr1* positively regulates genes encoding chemokines, cytokines, ECM proteins and adhesion molecules (Vallecillo García et al., in revision). By analogy with the previous observations related to *Tcf4*, *Tbx4* and *Tbx5* misexpression, MCT may contribute to establish a resident environment during limb development.

B. From a single gene to a multi-layered network

All cells constituting a multicellular eukaryotic organism share the same genetic information. However, the choice of differentiation taken by a progenitor cell encompasses dramatic phenotypic changes that differ between the hundreds of different cell types within a multicellular organism. For instance, mesenchymal progenitor cells derived from the lateral plate mesoderm or the somites will give rise to the different tissues of the limb musculoskeletal system, such as muscle, cartilage, tendon or MCT. All these cells possess the same DNA sequence in their nucleus but the retrieval of this information underlies a precise temporal and spatial control. Expression of the genetic information is therefore achieved by the realization of specific

developmental regulatory programs taking place that will drive the cells to follow the appropriate differentiation process. These developmental programs consist in a fine tuning between the repression and the expression of a given set of genes in response to certain extrinsic or intrinsic signals, at a specific location and/or at a precise time. The orchestrated process leading to the activation or inactivation of a gene cannot therefore rely only on the binding or the release of the RNA polymerase II (Pol II), respectively. It rather involves a plethora of checkpoints ranging from the chromatin structure to the base composition along the gene locus ultimately leading to the expression of a given gene.

Regulation of gene transcription

Gene expression is defined according to two biological processes, namely transcription and translation, each being dependent on an additional constant, the mRNA and protein degradation rate, respectively (Figure 7). The abundance of proteins is a balance between protein half-life and the number of proteins synthesized from the transcripts, which are themselves dependent on their half-life and the rate of gene transcription. A recent study has quantified protein and transcript abundances for ~5,000 protein-coding genes in mouse fibroblasts revealing a median transcription rate of 2 mRNA molecules per hour (Schwanhäusser et al. 2011). Although mRNA and protein half-lives are not correlated as compared to the number of transcripts vs. number of proteins, genes with similar mRNA and protein stabilities share common biological functions. Indeed, genes with both stable mRNAs and proteins are generally involved in housekeeping functions, whereas genes with stable mRNAs but unstable proteins mainly encode secreted proteins, adhesion molecules and ECM components. By contrast, TFs and chromatin-modifying enzymes have a short half-life and are synthesized from unstable transcripts, which is consistent with their dynamic function in gene regulation.

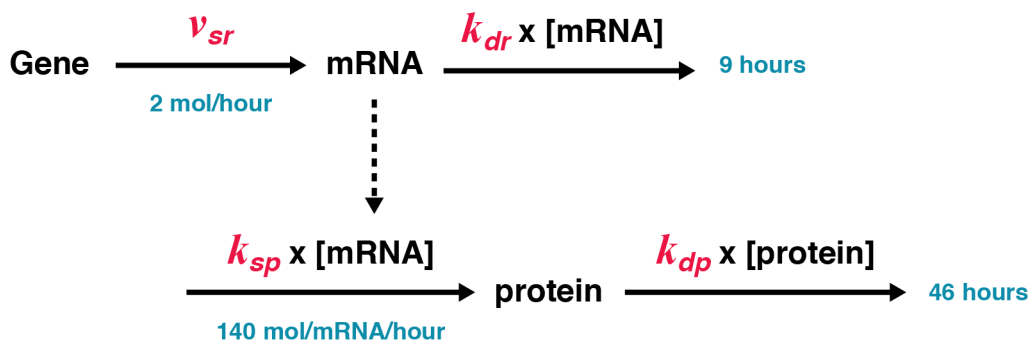


Figure 7. Quantitative model of gene expression. Transcripts are synthesized with the transcription rate v_{sr} and degraded according to the constant k_{dr} . Protein abundances are balanced between the translation rate k_{sp} and the degradation constant k_{dp} . Median synthesis rates and half-lives (blue) are indicated according to Schwanhäusser et al. 2011.

The first step of mRNA transcriptional initiation is marked by the recruitment of the pre-initiation complex (PIC) at the promoter (Figure 8A). The PIC consists in a multiprotein complex formed between the Pol II, a 12-subunit enzyme, and the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIIE, TFIIIF and TFIIH (reviewed in Sainsbury et al. 2015). The Pol II undergoes afterwards an elongation phase to synthesize the mRNA until termination, which corresponds to the release of the transcript from the Pol II. Repetition of these three steps then determines the gene expression levels. However, emerging evidence have highlighted that mRNA transcription is not a linear process and is rather precisely regulated at each step in order to ensure the proper timing of mRNA synthesis. One of the sophisticated mechanisms limiting the transcription rate is associated with the pausing of the Pol II 30-60 nucleotides downstream of the gene transcriptional start site (TSS; Figure 8B). Although several reports on a limited number of genes evidenced blocked Pol II at promoters (reviewed in Adelman and Lis 2012), the first genome-wide analysis of Pol II occupancy in mammalian cells revealed that accumulation of the PIC at the gene TSS is a promoter-proximal common feature (Kim et al. 2005). Later, a study in human embryonic stem cells correlated Pol II accumulation with chromatin signatures associated with transcriptional activity, suggesting that paused Pol II had previously initiated transcription (Guenther et al. 2007). The notion of pausing was finally demonstrated by a genome-wide sequencing of nascent transcripts in human lung fibroblasts since paused Pol II had the ability to re-initiate transcription *in vitro* (Core et al. 2008). Pol II pausing is broadly observed at genes with a wide range of expression levels in metazoans, suggesting that it is a general pausing mechanism to control the expression of active genes rather than an inactivation mechanism (reviewed Adelman and Lis 2012). Stability of Pol II promoter-proximal pausing seem to be mediated by the core promoter elements (Kwak et al. 2013; Li and Gilmour 2013), and by the cooperation of TFs with the negative elongation factor (NELF) and the DRB-sensitivity-inducing factor (DSIF) (reviewed in Adelman and Lis 2012). In addition to control gene expression levels, Pol II pausing is thought to keep the chromatin in an opened state accessible for TFs by preventing nucleosome rearrangement as well as to facilitate 5'-end capping of nascent transcripts prior to productive elongation (reviewed in Adelman and Lis 2012). Release of paused Pol II is under the control of the positive transcription elongation factor b (P-TEFb) that phosphorylates the Pol II carboxy-terminal domain, NELF and DSIF, thus freeing the Pol II for productive mRNA synthesis (reviewed in Jonkers and Lis 2015; Figure 8C).

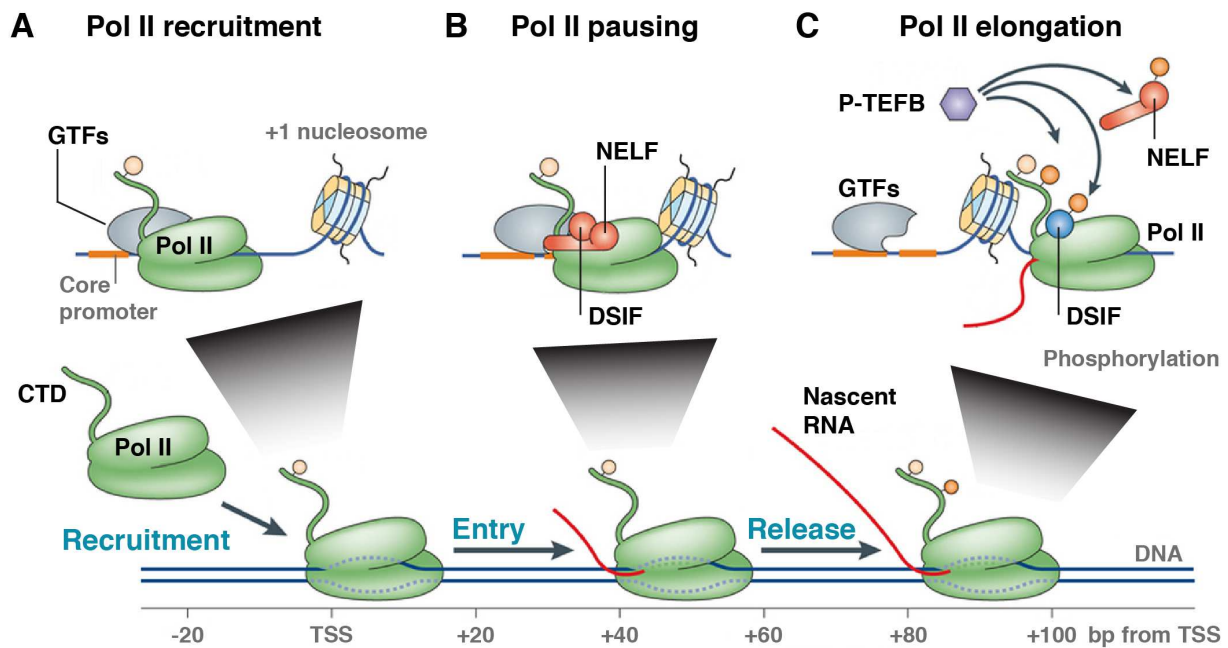


Figure 8. Schematic representation of initiation of gene transcription. (A) Recruitment of the Pol II at the core promoter elements by the GTFs resulting in the formation of the PIC. (B) Transcription initiation followed by Pol II pausing 30-60 bp downstream of the gene TSS mediated by the factors NELF and DSIF. (C) Release of paused Pol II for productive elongation following phosphorylation of the Pol II carboxy-terminal domain, NELF and DSIF orchestrated by P-TEFB. CTD, carboxy-terminal domain; GTF, general transcription factor; TSS, transcriptional start site.

Transcript synthesis through elongating Pol II is not a unidimensional process and the elongation rate is subjected to variations. Indeed, elongation speed is rather non-optimal at the transcription initiation of mammalian genes. Gradual accumulation and post-translational modifications (PTMs) of transcriptional machinery components seem to enhance the elongation rate to reach its maximum after a few kilobases (reviewed in Jonkers and Lis 2015). In addition, impediments encountered by the Pol II along the gene body induce elongation rate fluctuations, resulting in a gene-dependent transcription rate. Exon density, GC-content and long-terminal repeats are overall negatively correlated with the elongation rate, whereas long genes with low-complexity sequence and distant from other simultaneously transcribed genes are generally associated with faster elongation rates (Jonkers et al. 2014; Veloso et al. 2014). Reduced Pol II elongation rate at exons seems to be related to alternative splicing and exon skipping (reviewed in Jonkers and Lis 2015). Once having reached the polyadenylation signal, the Pol II pauses to seemingly increase the efficiency of the transcriptional termination process (reviewed in Porrua and Libri 2015). In metazoans, transcription termination corresponds to the release of the nascent mRNA transcript mediated by a cleavage complex composed of CPSF, CstF, CFI and CFII, followed by the dismantling of the elongation complex. Pol II pausing appears to be associated with formation of RNA-DNA hybrids (referred as R-loops) and further recruitment of the RNA exonuclease XRN2, which degrades the 3'-end of the nascent transcript remaining after cleavage.

DNA regulatory elements

To elucidate the complexity behind gene expression, the US National Human Genome Research Institute (NHGRI) initially launched in 2003 the Encyclopaedia of DNA Elements (ENCODE) project. This worldwide consortium intended to identify the functional elements that composed the human genome besides the 1.22% of the DNA sequence encoding proteins (ENCODE Project Consortium 2012). By 2012, the ENCODE consortium had generated over 1,600 data sets covering 147 different human cell types and claimed that 80.4% of the human genome is functional. This controversial term of “functional” engendered an unprecedented debate among the scientific community. Indeed, the definition of functionality is contradictory with the long-standing notion that the human genome is mainly composed of non-coding regions, the so-called “junk DNA” (Ohno 1972). The DNA elements thus defined as functional by the ENCODE consortium corresponds to the regions of DNA that have a biochemical function, which is distinguishable from the evolutionary functional regions (Germain et al. 2014). In other words, the genome does not carry only functional elements giving rise to proteins, but also numerous regulatory elements that contribute to gene expression and repression. However, to the same extent as for limb development, gene regulation processes ultimately leading to gene expression or repression remain only partially resolved.

Besides exonic sequences, the genome is composed of multiple domains that ensure the regulation of gene expression. These control elements can be divided into two main categories: promoters and enhancers. Basically, promoters correspond to domains located near the TSSs where Pol II binds and initiates transcription. Tetrapod promoters can be divided into two main classes depending on their base composition: GC- and AT-rich promoters (Carninci et al. 2006; Yamashita et al. 2005). GC-rich promoters are characterized by the presence of CpG islands and are related to both ubiquitous and differentially regulated genes. In contrast, AT-rich promoters contain the TATA-box and are frequently associated with tissue specificity (Barrera et al. 2008). Enhancers are *cis*-regulatory elements that can be located in close proximity but also distant to their targeted gene and that contribute to regulate the transcriptional activity independently of their position, orientation and distance (Banerji et al. 1981, 1983). Long-range regulation between a distal enhancer and its target gene is exemplified with the *SHH* locus. Mutation in an enhancer located nearly 1 Mb upstream of *SHH* is responsible for preaxial polydactyly in human (Lettice et al. 2003). Deletion of this enhancer in the mouse results in truncated limbs (Sagai et al. 2005). Likewise, *Sox9* and *Myf5* expression are also under the control of distant-acting enhancers (Carvajal et al. 2001; Wunderle et al. 1998).

Promoters and enhancers are marked by the recruitment of protein effectors that bind directly to DNA (activators) or indirectly through protein-protein interactions (coactivators). The repertoire of genes composing the human genome encodes for approximately 7-10% of DNA-binding protein effectors (Maston et al. 2006). TFs are protein effectors that are typically composed of three domains: (i) a DNA-binding domain, which recognizes a specific sequence motif; (ii) a trans-activating domain, which ensures the activation or silencing function; and (iii) a protein-protein interaction domain, which allows the cooperation with other proteins. TFs are gathered together in families based on the structural similarities within their DNA-binding domain such as the presence of a homeodomain (Hox, Pax), a basic helix-loop-helix structure (MyoD, myogenin) or a zinc-finger domain (Krüppel, Engrailed). The TF specificity of DNA binding is conferred by a short and degenerate recognition motif of 6-20 bp (reviewed in Luscombe et al. 2000). In general, the degeneracy level of recognition site does not prevent TFs to bind but influences the binding affinity of the TF for a given DNA sequence (Badis et al. 2009). However, TF recognition sites contain usually 4-6 constrained positions that truly dictate specificity since their modification or mutation can alter TF binding (Ibrahim et al. 2013; Wienert et al. 2015). In addition, TF-TF cooperativity is frequently observed with the formation of homo- or hetero-dimers between related TFs but also between TFs of distinct structural families (Jolma et al. 2015). The resulting recognition site is then usually composed of two half-motifs, which can be different from the preferential binding motif of each interacting partner. In *Drosophila* for instance, Extradenticle (Exd) interacts with several Hox proteins such as Labial (Lab) and Ultrabithorax (Ubx) during body axis patterning. The heterodimers thus formed recognize distinct consensus sequences that differ from the specific motif when they bind individually to DNA, which could then contribute to the specificity of Hox factors *in vivo* (Slattery et al. 2011). However, genome-wide analyses of TF occupancy have revealed that a high proportion of binding sites are devoid of a canonical binding motif, indicating that the specificity of TFs does not solely rely on the DNA sequence (Ernst and Kellis 2013). On the other hand, by considering the core binding site of 4-6 bp and the large size of eukaryotic genome, a TF has the potential to bind over hundreds of thousands positions (Wunderlich and Mirny 2009). Instead, TFs occupy a limited number of sites, suggesting that additional constraints must exist to prevent TFs to bind in an uncontrolled manner along the genome. A recent study in the sea squirt *Ciona intestinalis* in regard to the occupancy of the TFs GATA and ETS along the Otx-a enhancer has revealed that enhancer specificity is mediated by a suboptimization of the TF-binding sites (TFBS) (Farley et al. 2015). Indeed, incorporation of higher affinity binding motifs and optimal adjustment of motif spacing resulted in aberrant

enhancer activity and ectopic expression of the gene reporter. Therefore, attenuated enhancer activity through suboptimal TF binding seems to mediate the specificity of gene expression.

Chromatin modelling

The regulation of gene expression depends on the capability of TFs to bind on promoter and enhancer regulatory domains. Therefore, recognition of the specific sequence motif and DNA binding of the TFs rely on the accessibility of the chromatin (Kaplan et al. 2011; Thurman et al. 2012). DNA is packaged into chromatin structures named nucleosomes that are tightly regulated. A nucleosome consists of an octameric protein complex composed of a dual core of the histone proteins H2A, H2B, H3 and H4, wrapped around by ~150 bp of DNA and stabilized together by the histone protein H1 (Luger et al. 1997). The level of compaction of the nucleosome structure defines the chromatin state and by extension the level of gene expression. Condensed nucleosomes, namely heterochromatin, are generally associated with silencing due to a high compaction level that makes the DNA sequence inaccessible. By contrast, euchromatin is rather related to active gene expression owing to its relaxed nucleosome structure thus providing accessibility to DNA sequence. Histone proteins are relatively small and composed of two domains, a globular core region and a N-terminal tail protruding from the nucleosome. Although the globular domain is also dynamically regulated (reviewed in Lawrence et al. 2016), histone tails are subjected to various PTMs such as acetylation, methylation, phosphorylation and ubiquitination (reviewed in Bannister and Kouzarides 2011; Figure 9).

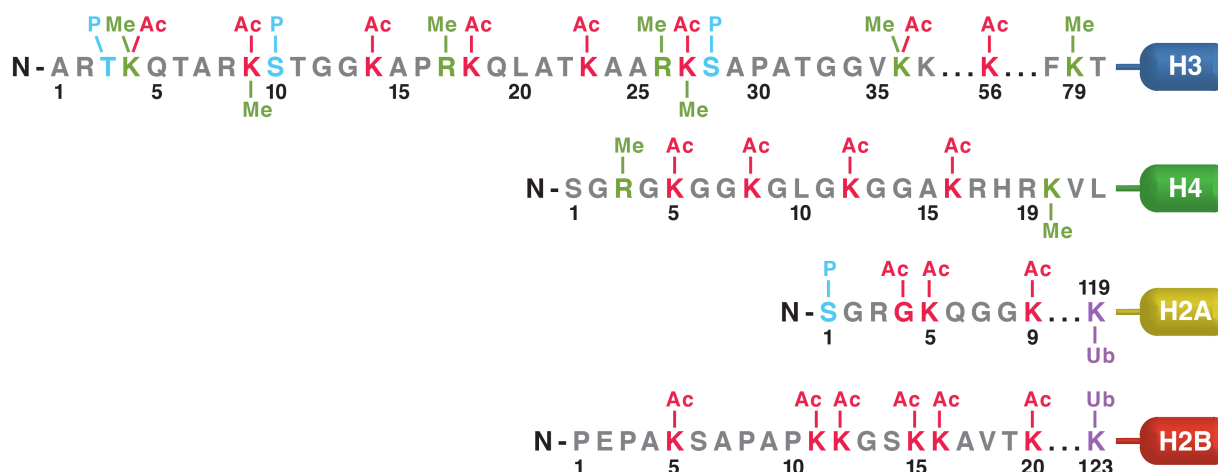


Figure 9. Post-translational modifications of histone protein tails. Overview of the main covalent histone modifications frequently found at the protruding tail of the histone proteins H2A, H2B, H3 and H4. Residues are depicted in grey. Amino-acid positions are depicted in black underneath starting from the N-terminal extremity. Histone PTMs associated with transcriptional activation are indicated above and those associated with transcriptional repression are displayed beneath. Ac, acetylation; Me, methylation; P, phosphorylation; Ub, ubiquitination. Adapted from Hamon and Cossart 2008.

The variety of histone modifications, affected residues and involved histone proteins constitute a code of chromatin structure (re-)modelling (Strahl and Allis 2000). Although this function has

been more associated with the globular domain (reviewed in Tessarz and Kouzarides 2014), PTMs on histone tails can also alter nucleosomal architecture. Acetylation of the lysine 16 of histone H4 (H4K16ac) loosens chromatin structure by preventing nucleosome cross-interactions (Shogren-Knaak et al. 2006), whereas trimethylation of H4K20 (H4K20me3) increases chromatin compaction *in vitro* (Lu et al. 2008). However, such role seems rather exceptional since the complete removal of histone tails does not affect nucleosome stability (Ausio et al. 1989). Histone tail-extension modifications can also promote the recruitment of protein effectors such as chromatin-modifying enzymes and TFs (Clements et al. 2003; Vettese-Dadey et al. 1996), as well as to prevent the access of chromatin remodelling complexes (Shao et al. 1999).

PTM of histone residues is a highly dynamic and reversible process regulated by histone-modifying enzymes. Histone methylation is under the control of histone methyltransferases (HMTs), whereas demethylation is controlled by histone demethylases (HDMs). Deposition of methyl groups can either mark an active, or a repressive chromatin state, depending on the histone residue methylated and the level of methylation (Cosgrove et al. 2004). For example, H3K4 and H3K36 trimethylations (H3K4me3, H3K36me3) are usually associated with euchromatin and active transcription (Bannister et al. 2005; Schneider et al. 2004), whereas H3K9 methylation marks constitutive heterochromatic regions (Noma et al. 2001). Constitutive heterochromatin is found in the vicinity of permanently silenced genes, such as those located near centromeres, or in gene deserts (Rosenfeld et al. 2009). By contrast, facultative heterochromatin is related to genes that are dynamically regulated through addition or removal of methyl groups. H3K27 trimethylation (H3K27me3) is strongly correlated with facultative heterochromatic regions. For example, H3K27me3 is localised at the TSS of *Hox* genes in embryonic stem cells or at early stage of development thus preventing their expression. KDM6A (also known as UTX) and KDM6B (also known as JMJD3) lysine demethylases are then recruited during differentiation and development to catalyse the demethylation of H3K27 residues and allow *Hox* gene transcription (Agger et al. 2007; Lan et al. 2007). In zebrafish, UTX has two homologues UTX1 and UTX2. *Utx1* morphants display defects of the posterior part of the trunk consistent with the moderate expression of the related *hox* genes normally involved in the patterning of the tail (Lan et al. 2007). Unlike histone methylation, deposition of acetyl groups loosens the chromatin structure due to their negative charge. Histone acetylation and deacetylation are mediated by histone acetyltransferases (HATs) and deacetylases (HDACs), respectively. H2BK5, H3K4, H3K9, H3K27 and H3K36 acetylations are for instances highly correlated with transcriptionally active genes (Wang et al. 2008). Additionally, chromatin signatures can directly affect gene transcription. For instances, H3K79me2 and H3K20me1 are correlated with higher elongation rates, whereas H3K36me3 appears to slowdown transcription

speed (Jonkers et al. 2014; Veloso et al. 2014). Likewise, deposition of H3K9me2 is associated with paused Pol II at transcription termination sites (reviewed in Porrua and Libri 2015). Taken together, histone modifications contribute therefore to mediate the accessibility or the obstruction to the DNA elements, and consist in a high-order control of gene expression.

Promoter and enhancer domains

Beyond the distance separating them from their target genes, promoters and enhancers can be distinguished by their local chromatin structure (Figure 10). Indeed, promoters are generally enriched for H3K4 trimethylated (H3K4me3); whereas enhancers are frequently associated with the presence of H3K4 mono-methylated (H3K4me1), the absence of H3K4me3 and binding of the transcriptional coactivator p300 (Barski et al. 2007; Heintzman et al. 2007; Visel et al. 2009). Promoter signatures such as H3K4me3 enrichment and DNase I hypersensitivity sites, which are characteristic of opened chromatin regions, are highly conserved across cell types (Roadmap Epigenomics Consortium et al. 2015; Thurman et al. 2012). Although this is particularly true for promoters associated with broadly-expressed genes, tissue-specific promoters tend to be less conserved across tissues and species (FANTOM Consortium et al. 2014; Young et al. 2015). By contrast, active enhancers are more restricted to certain tissues and species (Andersson et al. 2014). Indeed, a comparison of the chromatin landscape between murine macrophages isolated from different tissues revealed that enhancers recapitulate better the distinct gene expression profiles observed than promoters (Lavin et al. 2014). Another study comparing active promoters and enhancers in liver tissue across 20 mammalian species separated by up to 180 million years converged towards similar observations. Proximal promoters depict a remarkable conservation in terms of H3K4me3 enrichment and underlying DNA sequence, whereas distal promoters rapidly diverged across lineages and species (Villar et al. 2015).

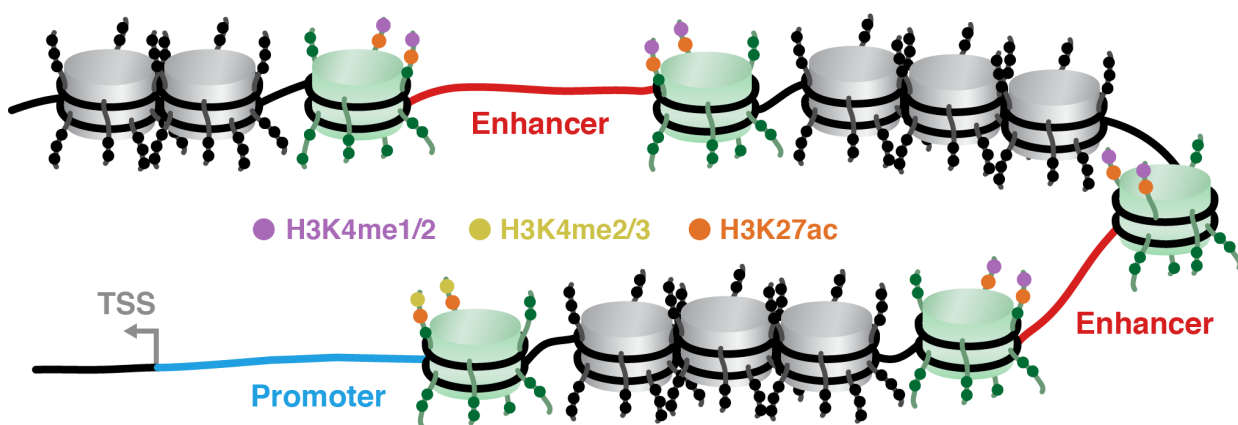


Figure 10. Schematic representation of promoter and enhancer architecture. Promoters are positioned near gene transcriptional start sites (TSSs), whereas enhancers can be located far away from their target gene. Mono- and tri-methylation of H3K4 (H3K4me1, H3K4me3) are highly enriched at enhancer and promoter regions, respectively. Ratio between both chromatin signatures and distance to closest TSS have been widely used to discriminate promoter and enhancer domains. Adapted from Kim and Shiekhhattar 2015.

However, despite their distinct features, promoters and enhancers share several architectural and functional commonalities. A typical example is related to the basic definition of promoters regarding their property to allow binding of the Pol II and further transcription initiation. Two independent genome-wide studies revealed that Pol II can be also recruited at active enhancer regions and initiate the synthesis of non-coding RNAs. These enhancer RNAs (eRNAs) were observed for the first time in neuronal cells and macrophages in response to membrane depolarization and lipopolysaccharide (LPS) stimulation, respectively (De Santa et al. 2010; Kim et al. 2010). Since then, eRNAs have been identified in other mammalian cells, including myoblasts and embryonic stem cells (reviewed in Lam et al. 2014; Mousavi et al. 2013). By using time course experiments, it has been shown that eRNA transcription is the earliest transcriptional event triggered during the commitment of stem cells and the differentiation of developmental progenitor cells (Arner et al. 2015). Besides RNA polymerase II binding, eRNA-producing enhancers share common features with conventional active enhancers such as cell-type specificity and identical, albeit more enriched, histone modifications, including H3K4me1, H3K27ac and H3K79me2 (Djebali et al. 2012). The majority of eRNAs are relatively short (< 2 kb), appear to be unspliced, can be polyadenylated or not and are retained in the nucleus. In addition, eRNA transcription can occur in both directions from the center of the enhancer domain (Kim et al. 2010). Transcriptional bidirectionality, also named divergent transcription, was initially described at promoters where it appeared that reverse-oriented nascent transcripts (referred as PROMPTs) were synthesized upstream of the transcription initiation site (Core et al. 2008; Preker et al. 2008; Seila et al. 2008). Divergent transcripts have a short lifetime due to rapid decay and are thought to prevent nucleosome repositioning and DNA supercoiling, to buffer the downstream gene transcription, or more recently to contribute to TF occupancy (Sigova et al. 2015). Although transcriptional bidirectionality has been considered as a general feature of eukaryotic active promoters, it seems that most of divergent transcripts arise actually from their own reverse-oriented core promoter (Duttke et al. 2015). In addition, intragenic enhancers are capable to serve as alternative promoters to synthesize non-coding transcripts sharing exons with their host gene (Kowalczyk et al. 2012).

By analogy with their transcriptional activities, promoters and enhancers share actually similar histone modifications. Ratio between H3K4me1 and H3K4me3 signals have been widely used to distinguish enhancers from promoters (Heintzman et al. 2009). However, H3K4me3 is also frequently detected at actively transcribed enhancers (Pekowska et al. 2011). The H3K4 methylation state seems most likely to simply reflect the transcriptional activity. Independently of transcript stability, the overall reduced transcription initiation level at enhancers as compared to promoters may contribute to the weaker enrichment of H3K4me3 mark at distal regulatory

regions (Core et al. 2014; Koch et al. 2011). Consistently, H3K79 bi- and tri-methylation (H3K79me_{2/3}), which are observed at actively transcribed gene bodies (Pokholok et al. 2005; Schübeler et al. 2004; Steger et al. 2008), are also detected at enhancers with transcriptional activity (Bonn et al. 2012; Djebali et al. 2012). DNase I hypersensitivity sites correspond to accessible DNA regions resulting from nucleosome depletion and are associated with TF-binding events. Both promoter and enhancer regions are highly correlated with DNase I hypersensitivity (Dong et al. 2012; Thurman et al. 2012). Whereas previous studies have highlighted that interacting promoter-enhancer pairs harbour similar TF-binding events (Bienz and Pelham 1986; Bohmann et al.; Parslow et al. 1987), recent genome-wide analyses have suggested the opposite pattern where bound TFs appear to be distinct between both interacting regulatory domains (Rada-Iglesias et al. 2011; Shen et al. 2012; Thurman et al. 2012). However, the distinct binding site pattern observed may result from the local base composition between promoters and enhancers. Indeed, enhancers display a lower GC content as compared to promoters that are in general enriched for CpG islands. TF-binding motifs identified in GC-low promoters actually resemble to those present in their interacting enhancers (Andersson et al. 2014). Besides promoter-enhancer interactions, high-order chromatin organization also exists through promoter-promoter interactions (Li et al. 2012). Unlike promoter-enhancer interactions that are mainly related to tissue-specific genes, promoter-promoter associations are common to both housekeeping and non-ubiquitous genes. Additionally, genes involved in such combinational associations display similar expression profiles suggesting a transcriptional cooperation. Altogether, these recent genome-wide studies infer that promoters can act as enhancer-like structures and vice versa, enhancers possess promoter-like capacities, which emphasize a high-order complexity to regulate gene expression and repression.

Transcription factor binding

Genome-wide analyses of TF occupancy reported from a few hundreds to several tens of thousands of binding sites per TF along eukaryotic genomes (ENCODE Project Consortium 2012; Johnson et al. 2007; Robertson et al. 2007). A recent *in vitro* study combining TF knockdown and gene expression profiling revealed a weak correlation between TFBS and differentially expressed genes upon TF knockdown (Cusanovich et al. 2014). According to this observation, most of the TFBS would be non-functional since only a few binding events have an impact on gene expression. Several evolutionary studies have attempted to decipher the TFBS functionality. Indeed, if the binding of a given TF at a certain location is functionally relevant, it is reasonable to suppose that the TF-binding sequence is under strong evolution constraint. A comparison of two liver-specific TFs CEPB α and HNF4 α across five vertebrates separated by up

to 80 million years showed a strong divergence in their binding sites although the recognition motif is highly conserved (Schmidt et al. 2010). Most of the discrepancy between species-related TFBS could be explained by the underlying sequence, since 60-85% of binding losses were related to substitution, insertion or deletion events. However, 40-50% of lost TFBS were compensated by another binding event within 10 kb. Similar observations were reported when comparing the occupancy of three hepatocyte TFs across five mouse species separated by up to 20 million years. The more the species are distant the higher is the divergence of the TF-binding profiles (Stefflova et al. 2013). Nevertheless, regions that were bound by several TFs were more conserved across rodent species suggesting a stronger sequence constraint for DNA elements with higher TF occupancy. By contrast, *Drosophila* species depict a generally higher conservation level in their TFBS as compared to mammals (Bradley et al. 2010; He et al. 2011; Paris et al. 2013). Although this could be correlated with a higher frequency of conserved elements across *Drosophila* genomes (37-53% vs. 3-8% for mammals; Villar et al. 2014), to a similar extent with mammalian species, regions with shared TFBS and located near genes were overall more evolutionary conserved (Paris et al. 2013; Wong et al. 2015). Taken together, these studies suggest that the cooperative regulation of tissue-specific TFs and their binding co-dependence depict the TFBS functionality. By contrast, numerous TFBS seem rather non-functional and tend to rapidly evolve. Although it is a matter of debate whether non-functional TFBS are truly non-functional, it has been proposed that these binding sites could contribute to buffer TF dosage (Veitia et al. 2013). On the other hand, extensive investigations on a higher number of TFs indeed revealed that TFs usually tend to bind as clusters in high-occupancy target (HOT) regions (Gerstein et al. 2010; modENCODE Consortium et al. 2010; Moorman et al. 2006; Nègre et al. 2011; Yip et al. 2012). HOT regions appear to be a good indicator of TFBS functionality since they are correlated with interacting loci, Pol II recruitment and spatiotemporal gene expression (Foley and Sidow 2013; Heidari et al. 2014; Kvon et al. 2012). By contrast, low-occupancy target regions tend to be rather non-functional since they mostly fail to drive expression in transgenic gene reporter assays (Fisher et al. 2012).

In addition to chromatin remodelling complexes, nucleosome condensation is mediated by TFs as well, the so-called pioneer factors (Figure 11). These effector proteins are thought to bind and penetrate compacted nucleosomes to facilitate the recruitment of chromatin-modifying enzymes and to pioneer the access of secondary TFs. For instance, MyoD cooperates with the homeobox TF Pbx1 to recruit the SWI/SNF remodelling complex at the promoter of *Myog* for further binding and transcription induction (Berkes et al. 2004; de la Serna et al. 2005). Similarly, Pax7 interacts with the Wdr5-Ash2L-MLL2 histone methyltransferase complex in the vicinity of *Myf5*. Its recruitment induces the trimethylation of H3K4 followed by transcriptional expression

of *Myf5* (McKinnell et al. 2008). In addition, TFs usually require the involvement of co-factors to successively bind to DNA elements or to orchestrate their regulatory potential (Figure 11). TF co-factors do not necessarily bind to DNA and are therefore recruited through protein-protein interactions. In the budding yeast *Saccharomyces cerevisiae* for instance, a complex is formed between the TF Cbf1 and the non-DNA-binding co-factors Met4 and Met28. The resulting complex binds to an extended sequence of the Cbf1-Met4 recognition motif therefore increasing its binding specificity (Siggers et al. 2011). Such latent specificity has been also characterized in hematopoietic cell specification. The non-DNA-binding co-factor Fog1 interacts with Gata1 to positively regulate the differentiation of megakaryocytes and erythrocytes (Tsang et al. 1997), whereas Fog1 acts negatively to the formation of eosinophils and mast cells mediated by Gata1 (Cantor et al. 2008; Querfurth et al. 2000). Disruption of the Gata1-Fog1 complex by mutating one residue of Gata1 prevents the binding of Gata1 at certain locations while enhancing new Gata1 binding sites (Chlon et al. 2012), indicating that a co-factor can act as an activator and/or as a repressor on TF occupancy.

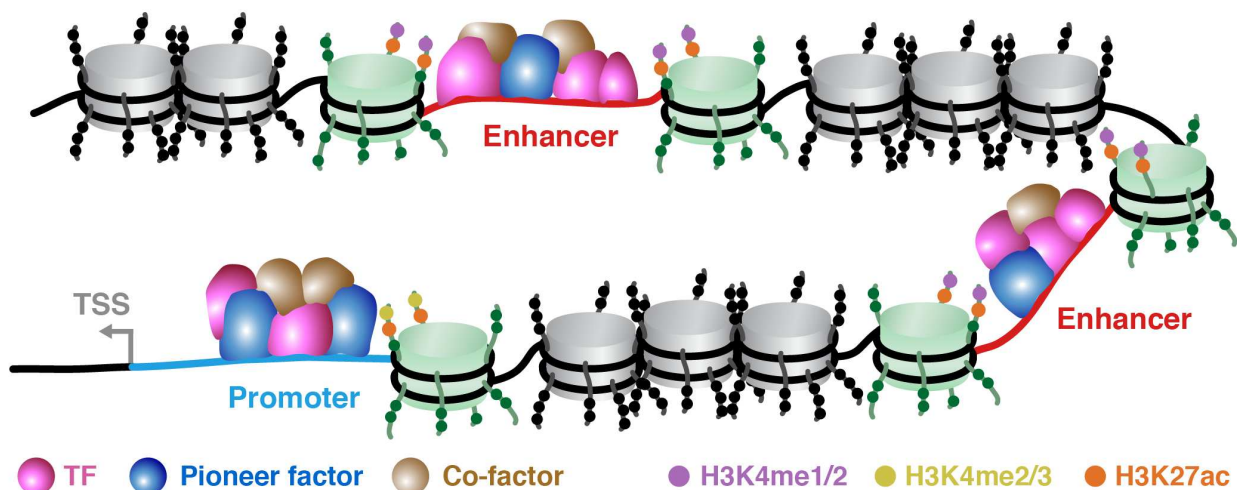


Figure 11. Schematic representation of transcription factor binding at DNA regulatory elements. Accessible promoters and enhancers are prone to the binding of TFs, which are recruited either via DNA-protein interactions by recognizing their own binding motif, or through protein-protein interactions. Pioneer factors (blue) have in general a high occupancy along the genome and promote the recruitment of specific TFs (pink) and co-factors (brown). Adapted from Kim and Shiekhattar 2015.

Hierarchical cooperativity

A time course experiment of murine primary dendritic cells stimulated with lipopolysaccharide (LPS) revealed a gradual molecular response via dynamic TF-binding events that correlated well with further recruitment of the Pol II and transcription initiation (Garber et al. 2012). This study established a hierarchical network composed of three distinct molecular states: (i) a static state corresponding to the pioneer factors Spi1 (also known as PU.1) and Cebp β that were already bound prior to cell stimulation; (ii) a prime response with TFs that rapidly bind upon LPS stimulation near genes that will be further activated; and (iii) a dynamic state associated with a

third wave of TFs that are additionally recruited and that induce the expression of antigen-response genes. Therefore, TFs seem to act in a coordinate manner via successive cooperative recruitments to orchestrate relevant cell-response programs upon environmental cues. It is noteworthy that pioneer factors seem to have a dual role in initiating gene expression. They can mediate chromatin accessibility through binding to nucleosomes and they can maintain nucleosome-free regions while being bound. On the other hand, it seems that transcriptional activity is rather regulated by the timing of TF occupancy than the timing of TF expression. Indeed, a comparison of MyoD occupancy profile between 50%-, 95%-confluent myoblasts and myotubes revealed that most of the MyoD binding sites are shared across all cell populations (Cao et al. 2010a). However, the intensity of MyoD enrichment differed between proliferating myoblasts and differentiated myotubes. For instance, genes involved in muscle formation and function depicted a higher MyoD occupancy in myotubes as compared to myoblasts. Several models have been proposed to illustrate the cooperative TF binding in *cis*-regulatory regions (Figure 12).

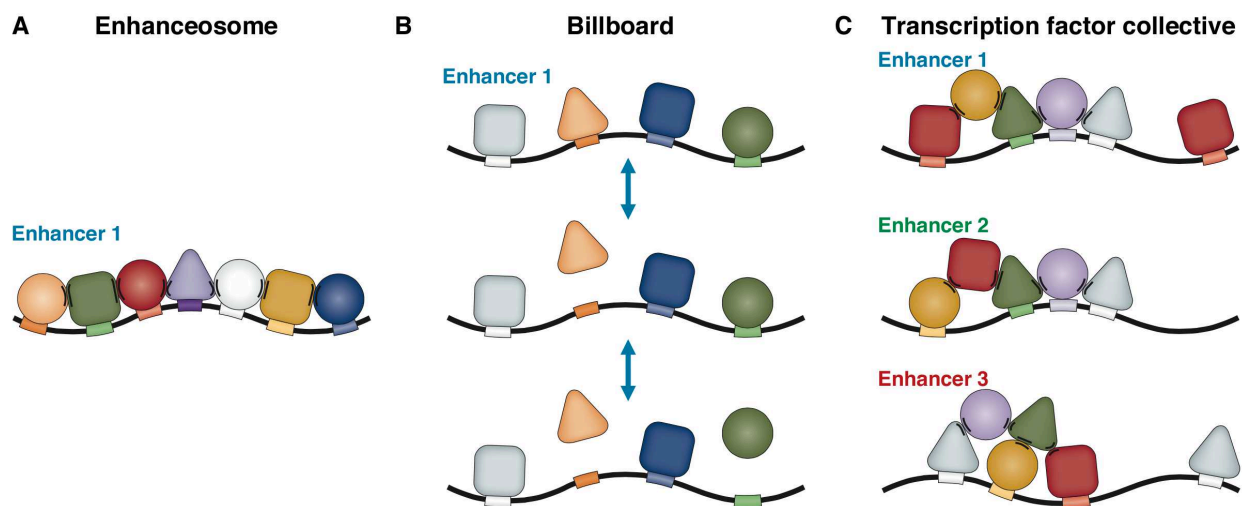


Figure 12. Models of transcription factor-binding at *cis*-regulatory regions. (A) Enhanceosome model. The TF cooperativity requires that all TF-binding motifs are present and spatially organized. (B) Billboard model. The spatial arrangement of the TF-binding motifs is looser and the TF cooperativity is less constraint. (C) TF collective model. The same cluster of TFs can bind to different enhancers independently of both motif composition and positioning. TF cooperativity depends on protein-DNA contacts and protein-protein interactions. Adapted from Spitz and Furlong 2012.

The enhanceosome model relies on the strict presence of the binding motif of all TFs and their appropriate positioning to ensure protein-protein interactions in addition to protein-DNA contacts (Panne et al. 2007; Thanos and Maniatis 1995; Figure 12A). This model involves a strong binding cooperativity between TFs since small sequence alterations disrupt enhancer activity. A more flexible cooperativity has been proposed later with the billboard model, in which some TFs bind cooperatively while others are recruited independently or additionally (Kulkarni and Arnosti 2003; Figure 12B). Therefore, enhancer activity is less restrictive to the spatial organization of the binding motifs and only driven by a subset of bound TFs. However,

accumulating evidence has suggested that enhancer activity is mediated by a very flexible TF motif content and relies on a collective TF recruitment (Junion et al. 2012; Figure 12C). In this model, enhancer regions harbour diverse binding motifs for only a subset of the TFs and the TF recruitment relies on both protein-DNA and protein-protein interactions. A cluster of TFs can then bind to several enhancers independently of the motif composition and removal of one TF reduces or even abrogates enhancer activity. Overall, cooperativity between TFs and their co-factors at specific enhancers occurs in a context-dependent manner, which emphasizes a combinatorial regulatory function to control gene expression (Stampfel et al. 2015).

Physical interaction between distal enhancers and their target genes is rendered possible via DNA folding (Figure 13). Chromatin is indeed organized into spatial subdivisions, the topologically associated domains (TADs), which bring together a gene promoter with its corresponding distal regulatory elements (Dixon et al. 2012). Such high-order chromatin architecture seems to be the first prerequisite to regulate gene expression during cell differentiation and proliferation (de Wit et al. 2013; Naumova et al. 2013). For instance, during mouse limb development, a permutation between two TADs in the *Hoxd* gene cluster coincides with the inversion of expression observed for the subsequent patterning of the arm/forearm and the hand (Andrey et al. 2013). Consequently, TAD structure appears to be a critical regulatory unit to promote enhancer-promoter interaction and it is not surprising that disruptions of TAD have been recently associated with developmental pathologies such as limb syndromes (Lupiáñez et al. 2015). However, chromatin organization alone is not sufficient to drive gene expression since chromosome conformation appears rather stable and static across cell types and during development (Ghavi-Helm et al. 2014; Jin et al. 2013). Therefore, transcription initiation is rather dependent on an accumulation of multiple environmental cues such as three-dimensional interactions, local chromatin states that render regulatory domains accessible, and cooperative binding of a cohort of context-dependent TFs. Although the demarcation of TAD boundaries and the precise mechanism by which enhancers are brought in close proximity to promoters remain mainly unsolved, protein-protein interactions between the activation domain of the TFs and the PIC seem to be mediated by the Mediator, a multiprotein complex composed of 26 subunits in mammals (Figure 13). Mediator subunit composition and structure are highly dynamic, which emphasizes an additional tuning to orchestrate gene expression (reviewed in Allen and Taatjes 2015). The large size of the Mediator complex constitutes an extensive molecular scaffold that transduces regulatory signals between the TFs and the PIC since they do not directly interact. Initiation of transcription is marked by the disruption of Pol II contacts with the Mediator and the PIC via phosphorylation of its carboxy-terminal domain by TFIIF,

followed by its release from the promoter (Kim et al. 1994). Pol II then undergoes a productive elongation phase ultimately leading to mRNA synthesis.

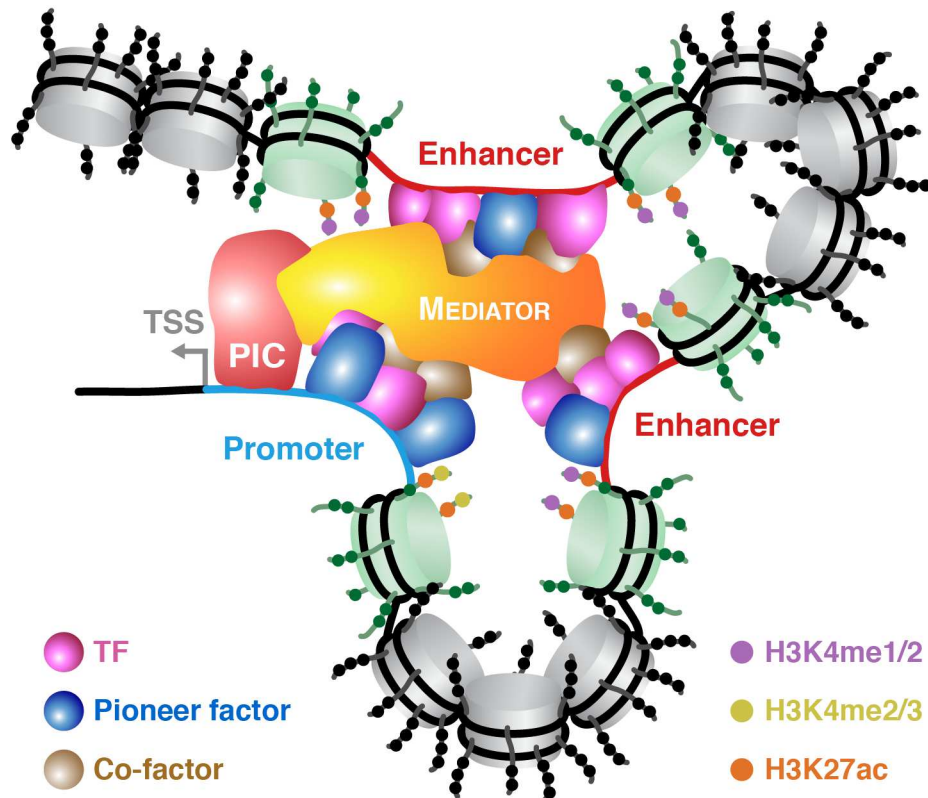


Figure 13. Schematic representation of RNA polymerase II recruitment. DNA looping brings enhancers located distally near their target gene. Although this mechanism remains unclear, CTCF and Cohesion are thought to contribute to promoter-enhancer interactions (not shown). The Mediator (yellow) is a large multiprotein complex that mediates the recruitment of the pre-initiation complex (PIC; red), which is composed of the Pol II and GTFs, for further transcription initiation. Adapted from Kim and Shiekhattar 2015.

C. PhD project outline

Fundamental and applied research on skeletal muscle has been mainly focused for decades by examining muscle cells, their progenitors and their properties. As aforementioned in the previous section, muscle development does not depend solely on muscle cells themselves. Several recent lines of evidence have indeed demonstrated that environmental cues coming from the surrounding tissues are crucial during muscle embryogenesis. However, connective tissue differentiation and formation, and their precise role in muscle development remain to date poorly investigated. The PhD thesis presented here intends to characterize the molecular mechanisms underlying the differentiation of connective tissue cells during limb development to provide a framework for future analyses of connective tissue and muscle interconnectivity. In this regard, five zinc-finger TFs were selected: OSR1, OSR2, EGR1, KLF2 and KLF4.

Selection of connective tissue-associated transcription factors

Odd-skipped related 1 and *2* (*OSR1*, *OSR2*) are expressed in the MCT in close vicinity to myogenic cells throughout limb development in chick and mouse embryos (Stricker et al. 2006, 2012). Both TFs induce the differentiation of mesenchymal cells into irregular connective tissue by repressing tendon-, bone- and cartilage-associated genes in culture systems (Stricker et al. 2012). *Osr1*-deficient mouse embryos die at E14.5 and depict defects in individual skeletal muscles and myotendinous junctions (Vallecillo García et al., in revision). During early limb development, *Osr1* and *Osr2* expression are partly restricted by FGF signalling since blocking of FGF receptors in forelimb bud explants extends their expression domain (Lewandowski et al. 2015). Besides their role during muscle patterning, *Osr1* and/or *Osr2* have been related to heart, kidney, joint, tooth and palate development in mice (Gao et al. 2011; James et al. 2006; Lan et al. 2004; Mugford et al. 2008; Xu et al. 2014; Zhang et al. 2009). Recently, *OSR1* has been characterized as a tumour suppressor in gastric cancer via activation of p53 and repression of TCF/LEF (Otani et al. 2014), while *osr2* has been shown to promote *sox9a* and *col2a1* expression as well as fin chondrogenesis in zebrafish (Lam et al. 2013). In *Xenopus*, *osr1* and *osr2* are involved in lung specification by regulating the BMP-WNT signalling crosstalk (Rankin et al. 2012).

Early growth response 1 (*EGRI*) is expressed in tendons close to muscle attachments during chick and mouse limb development (Lejard et al. 2011). Ectopic expression of *EGRI* in chick neural tubes or in murine mesenchymal stem cells is sufficient to induce the expression of tendon-associated genes and collagens, and to promote tendon differentiation (Guerquin et al. 2013; Lejard et al. 2011). Consistently, *Egr1*-deficient mice display tendons with disorganized collagen fibrils, reduced expression of tendon-associated genes and decreased biomechanical properties, as well as impaired healing capacity following injury (Guerquin et al. 2013). In addition to its role in tendon differentiation and repair, *Egr1* regulates cell growth, survival, migration and apoptosis in response to external stimuli including mechanical forces and stress (reviewed in Bhattacharyya et al. 2011; Pagel and Deindl 2011). EGR1 has been associated with thymocyte, brain, kidney, bone and adipose tissue development as well (Bettini et al. 2002; Boyle et al. 2009; Hansson et al. 2012; Knapska and Kaczmarek 2004; Suva et al. 1991).

Krüppel-like factors 2 and *4* (*KLF2*, *KLF4*) are expressed in chondrogenic condensation regions in limb buds of chick and mouse embryos (Antin et al. 2010; Cameron et al. 2009). Although their role in limb development is presently not elucidated, both KLF factors have been associated with various biological processes such as haematopoiesis and the development of lung, thymocytes, heart, gut, skin and vessels (reviewed in Cao et al. 2010b; McConnell and Yang

2010; Park et al. 2016). Most importantly, both KLF factors are able to maintain the pluripotency of embryonic stem cells (Jiang et al. 2008), especially KLF4, which is one of the four key factors sufficient for cell reprogramming (Takahashi and Yamanaka 2006). In addition, Klf2 and Klf4 have been associated with adipogenesis. Klf2 inhibits adipogenesis by preventing the expression of *PPAR γ* and *Klf2*-null mouse embryonic fibroblasts (MEFs) differentiate faster into adipocytes than wild-type cells (Sen Banerjee et al. 2003; Wu et al. 2005). By contrast, Klf4 promotes adipogenesis of murine pre-adipocyte cells by activating *C/EBP β* in conjunction with *Egr2*, while *Klf4* knockdown prevents their differentiation into adipocytes (Birsoy et al. 2008). More recently, both factors have been shown as being involved in myoblast fusion through the ERK5 signalling pathway. Indeed, *Klf2* or *Klf4* knockdown in myoblasts decreases myotube formation, while not affecting the expression of muscle differentiation genes. Reciprocally, overexpression of both Klf factors enhances myoblast fusion by generating larger myotubes with higher number of nuclei (Sunadome et al. 2011).

Deciphering the molecular aspects of connective tissue cell differentiation

Considering the close association existing between connective tissues (MCT, tendon) and muscle, the five zinc-finger TFs, OSR1, OSR2, EGR1, KLF2 and KLF4, constitute therefore good candidates to decipher the influence of connective tissue during limb muscle development. For this study, the chick embryo was chosen as animal model given its ease to manipulate *in vitro* and *in vivo* (e.g. Biau et al. 2007; Bourgeois et al. 2015; DeLise et al. 2000). First, the endogenous expression of the five TFs was investigated during chick limb development. Second, their influence on cell differentiation processes was evaluated upon overexpression of each TF *in vivo* in chick forelimbs, as well as *in vitro* in a high-density culture of mesenchymal cells extracted from chick limb buds. To further evaluate their influence at the molecular level, a whole-transcriptome sequencing (RNA-seq) approach was performed to identify the gene regulatory programs orchestrated by each TF. This approach was supplemented by a genome-wide strategy of chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq) to determine two types of protein-DNA interactions: (i) chromatin architecture, by assessing histone tail PTMs that are characteristic of promoters, enhancers and facultative heterochromatin; and (ii) TF occupancy, so as to distinguish between indirect and direct target genes. The combination of these three levels of gene regulation allowed me to design a novel, unique and unexplored global regulatory network underlying connective tissue cell differentiation. Finally, a few interesting candidate target genes were selected and concisely characterized to validate the gene regulatory model.

MATERIALS AND METHODS

A. *Experimental procedures*

Fertilized chicken eggs

Fertilized eggs used for *in situ* hybridization were provided by the Institut de Sélection Animale (JA 57 strain, Lyon, France). Chicken micromass cultures were prepared from fertilized eggs obtained from VALO BioMedia (Lohmann Selected Leghorn strain, Osterholz-Scharmbeck, Germany). White Leghorn fertilized eggs (HAAS, Strasbourg, France) were used for the preparation of chicken embryo fibroblasts and grafts. Embryos were staged according to the number of days *in ovo* and the Hamburger-Hamilton (HH) stages (Hamburger and Hamilton 1951).

***In situ* hybridization probes**

To target the endogenous expression of the transcription factors (TFs), the following probes were used: *cOSR1* and *cOSR2* (Stricker et al. 2006); *cEGRI* (Lejard et al. 2011); *cKLF2* and *cKLF4* (Antin et al. 2010). These probes were also used to detect the expression of the TFs from the RCAS-BP(A) retroviruses, except for *cOSR1* and *cKLF4* probes, which are located outside of the coding sequences (CDSs) cloned. Primers were designed to generate new probes for these both factors (Supplemental table S1). Expression of tendon and myogenic markers were assessed with the following probes: *cSCX* (Schweitzer et al. 2001); *cTNMD* (chEST332f24, Source BioScience); *cMYOD* (Pourquié et al. 1996); and *cMYOG* (Delfini and Duprez 2004). Antisense digoxigenin-labelled mRNA probes were synthesized by using SP6 or T7 RNA polymerases and the Riboprobe *in vitro* transcription system (Promega), and finally purified by using Illustra ProbeQuant G-50 columns (GE Healthcare). Successful probe synthesis was monitored on a 1% agarose gel.

***In situ* hybridization on paraffin sections**

Fertilized chicken eggs were incubated at 37°C for 4.5 days (HH24-25) or 9.5 days (HH35-36) to investigate the endogenous expression of the TFs. Forelimbs and/or hindlimbs were harvested and fixed overnight at 4°C in 60% ethanol, 30% FA and 10% acetic acid. Limbs were washed twice with 70% ethanol and either stored at 4°C, or incubated for 30 min to 1 hour in 70% ethanol according to their embryonic stage. Limbs were dehydrated by successive baths in 95% ethanol (once) and 100% ethanol (twice), for 30 min to 1 hour, depending on the embryonic

stage. Limbs were transferred in a first bath of 100% toluene until falling down and transferred into a second bath of 100% toluene until becoming transparent. Limbs were then incubated overnight at 65°C in Paraplast (Sigma Aldrich). Limbs were then embedded in Paraplast and stored at room temperature until further processing. Paraffin blocks were cut by using a Microtome (Microm) with a thickness of 8-10 µm. Tissue sections were incubated at 37°C for at least 24 hours before further processing.

Paraffin was first removed from tissue sections by three successive baths of 100% toluene for 5 min. Sections were then rehydrated by successive baths of 5 min with decreasing concentrations of ethanol (from 100% to 30%) and kept in 1X PBS. Tissue sections were permeabilized by treatment with 1 µg/µL of proteinase K in 1X PBS for 7 min at 37°C, washed 5 min in 1X PBS, fixed 20 min with 4% PFA in 1X PBS, washed again 5 min in 1X PBS and kept in SSC 2X for at least 5 min. 1/100th volume of antisense digoxigenin-labelled mRNA probes in hybridization buffer (50% formamide, 10% dextran sulphate, 1 mg/mL yeast RNA, 1X Denhardt's solution, 3 mM Tris-base, 9 mM Tris-HCl, 200 mM NaCl, 5 mM NaH₂PO₄-H₂O, 5 mM Na₂HPO₄, 5 mM EDTA) was added to each tissue section and incubated overnight at 65°C in a wet chamber containing paper soaked with 50% formamide and 2X SSC. Sections were then washed in three baths of 50% formamide, 1X SSC, 0.1% Tween 20, and blocked for at least 2 hours in 1X MAB supplemented with 0.1% Tween 20, 2% blocking reagent (Roche) and 20% goat serum. Sections were then incubated with anti-digoxigenin-AP antibody (Roche, 11093274910) diluted 1:2,000 in blocking solution. Tissue sections were washed five times with 1X MAB and 0.1% Tween 20, twice with 1X NTMT (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM MgCl₂, 0.1% Tween 20). Detection was performed at 37°C in the dark in 1X NTMT containing 22.5 µg/mL NBT (Roche) and 120 µg/mL BCIP (Roche). Sections were finally washed in 1X PBS and mounted with Aquatex (Merck).

If desired, muscle fibres were labelled by immunohistochemistry after the *in situ* hybridization procedure prior to mounting. Tissue sections were washed three times in 1X PBS, incubated for 2 hours with mouse anti-MF20 antibody (Hybridoma Bank) recognizing sarcomeric myosin heavy chains, washed three times in 1X PBS, incubated for 1 hour with goat anti-mouse IgG2b-HRP (Southern Biotech, 1090-05) diluted 1:50 in blocking solution, washed three times in 1X PBS and stained by using DAB and H₂O₂.

Molecular cloning of the transcription factors

The protein CDSs of the TFs OSR1 (NM_001168709.1), OSR2 (NM_001170344.1), EGR1 (NM_204136.2), KLF2 (XM_418264.4) and KLF4 (XM_004949369.1) were amplified by PCR from chick embryo cDNA samples by using primers spanning from the translational start site to

the last triplet before the stop codon and carrying a digestion site for NcoI or BsmBI, and EcoRI, respectively (Supplemental table S1). TF coding sequences were digested with the corresponding restriction enzymes and ligated into an in-house modified version of the pSlax-13 vector (Morgan and Fekete 1996) containing a triple FLAG-tag and a digestion site for SpeI, previously linearized with NcoI and EcoRI. This intermediate vector enabled to fuse the triple FLAG-tag C-terminally to the CDS of each TF. Directional cloning was finally performed by using ClaI and SpeI in order to insert the fused protein-coding sequences into the RCAS-BP(A) vector (Hughes 2004). Sanger sequencing was performed to validate the integrity of the five inserts by using the BigDye v3.1 kit and an ABI 3700 capillary sequencer (Applied Biosystems). All the molecular biology procedures were carried out according to Sambrook and Russel 2001.

Grafts of virus-producing cells into chick embryo limbs

Chicken embryo fibroblast (CEF) cells extracted from E10 (HH36) chick embryos were transfected with each RCAS-BP(A) construct carrying the CDS of the TFs by using the Calcium Phosphate Transfection kit (Invitrogen), and passaged for at least one week at 37°C in DMEM medium (Gibco) containing 4.5 g/L of glucose and supplemented with 10% FBS (Sigma-Aldrich), 1% chicken serum (Gibco), 1% L-glutamine (Gibco) and 1% penicillin/streptomycin (Gibco). On the last day, cells were passaged into a non-coated petri dish to form aggregates due to their inability to adhere on the plastic. Cell aggregates of approximately 50-100 µm in diameter were grafted into forelimb buds of E2.5 (HH17-18) chick embryo as previously described (Duprez et al. 1996b). Grafted and non-grafted contralateral wings were harvested at E9.5 (HH35-36) for either whole-mount *in situ* hybridization against *MYOG*, or *in situ* hybridization on paraffin-embedded tissue sections. Whole embryos were harvested at E11.5 (HH37-38) for skeletal preparation.

Whole-mount *in situ* hybridization

Grafted embryos were harvested at E9.5 (HH35-36). Heads were cut off, viscera were removed and embryos were fixed overnight at 4°C with 4% FA in 1X PBS. Embryos were rinsed twice with PBT (1X PBS, 0.1% Tween 20), incubated for 15 min once in 50% methanol in PBT and twice in 100% methanol. Embryos were stored at -20°C until further processing. Rehydration of embryos was performed by successive bath of 15 min with decreasing concentration of methanol in PBT (from 75% to 25%). Embryos were finally washed twice for 15 min in PBT. Membrane permeabilization was performed for 20 min with 20 µg/µL proteinase K in PBT, followed by 2 rinsing in PBT. Embryos were fixed for 20 min with 4% FA and 0.1% glutaraldehyde in PBT, followed by one rinsing in PBT and one bath of 5 min in PBT. Embryos were rinsed with hybridization mix (50% formamide, 1.3X SSC, 5 mM EDTA, 50 µg/mL yeast RNA, 0.2%

Tween 20, 0.5% CHAPS, 100 µg/mL heparin) and incubated for 1 hour in hybridization mix. Embryos were then incubated for 36 hours at 70°C with 1/100th volume of antisense digoxigenin-labelled mRNA probes in hybridization mix. Embryos were rinsed twice and washed three times for 30 min at 70°C in hybridization mix, followed by an incubation for 20 min at 70°C in 0.5X MAB, 0.05% Tween 20, 50% hybridization mix. Embryos were then rinsed twice and washed twice for 30 min at room temperature in 1X MAB, 0.1% Tween 20. Embryos were incubated for at least one hour in blocking solution containing 1X MAB, 0.1% Tween 20, 2% blocking reagent (Roche) and 20% goat serum. Embryos were incubated overnight with anti-digoxigenin-AP antibody (Roche, 11093274910) diluted 1:2,000 in blocking solution. Embryos were rinsed three times and washed for 2 days with rocking in 1X MAB, 0.1% Tween 20. Embryos were washed twice for 10 min in 1X NTMT (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM MgCl₂, 0.1% Tween 20). Detection was finally performed at 37°C in the dark in 1X NTMT containing 22.5 µg/mL NBT (Roche) and 120 µg/mL BCIP (Roche). Staining was stopped in 1X PBT, embryos were fixed for 30 min with 4% PFA in 1X PBS and stored at 4°C.

Skeletal preparation

E11.5 (HH37-38) embryos were harvested for skeletal preparation as previously described (Ojeda et al. 1970; Simons and van Horn 1971). Viscera and eyes were removed, embryos were washed with 1X PBS and fixed for 24 hours in 80% ethanol, 20% acetic acid and 15 mg/mL Alcian blue 8GX (BDH) for cartilage staining. Embryos were dehydrated by several washes in 100% ethanol up to one week. Bones were coloured by using a solution of 0.1 mg/mL Alizarin Red S (Sigma-Aldrich) and 0.5% KOH. Embryos were then incubated in a solution of 1% KOH and 20% glycerol until tissues became transparent and washed in 20% glycerol. Finally, skeletal preparations were incubated in successive baths of increasing concentration of glycerol (from 40% to 100%).

Retroviral particle production

RCAS-BP(A) viruses carrying each of the TF coding sequences were produced with chicken DF1 cells (ATCC). DF1 cells were transfected by using PEI (Polysciences) and kept in culture for one week in DMEM medium (Gibco) containing 4.5 g/L of glucose and supplemented with 10% FBS (Biochrom), 2% chicken serum (Sigma-Aldrich), 1% L-glutamine (Lonza) and 1% penicillin/streptomycin (Lonza). Retrovirus-infected cells were then maintained under starvation conditions for three days in DMEM medium containing 1 g/L of glucose and supplemented with 1% FBS, 0.2% chicken serum and 1% penicillin/streptomycin. Retroviruses were harvested every day by pipetting the supernatant, flash-frozen in liquid N₂ and stored at -80°C. Virus supernatants were thawed at 37°C, filtered by using a 0.45-µm vacuum system (Millipore) and

ultra-centrifuged for three to four hours at 4°C at 22,000 rpm. Supernatants were discarded and pellets were resuspended in the remaining drops by shaking for 1 hour at 4°C on ice. Concentrated viruses were aliquoted, flash-frozen in liquid N₂ and stored at -80°C.

Virus titration was performed by immunocytological staining using an in-house antibody directed against the RCAS-BP(A) envelope glycoprotein (mouse hybridoma cell line 3C2). DF1 cells were infected with viruses serially diluted for 24 hours. Cells were fixed with 4% PFA in 1X PBS for 15 min, washed three times in DPBS (Gibco) and blocked for 30 min in MST solution (DMEM 1 g/L of glucose, 10% FBS, 0.2% Triton X-100). Cells were then incubated overnight at 4°C with the antibody directed against the RCAS envelope protein diluted 1:5 in MST blocking solution. Cells were washed three times in DPBS and incubated for 30 min with a biotinylated anti-mouse IgG antibody (Vector Laboratories) diluted 1:500 in MST blocking solution containing 0.005% of horse serum. Cells were washed three times in PBS supplemented with 0.1% of Tween-20, cell membranes were permeabilized for 30 min in 0.3% H₂O₂/methanol and washed three times in PBS and 0.2% Tween-20. Cells were finally stained by using the Vectastain Elite ABC and the DAB Peroxidase Substrate kits (Vector Laboratories). Virus titre was calculated by counting the number of brown cells stained by the peroxidase reaction for a given viral dilution.

Chicken micromass cultures

Chicken micromass (chMM) cultures were prepared as previously described (DeLise et al. 2000; Ibrahim et al. 2013; Solursh et al. 1978). Fertilized chicken eggs were incubated at 37.5°C and 60% of humidity for 4.5 days (HH24-25). Fore- and hindlimb buds were dissected and collected in DPBS (Gibco). Limb buds were washed three to five times in prewarmed Hanks' Balanced Salt Solution (HBSS; Lonza) and digested for 15 min at 37°C in prewarmed Dispase solution (Gibco) at 3 mg/mL to dissociate the ectoderm from the mesenchyme. Limb buds were washed eight to ten times with prewarmed HBSS to discard the ectoderm. Limb mesenchyme was then incubated for 30 min at 37°C in 1 mL of a prewarmed digestion solution composed of 0.1% Collagenase type Ia (Sigma-Aldrich), 0.1% Trypsin (Gibco) and 5% FBS (Biochrom) in DPBS. Mesenchymal tissues were then pipetted up and down to obtain a single-cell suspension. Cells were transferred in 9 mL of prewarmed chMM medium consisting of DMEM/Ham's F-12 (1:1) medium (Biochrom) supplemented with 10% FBS, 0.2% chicken serum, 1% L-glutamine and 1% penicillin/streptomycin. Cell suspension was passed through a 40-µm strainer (Fischerbrand) to filter single cells and 10 additional mL of prewarmed chMM medium were used to wash the cell strainer. Cells were spun down 5 min at 1,000 rpm, resuspended in 5 mL of prewarmed chMM medium and counted by using a Neubauer chamber. Volume was adjusted to reach 2×10^7

cells/mL with prewarmed chMM medium and RCAS-BP(A) retroviral particles were added at a ratio 1:1 to obtain a final concentration of 1.6×10^7 cells/mL. 10 μ L of this virus/cell suspension were seeded per well on a 24-well cell culture plate and incubated for 2 hours at 37°C until cells adhere. Finally, 1 mL of chMM medium was carefully added to each drop. Cultures were kept at 37°C for five days and medium was renewed every two days.

Histological staining

Histochemistry was used to assess viral dissemination and cell differentiation processes within the chMM cultures after 5 days of infection with each of the produced retroviruses. Viral flagged-TF expression was monitored by using a mouse antibody directed against the triple-FLAG tag (Sigma-Aldrich, F1804). Muscle differentiation was assessed by using a mouse anti-MF20 antibody (Hybridoma Bank). Immunohistological staining was performed as described for the retrovirus titration procedure by using the Vectastain Elite ABC and the DAB Peroxidase Substrate kits (Vector Laboratories). Anti-FLAG and anti-MF20 antibodies were used at a dilution of 1:500 and 1:100, respectively. Muscle differentiation was measured by determining the percentage of culture area covered by myotubes in four distinct regions of 500×500 pixel² for each chMM culture using ImageJ (Schneider et al. 2012). To visualize the overall morphology of RCAS-infected chMM cultures, an Eosin staining was performed. Cultures were fixed with 4% PFA in 1X PBS at 4°C overnight, washed three times in 1X PBS, stained for 2 min with 2.5 g/L of Eosin (Sigma-Aldrich) in 80% ethanol and 0.5% acetic acid, and finally washed five times in 1X PBS. Cartilage differentiation in chMM cultures was assessed by Alcian blue staining. Cultures were fixed for 30 min with Kahle's fixation solution (1% formalin, 30% ethanol and 4% acetic acid), washed three times with 1X PBS and stained overnight at 4°C in 1% Alcian blue in 0.1 M HCl. Cultures were then washed five times in 1X PBS and dried at room temperature. Chondrogenic matrix areas were measured by using ImageJ (Schneider et al. 2012).

Western blot

The integrity of the retroviral TF expression was validated by Western blot. Protein extracts were obtained from infected chMM cultures (6 wells) by harvesting the cells in RIPA buffer containing 50 mM HEPES-KOH pKa 7.55, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-deoxycholate supplemented with protease inhibitors (Roche), and passing them at least five times through a 0.4-mm needle. Protein concentration was determined by using the Micro BCA protein assay kit for microplates (Thermo Fischer). 10 μ g of each protein sample were mixed with 4X boiling Laemmli loading buffer (62.5 mM Tris-HCl pH 6.8, 1% SDS, 10% glycerol, 0.005% bromophenol blue, 355 mM 2-mercaptoethanol) and incubated for 10 min at 95°C. Protein samples were spun down for 3 min at 3,000 x g and separated by SDS-PAGE using a

stacking gel with 5% acrylamide/bis-acrylamide (125 mM Tris-HCl pH 8.8, 0.1% SDS, 0.1% APS, 0.1% TEMED) and a separation gel with 12% acrylamide/bis-acrylamide (0.4 M Tris HCl pH 8.8, 1% SDS, 1% APS, 0.1% TEMED) in 1X migration buffer (25 mM Tris-HCl pH 7.5, 250 mM glycine, 0.1% SDS). Transfer was performed overnight at 4°C at 20 V in 50 mM Tris-HCl, 50 mM glycine, 20% methanol and 0.04% SDS, on a Immobilon PVDF membrane (Millipore, 0.45- μ m pore size) previously activated for 3 min in 100% methanol. Membranes were washed twice in 1X TBS (10 mM Tris-HCl pH 7.5, 150 mM NaCl) and blocked overnight at 4°C in 5% milk in 1X TBS containing 0.1% Tween 20. Membranes were washed four times in 1X TBS and 0.1% Tween 20, and incubated 2 hours at room temperature with either a mouse anti-FLAG antibody (Sigma-Aldrich, F1804) diluted 1:1,000, or a rabbit anti-Histone H3 antibody (Abcam, ab1791) diluted 1:10,000 in 5% milk in 1X TBS and 0.1% Tween 20. After four washes in 1X TBS and 0.1% Tween 20, membranes were incubated for 1 hour at room temperature with either a goat anti-mouse IgG peroxidase conjugate antibody (Calbiochem, DC08L), or a goat anti-rabbit IgG peroxidase conjugate antibody (Calbiochem, DC03L), diluted 1:1,000 in 5% milk in 1X TBS and 0.1% Tween 20. Membranes were finally washed four times in 1X TBS and 0.1% Tween 20. Proteins were detected by using the Western Lightning Plus-ECL (PerkinElmer) reagents.

Gene expression profiling

Two biological replicates of chMM cultures were prepared from two independent pools of 4.5-day-old limb buds and infected for 5 days with RCAS-BP(A) retroviruses carrying each of the TFs or no recombinant protein as control. For each replicate, RNA extracts were obtained by harvesting 6 culture wells with RLT buffer (Qiagen). Total RNAs were purified by using the RNeasy mini kit (Qiagen) in combination to a DNase I (Qiagen) treatment to prevent genomic DNA contamination. Quality of RNA extracts was monitored on a 1% agarose gel. RNA samples were given to the sequencing facility of the Max Planck Institute for Molecular Genetics for high-throughput sequencing. RNA libraries were prepared by using the TruSeq Stranded mRNA Library Preparation kit (Illumina), which enables to preserve the RNA strand orientation. Strand-specific 50-bp paired-end reads were generated by using a HiSeq 2500 sequencer (Illumina) with a mean insert size of ~150 bp (Supplemental table S2).

Cell cross-link

RCAS-BP(A)-infected chMM cultures from the two independent biological replicates used for RNA-sequencing were also harvested for cross-linking. After 5 days of growth, chMM cultures were digested for 60 min at 37°C in a solution of 0.1% Collagenase type Ia in chMM medium. Cultures were scratched by using a scraper and pipetted up and down to obtain a cell suspension.

Cell suspensions were spun down for 5 min at 1,000 rpm, washed with DPBS and spun down again for 5 min at 1,000 rpm. Pellets were resuspended in 10 mL of cold chMM medium and kept on ice. Cells were cross-linked in 1% FA for 10 min on ice. Fixation was quenched with 125 mM of glycine. Cross-linked cells were spun down for 5 min at 1,000 rpm and washed twice with cold DPBS. After a final centrifugation step of 5 min at 1,000 rpm, pellets were flash-frozen in liquid N₂ and stored at -80°C.

Nuclear extraction

Nuclear extracts were prepared according to Ibrahim et al. 2013; Lee et al. 2006. Cross-linked cells were thawed and incubated for 10 min at 4°C with gentle rocking in 10 mL of Lysis Buffer 1, containing 50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and protease inhibitors (Roche). Cells were spun down for 5 min at 4°C at 2,700 rpm and incubated for 10 min at room temperature with gentle rocking in 10 mL of Lysis Buffer 2, containing 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors (Roche). Cell nuclei were spun down for 5 min at room temperature at 2,700 rpm followed by resuspension in 1.5 mL of cold Lysis buffer 3, containing 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine and protease inhibitors (Roche).

Chromatin sonication

Nuclear extracts were sonicated by using a Bioruptor NexGen (Diagenode) for 45 cycles of 30-sec pulse and 30-sec pause with high intensity at 4°C. 150 µL of Triton X-100 were added to each sonicated fraction and spun down for 10 min at 4°C at 16,000 x g to remove cell debris. Supernatants were carefully retrieved and transferred into a fresh tube. 100 µL were taken for chromatin concentration assay, while the rest of each sonicated sample was flash-frozen in liquid N₂ and stored at -80°C. 5M NaCl was added to each chromatin aliquot to reach a final concentration of 0.5M NaCl. Chromatin aliquots were incubated for 15 min at 99°C with shaking (1,300 rpm) to reverse the cross-link. 4 µL of RNase A were added, samples were briefly vortexed and incubated for 30 min at 37°C to digest RNAs. 4 µL of proteinase K were added, samples were briefly vortexed and incubated for 30 min at 55°C to digest proteins. 4 µL of 5 mg/mL glycogen were finally added to each chromatin sample followed by 2.5 volumes of cold 100% ethanol and incubated for 2 hours at -20°C. Chromatin samples were then spun down for 30 min at 4°C at 15,000 x g, pellets were washed with cold 70% ethanol and spun down again for 10 min at 4°C at 15,000 x g. Supernatants were removed, pellets were dried at room temperature and resuspended in 20 µL of ddH₂O. Chromatin samples were incubated 30 min at

37°C for complete resuspension. Concentrations were measured and used to estimate the concentration of chromatin in the nuclear extracts according to the following formula:

$$\frac{C_{\text{aliquot}} \times V_{\text{resuspended}} \times \frac{V_{\text{nuclear extract}}}{V_{\text{aliquot}}}}{V_{\text{nuclear extract}}}$$

Efficiency of the sonication was monitored by analysing the remaining chromatin aliquots on a 1% agarose gel. Chromatin sonication was considered as sufficient and successful if DNA was fragmented between 200 and 500 bp.

Chromatin immunoprecipitation

Nuclear extracts from both biological replicates were used for chromatin immunoprecipitation (ChIP). Histone modifications were investigated in the chMM cultures infected with empty RCAS-BP(A) retroviruses carrying no recombinant protein, whereas the chMM cultures infected with RCAS-BP(A) particles carrying each of the TF CDS were used to investigate the TF-binding sites. The following antibodies and volumes were used to establish the chromatin landscape within the chMM cultures: 4 µL (4 µg) of mouse anti-H3K4me1 (Abcam, ab8895); 8 µL of mouse anti-H3K4me2 (Abcam, ab32356); 4 µL of mouse anti-H3K4me3 (Millipore, 07-473); 4 µL (4 µg) of mouse anti-H3K27ac (Abcam, ab4729); and 4 µL (4 µg) of mouse anti-H3K27me3 (Millipore, 07-449). 10 µL (10 µg) of mouse anti-FLAG (Sigma-Aldrich, F1804) were used to target the retroviral TFs.

10 µg (~8 culture wells) and 30 µg (~24 culture wells) of chromatin extracts were used for ChIP against the histone marks and against the TFs, respectively. Aforementioned volumes of antibody were mixed to the chromatin fractions and incubated overnight at 4°C with gentle rocking. 50 µL of each chromatin extract were kept as input control and stored at -20°C until cross-link reversal. 40 µL of magnetic beads (Dynabeads protein G; Thermo Fischer) were used per ChIP. Prior to add them to the chromatin samples, magnetic beads were first blocked by four successive washes with 0.25% BSA in DPBS and finally resuspended in 100 µL of this blocking solution. Coated beads were then mixed to the antibody-chromatin complexes and incubated for at least 3 hours at 4°C with gentle rocking. Beads were then captured with a magnet, supernatant was discarded, and washed six times with cold RIPA buffer, containing 50 mM HEPES-KOH pKa 7.55, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-deoxycholate and protease inhibitors (Roche). Complexes were washed once with 500 µL of cold TE buffer, containing 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 50 mM NaCl and protease inhibitors (Roche). Supernatant was removed and complexes were spun down for 3 min at 4°C at 1,000 x g. Remaining supernatant was removed, complexes were resuspended in 210 µL of Elution buffer (50 mM

Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and incubated for 30 min at 65°C with shaking (900 rpm). Eluates were spun down for 1 min at 16,000 x g and 200 µL of supernatant was carefully retrieved.

Input controls and ChIP eluates were then handled in parallel. 1/10th volume of 5M NaCl was added to each input and ChIP samples followed by an incubation overnight at 65°C to reverse the cross-link. Both sample types were then processed as for the chromatin concentration assay with successive incubations with RNase A and proteinase K to remove RNAs and proteins, respectively, and followed by ethanol precipitation and resuspension in ddH₂O. Precipitated ChIP and input samples were finally given to the sequencing facility of the Berlin-Brandenburg Center for Regenerative Therapies for high-throughput sequencing. Libraries were prepared by using the NEBNext Ultra DNA Library Preparation kit for Illumina (New England Biolabs). 50-bp single-end reads were generated by using a HiSeq 1500 sequencer (Illumina; Supplemental tables S3, S4).

Expression pattern of interesting targeted genes

Primers listed in Supplemental table S1 were used to generate probes detecting the following targets: *ADGRG2* (XM_015272749.1); *ANXA1* (NM_206906.1; GEISHA ID, ANXA1.UApcr); *CBFA2T2* (NM_001011689.1); *NTN1* (L34549.1; Murakami et al. 2010); *WNT4* (NM_204783.1; GEISHA ID, WNT4.UApcr); *WNT11* (NM_204784.1; GEISHA ID, WNT11.UApcr). Products amplified by PCR from chick embryo cDNA samples were cloned into the pCRII-TOPO vector (Invitrogen). Depending on the insert orientation, plasmids were linearized by either SpeI (Thermo Fischer), or NotI (Thermo Fischer). Probe synthesis was performed as aforementioned. Endogenous expression of the listed putative targeted genes was investigated by *in situ* hybridization on whole-mount at E5.5 (HH27-28) and paraffin-embedded tissue sections at E8 (HH34).

B. Computational analysis

Transcript-discovery approach

RNA-seq data obtained from both control biological replicates were used to more accurately define the gene expression profile within the chMM culture. Firstly, strand-specific read pairs were mapped against the chicken genome galGal4 (Hillier et al. 2004) by using TopHat2 v0.14 (Kim et al. 2013) with the following parameters: -r 150; -N 3; --read-edit-dist 3; --library-type fr-firststrand; -i 50; -G. UCSC (galGal4) and Ensembl (release 75) annotations were downloaded from Illumina iGenomes and compared by using Cuffcompare from the Cufflinks suite v2.1.1

(Trapnell et al. 2010). Identical genes were retrieved only once and merged with the unique genes from each annotation. In case of discordant genes, the gene annotation with the best coverage was selected. The resulting gene annotation model composed of 17,318 genes was used as input for TopHat2 mapping. Transcript discovery was performed for each replicate by using Cufflinks (-b; -u; -library-type, fr-firststrand; -g) and the combined gene annotation model as guide. Resulting annotations were merged into a single model by using the Cufflinks tool Cuffmerge.

In addition, a second transcript-discovery approach was led independently of the genome sequence. Low-quality RNA-sequencing reads from control chMM cultures were first filtered out by using the FASTX-Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit). Reads with a median quality value lower than 28 were discarded. Filtered read pairs were then trimmed by using Trimmomatic v0.32 (Bolger et al. 2014) with the following parameters: ILLUMINACLIP TruSeq3 paired-end for HiSeq, seedMismatches 2, palindromeClipThreshold 30, simpleClipThreshold 10; LEADING 5; TRAILING 5; MINLEN 36. Complete read pairs were then assembled by using Trinity r20140717 (Grabherr et al. 2011) with default parameters except for the library type set at RF. Resulting contigs were compared to the gene sequences obtained by the first approach by using BLAST v2.2.31+ (Camacho et al. 2009) (-strand plus; -dust no; -soft_masking no). Contigs were assigned to a given gene if they matched at least 40 bp with a percentage of identities higher than 90%. Assigned contigs that were not fully covered by a given gene were further processed to extract continuous uncovered regions of at least 400 bp. Remaining contigs were mapped against the galGal4 genome by using BLAST. Contigs were assigned to a given gene if they were located between two gene features, potentially corresponding to an exon missed by Cufflinks, or in the vicinity of a first or last exon, potentially corresponding to a missing 5'- or 3'-untranslated region (UTR), respectively. Remaining unmapped contigs were retrieved as they could correspond to non-defined genomic regions. Unmapped, unassigned and non-covered contigs or regions were compared to each other to remove redundant regions. Unique contig sequences were gathered together as an additional chromosome and separated to each other by 250 bp.

Gene sequences retrieved from both transcript-discovery approaches were then compared to existing databases for gene name assignment. First, genes were compared to the NCBI RefSeq transcript database by using BLASTN (-strand plus; -dust no; -soft_masking no). Comparison was limited to Aves (birds) sequences (taxid 8782), including the chicken annotation galGal5 lately released (http://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/103). Genes with a percentage of identities higher than 75% and 90% for bird and chicken genes,

respectively, and covering or covered by at least 50% of their length were assigned to the corresponding hits. Non-annotated gene sequences were then compared against the NCBI human (taxid 9606) and mouse (taxid 10090) non-redundant protein database by using BLASTX (-strand, plus; -seg, no). Genes with a percentage of homology of at least 30% and covered by at least 50% of their length were filtered. Matching protein accession numbers were converted into gene accession numbers by using the Hyperlink Management System (Imanishi and Nakaoka 2009). Open reading frame (ORF) prediction was finally performed on remaining genes by using TransDecoder v2.1.0 (Haas et al. 2013) with strand specificity (-S). ORFs of at least 100 amino acids were annotated by using Trinotate v3.0.1 (<https://trinotate.github.io>). Functional annotation was based on the following protein predictions: (i) BLASTX and BLASTP homology search against the SwissProt database (Bairoch et al. 2004); (ii) protein domain prediction against the Pfam database (Punta et al. 2012) by using HMMER v3.1b2 (Finn et al. 2011); (iii) peptide signal prediction by using SignalP v4.1 (Petersen et al. 2011); and (iv) transmembrane domain prediction by using tmHMM v2.0c (Krogh et al. 2001). Resulting functional annotation was divided into three categories: (i) putative proteins, for which at least one protein domain could be identified; (ii) uncharacterized proteins, corresponding to ORFs for which no protein domain could be identified; and (iii) non-coding RNAs, corresponding to genes with an ORF shorter than 100 amino acids.

Altogether, this dual transcript-discovery approach enabled to define an annotation model of 21,347 unique genes that was used afterwards as reference for the analysis of the chMM culture expression profiles.

Differential expression analysis

RNA-seq strand-specific read pairs generated for each chMM culture condition and replicate were first mapped against the chicken genome galGal4 by using TopHat2 (-r 150; -N 3; --read-edit-dist 3; --library-type fr-firststrand; -i 50; -G) and the annotation model from the transcript-discovery approach as guide. Alignment maps for each replicate of each condition were split by strand by using SAMtools v1.2 (Li et al. 2009b) according to the FLAG field (strand plus: -f 128 -F 16, -f 80; strand minus: -f 144, -f 64 -F 16). Fragments (both reads of a pair) mapped on gene features were counted by using featureCounts v1.4.6-p3 (Liao et al. 2014) with the following parameters: -p; -s 2; --ignoreDup; -B; -R. Chimeric fragments aligned on different chromosomes were taken into consideration to overcome the gene fragmentation due to the location of gene parts on multiple chromosome contigs. Reads unmapped on the galGal4 genome and mapped but unassigned due to the absence of gene feature were extracted and parsed to remove singletons. Resulting complete read pairs were then mapped against the corresponding RCAS-BP(A)

construct sequence by using TopHat2 (same parameters, except that no gene annotation model was provided). Remaining unmapped read pairs were aligned by using TopHat2 (same parameters) against the created contig containing all the *de novo* gene sequences identified by Trinity. Alignment maps were split by strand by using SAMtools and fragments were counted only for the strand plus by using featureCounts with the same options, except that chimeric read pairs were ignored (-C).

Fragment counts for each biological replicate of each chMM culture condition were summed and normalized by using DESeq2 v1.8.1 (Love et al. 2014). The DESeq2 normalization method is based on the assumption that most genes are not differentially expressed (DE) when comparing two conditions. First, DESeq2 calculates the geometric mean for each gene across all conditions and replicates. Second, it divides the fragment count of each gene for each sample by this mean. A scaling factor is finally defined for each sample as the median of these ratios, which is then applied to all genes of the corresponding sample. This normalization method thus computed corrects for library size and RNA composition bias between samples (Supplemental figures S1A,B) and has been demonstrated to be more sensitive and powerful as compared to using raw fragment counts (Dillies et al. 2013). To evaluate the discrepancy among biological replicates and conditions, a regularized-logarithm transformation was applied to normalized fragment counts (Love et al. 2014). This additional normalization shrinks together genes with low fragment counts among samples while genes with high fragment counts remain unaffected (Supplemental figure S1C), thus stabilizing both fragment count variance and biological noise. Hierarchical clustering was then performed among chMM culture expression profiles by computing the Euclidean distances between the biological replicates across all conditions. Differential expression analysis was finally carried out by comparing the normalized fragment counts of chMM cultures overexpressing each of the TFs against the control chMM cultures by using DESeq2 and a false-discovery rate (FDR, α) of 0.01. Genes with an absolute fold change of at least 2 and a Benjamini-Hochberg adjusted p-value (Benjamini and Hochberg 1995) below 0.01 were considered as being differentially expressed, resulting in a total of 10,712 DE genes across all conditions.

Estimation of transcript abundances

Transcripts per million (TPM) values (Li et al. 2009a; Wagner et al. 2012) were preferred to Reads/Fragments per kilobase of transcript per million (RPKM/FPKM) values (Mortazavi et al. 2008) to estimate the transcript abundances within each biological replicate. Both calculation methods are based on the normalization of fragment counts by the gene length but they differ in the normalization factor used. TPM values are estimated by using the total number of

reads/fragments overlapping gene features, whereas RPKM/FPKM values are normalized by the total number of sequenced reads/fragments. Nevertheless, the latter estimation method has been shown to be inconsistent among samples (Dillies et al. 2013). TPM values were then calculated by using the DESeq2 normalized fragment counts to limit the library size effect across samples and the following formula (Wagner et al. 2012):

$$\frac{Gene_{fragment\ count} \times Read_{length}}{Gene_{length} \times \sum \frac{Gene_{fragment\ count} \times Read_{length}}{Gene_{length}}} \times 10^6$$

Genes with a TPM value of at least 1 were considered as being expressed.

Prediction of cell type abundances

To assess the propensity of the TFs to influence particular cell fates within the chMM cultures, transcript abundances of genes related to differentiation and developmental processes were investigated. Lists of genes belonging to given Gene ontology (GO) terms were extracted by using the QuickGO tool (Binns et al. 2009) and restraining the research to the Tetrapoda taxon (tetrapods, taxid 32523). The following cell types and tissues and their associated GO terms were investigated: (i) bone (GO:0060348) and bone cell (GO:0098751) development; (ii) cartilage development (GO:0051216) and chondrocyte differentiation (GO:0002062); (iii) muscle cell development (GO:0055001) and differentiation (GO:0042692), myoblast development (GO:0048627) and differentiation (GO:0045445); (iv) tendon development (GO:0035989) and tendon cell differentiation (GO:0035990); (v) adipose tissue development (GO:0060612) and fat cell differentiation (GO:0045444); and (vi) embryonic skeletal system development (GO:0048706) and morphogenesis (GO:0048704). The list of genes associated with tendon development was supplemented with the top 100 genes enriched in mouse *Scx*-expressing cells (Havis et al. 2014). Unique genes were extracted and filtered to keep only those with a TPM value of at least 1 across all chMM culture samples. Paired Wilcoxon rank-sum test was used to assess separately both alternative hypotheses, greater or less, when comparing for a given GO term the transcript abundances of TF-overexpressing chMM cultures against control cultures.

Estimation of overexpression levels

Retroviral expression of each TF within chMM cultures was estimated by mapping the raw RNA-seq paired-end reads against the corresponding RCAS-BP(A) construct sequence with TopHat2 (same parameters as previously). Mapped read pairs from the strand plus were extracted by using SAMtools and counted by using featureCounts (-p; -s 2; --ignoreDup; -B; -C; -M; --fraction; --minReadOverlap 50; -R). Three regions common to all constructs and unique to each viral splicing variant were selected: *gag* and *env* CDS (first and second splicing variant,

respectively) and a region spanning from the triple-FLAG tag to the 3'LTR (third splicing variant). Fragment counts were normalized by the corresponding DESeq2 size factor and the region length. Proportion of each splicing variant was determined by subtracting the normalized fragment count between the three regions. Overexpression level was estimated based on the assumption that the endogenous expression of a given TF is constant when an infected cell expresses the corresponding TF virally. Therefore the following formula was applied:

$$\frac{(TPM_{virus} - TPM_{control}) \times Proportion_{third\ splicing\ variant}}{TPM_{control}}$$

K-means gene clustering

DE genes identified for each TF-overexpressing chMM culture were gathered together. The resulting 4,298 non-redundant DE genes were clustered by using *K*-means (Hartigan and Wong 1979) based on the DESeq2 normalized fragment counts across samples. *K*-means is a centroid clustering method used to partition a number of observations into a given number of clusters such that observations in a same cluster are as similar as possible to each other as compared to observations in another cluster. In term of gene expression, a given cluster would contain genes with a similar expression pattern such as up- or down-regulation within a certain number of conditions that would not be shared by the remaining conditions and that would be only observed in this particular cluster. *K*-means clustering was performed by using GENE-E (<http://www.broadinstitute.org/cancer/software/GENE-E>) with a row distance metric set at 1 minus Pearson correlation and 2,000 iterations. The number of *K* clusters was defined at 8 because lower values did not separate distinct gene clusters and higher values subdivided meaningful gene clusters.

Gene ontology analysis

Gene ontology (Ashburner et al. 2000) analyses were performed for given gene lists by using the PANTHER statistical overrepresentation test r20160321 (Mi et al. 2010) and the Bonferroni correction for multiple testing (Dunn 1961). The following annotations were interrogated: PANTHER version 10.0 released on 2015-05-15 for GO-slim biological process, GO-slim cellular component, pathways and protein class; GO ontology database released on 2016-04-23 for GO biological process complete.

ChIP-sequencing

Two types of protein-DNA associations were investigated genome-widely within the chMM cultures. Firstly, both biological replicates from the chMM cultures infected with RCAS-BP(A) retroviral particles carrying no recombinant protein were used to characterize the chromatin

landscape. The following histone modification profiles were explored: H3K4me1, H3K4me2, H3K4me3, H3K27ac and H3K27me3. Secondly, the binding sites of the five TFs were investigated within both biological replicates of chMM cultures infected with RCAS-BP(A) retroviruses carrying the corresponding TF CDS. For both ChIP-seq series, sonicated DNA from the nuclear fractions used for each sample was sequenced as input control.

50-bp single-end reads generated for each ChIP and input fractions were first filtered on their quality by using the FASTX-Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit). Reads with a median quality value of minimum 28 were retrieved and mapped against the chicken genome galGal4 by using BWA v0.5.9 (Li and Durbin 2009) with default parameters. Uniquely mapped reads were then extracted by parsing the alignments containing the tag “XT:A:U”. Duplicated reads were finally removed by using the tool rmdup from SAMtools v1.2 (Li et al. 2009b). Resulting alignment maps were then separately processed depending on the investigated profile, histone marks or transcription factor-binding sites (TFBS).

Histone modification peak calling

Peak calling for the histone mark ChIP-seq was performed as suggested by the ENCODE consortium and the Roadmap Epigenomics project (Kellis et al. 2014; Roadmap Epigenomics Consortium et al. 2015). For each histone modification, peaks were called independently for each biological replicate and for the pooled biological replicates, each time against the merged input control of both replicates, by using MACS2 v2.1.0.20140616 (Zhang et al. 2008) and the following parameters: --bw 400 (according to the sonicated DNA migration gel); -g 1.0e9; --to-large. Except for the H3K27me3 mark, peak calling was performed twice for each replicate and pooled replicate: (i) narrow peaks passing a p-value (-p) of 0.01; and (ii) broad peaks passing an additional broad-peak p-value (-p 0.01; --broad; --broad-cutoff) of 0.1. Only broad peaks were called for the H3K27me3 ChIP-seq due to its diffused signal. Broad peaks detected for each replicate and pooled replicate that contain at least one narrow peak were extracted by using BEDtools intersect v2.24.0 (Quinlan and Hall 2010). Final sets of peaks for each histone modification were obtained by filtering broad peaks called for the pooled replicates that are shared between both biological replicates independently.

Identification of regulatory domains

Regulatory domains were defined according to the combination of the different histone modification profiles obtained by ChIP-seq, independently of the gene annotation model. Domains were divided into three categories: (1) promoters; (2) enhancers; and (3) repression islands. (1) Promoters were defined according to the presence of H3K4me3 signal. (2) Enhancers

corresponded to regions enriched for H3K4me1 and devoid of H3K4me3 signal. (3) Repression islands were distinguished by the unique presence of H3K27me3 signal. Regions enriched for H3K4me2 but with no detectable H3K4me1 signal were classified as promoters, whereas regions containing both H3K4me1/2 marks were defined as enhancers. Promoter and enhancer domains were further subcategorised into four distinct states according to the active marks H3K4me3 and H3K27ac, and the repressive mark H3K27me3: (i) inactive, no active and repressive signal detected (H3K4me3⁻, H3K27ac⁻, H3K27me3⁻); (ii) poised, no active mark but repressive signal detected (H3K4me3⁻, H3K27ac⁻, H3K27me3⁺); (iii) active, only active mark detected (H3K4me3⁺ and/or H3K27ac⁺, H3K27me3⁻); and (iv) bivalent, both active and repressive marks detected (H3K4me3⁺ and/or H3K27ac⁺, H3K27me3⁺).

Quality control of transcription factor-binding profiles

The ENCODE consortium defined a series of quality metrics intending to validate ChIP-seq TF-binding profiles prior to biological interpretation (Landt et al. 2012). Strand cross-correlation analysis was performed by using the `get.binding.characteristics` function (srange 0,400; bin 5; cluster 2; accept.all.tag F) from SPP v1.11 (Kharchenko et al. 2008). Quality of the TF ChIP-seq data was evaluated following the ENCODE consortium recommendations:

- i. PCR bottleneck coefficient (PBC; ≥ 0.8 for 10 million uniquely mapped reads), calculated by dividing the number of non-redundant uniquely mapped reads by the number of uniquely mapped reads. Low PBC value is indicative of a low-complexity library;
- ii. Normalized strand correlation (NSC; ≥ 1.05), determined by the cross-correlation analysis and corresponding to the ratio between the ChIP peak and the background signal. Low NSC value is indicative of a weak enrichment;
- iii. Relative strand correlation (RSC; ≥ 0.8), determined by the cross-correlation analysis and corresponding to the ratio between the ChIP peak and the phantom peak (read-length peak). Low RSC value is indicative of a weak signal-to-noise ratio.

Irreproducible discovery rate analysis

In addition, the ENCODE consortium developed a method to evaluate the consistency between biological replicates called the irreproducible discovery rate (IDR) analysis (Landt et al. 2012; Li et al. 2011a). The IDR analysis is based on the assumption that the most significant peaks of two biological replicates are most likely consistent between both replicates and therefore correspond to a real signal. In contrast, less significant peaks of two biological replicates are expected to be detected in only one replicate and therefore correspond to a background (noise) signal. This

method thus relies on the reproducibility between independent biological replicates rather than a fixed threshold during the peak calling procedure.

The IDR analysis was performed on the TF-binding profiles generated for both biological replicates of the chMM cultures overexpressing each of the TFs. The ENCODE consortium recommends to assess the IDR on three different levels: (i) the true replicate consistency, by comparing both biological replicates; (ii) the self-replicate consistency, by randomly subsampling the signal of each biological replicate into two pseudo-replicates, which is a good indicator of the signal-to-noise ratio; and (iii) the pooled-replicate consistency, by merging the signal of both biological replicates and randomly partitioning it into two pseudo-replicates. For each replicate, self pseudo-replicate and pooled pseudo-replicate, peaks were called by using MACS2 v2.1.0.20140616 (Zhang et al. 2008) with low-stringency parameters to obtain a significant list of peaks (--bw 130|135, as determined by the cross-correlation analysis; -g 1.0e9; --to-large; -p 0.025). IDR analysis was performed on the top 125,000 peaks according to their p-value by using the following parameters: peak.half.width -1; min.overlap.ratio 0; is.broadpeak F; ranking.measure p.value. The number of peaks passing an IDR threshold of 0.01 were retrieved for each replicate and pseudo-replicate and compared to each other. The ENCODE consortium recommends that the number of peaks between each comparison is within a factor of 2.

Determination of the transcription factor-binding sites

The IDR analysis is dominated by the replicate with the weakest signal. Consequently, a significant number of peaks with real signal found in the replicate of higher quality are rejected by the IDR analysis because they are not detected in the replicate of lower quality. Therefore, peak calling was performed on the pooled biological replicates for each TF. The final set of TFBS was determined by selecting the number of peaks with an IDR threshold below 0.01 obtained from the pooled-replicate consistency analysis. The global ChIP enrichment for each TF was finally determined by computing the fraction of sequencing reads located in the final set of called peaks (FRiP; ≥ 0.01) using BEDtools coverage v2.24.0 (Quinlan and Hall 2010).

Binding motif analysis

Motif analysis was performed by using DREME v4.11.2 (Bailey 2011) with default parameters on the 150-bp sequences surrounding the summits of the 1,000 most significant TF peaks that overlap with promoters and enhancers. Recognition motifs thus identified were then compared against motif databases by using Tomtom v4.11.2 (Gupta et al. 2007) with default parameters.

Assessment of the similarity between the ChIP-seq profiles

Histone mark and TF coverage profiles were generated by using the tool `bdgcmp` from MACS2 v2.1.0.20140616 (Zhang et al. 2008). ChIP-seq signal was normalized independently for each biological replicate against the pooled input controls of both replicates according to the negative \log_{10} of the Poisson p-value (`-m ppois`). Normalized signal coverage was then determined genome-widely in 500-bp non-overlapping windows. For histone marks, ChIP-seq coverage profiles were compared across all histone modifications and all biological replicates by using the Pearson correlation. For TF-binding profiles, similarity was assessed between all biological replicates and TFs by using a principal components analysis (PCA) (Hotelling 1933) with the R function `prcomp` (`center TRUE`; `scale. TRUE`).

Data visualization

RNA-seq alignment maps of both biological replicates for each chMM culture condition were merged and split by strand by using `SAMtools` v1.2 (Li et al. 2009b). Strand-specific maps of pooled replicates were then converted into `bedgraph` format by using the `genomeCoverageBed` tool v2.24.0 (Quinlan and Hall 2010) with the following parameters: `-bg`; `-ibam`; `-split`; `-g`; `-scale`. Read coverage was normalized between conditions by using the `DESeq2` size factor of each biological replicate and calculated as follows:

$$\left(\frac{0.5}{DESeq2\ factor_{replicate\ A}} \times \frac{0.5}{DESeq2\ factor_{replicate\ B}} \right) / 2$$

ChIP-seq profiles of both biological replicates for each histone modification and TF were merged and normalized against the corresponding pooled input controls by using the tool `bdgcmp` from MACS2 v2.1.0.20140616 (Zhang et al. 2008). Signal normalization was performed by using the negative \log_{10} of the Poisson p-value (`-m ppois`).

RNA- and ChIP-seq coverage profiles were finally converted into `bigwig` format by using the `bedGraphToBigWig` tool (Kent et al. 2010) with default parameters. Resulting files were uploaded to an internal server of the Max Planck Institute for Molecular Genetics and linked to the UCSC genome browser (Kent et al. 2002) as custom tracks. Data can be accessed and visualized via the following link: <https://goo.gl/85kn3x>

A. Characterization of the transcription factors

The five transcription factors (TFs) *OSR1*, *OSR2*, *EGR1*, *KLF2* and *KLF4* were initially selected based on their involvement or supposed involvement in the differentiation of mesenchymal cells towards distinct musculoskeletal lineages during limb development.

The TFs are related to subcompartments of the musculoskeletal system

Endogenous expression of the TFs has been mainly explored by whole-mount *in situ* hybridization (ISH) at early stages of chick limb development. *OSR1* and *OSR2* are strongly and broadly expressed in forelimb and hindlimb mesenchyme between E3.5 and E6 (HH22-28) (Stricker et al. 2006). *EGR1* is expressed in tendon close to muscle attachments in the chick forelimb at E7 (HH30) (Lejard et al. 2011). *KLF2* and *KLF4* are expressed locally in the proximal part of the chick hindlimb at E5-5.5 (HH26-27) (Antin et al. 2010). *KLF2* expression is restricted to the anterior proximal mesoderm, while *KLF4* expression is detected in the dorsal and ventral proximal mesoderm. Both factors are detected at chondrogenic condensation regions where bone will form at E5. To identify more precisely the expression pattern of each TF, ISH were performed on tissue sections. Two developmental stages of the chick limb were analysed: E4.5 (Figure 14; Supplemental figure S2) and E9.5 (Figure 15; Supplemental figure S3).

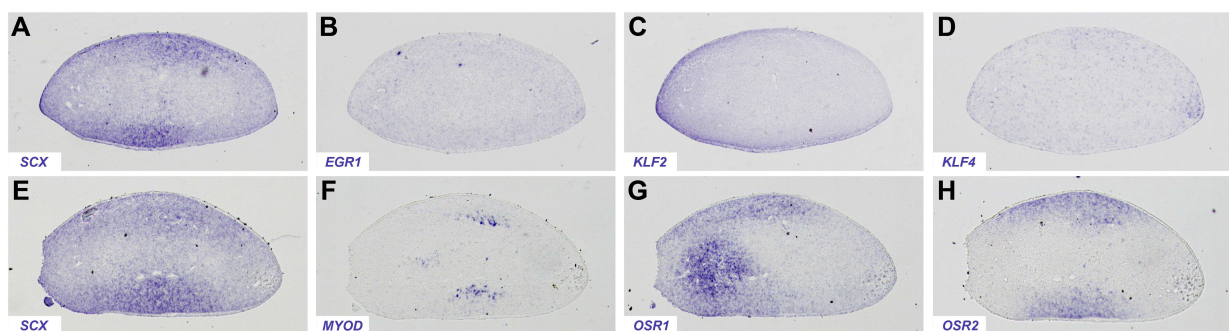


Figure 14. Endogenous expression of the transcription factors in hindlimbs of E4.5 chick embryos. ISH for *SCX* (A,E), *EGR1* (B), *KLF2* (C), *KLF4* (D), *MYOD* (F), *OSR1* (G) and *OSR2* (H) genes in hindlimbs of E4.5 (HH24-25) chick embryos. (A-D) Adjacent and transverse sections were hybridized with *SCX*, *EGR1*, *KLF2* and *KLF4* probes (blue). Magnification 5X. (E-H) Adjacent and transverse sections were hybridized with *SCX*, *MYOD*, *OSR1* and *OSR2* probes (blue). Magnification 5X. Top, dorsal; left, posterior. *SCX* is a tendon-specific marker; *MYOD* is a muscle-specific marker.

In contrast to the tendon marker *SCX*, the expression of *EGR1*, *KLF2* and *KLF4* was not detected in E4.5 limb buds (Figures 14A-D; Supplemental figures S2A-B). Both *OSR1* and *OSR2* transcripts were detected in the dorsal and ventral parts of E4.5 limb buds, partly overlapping with *SCX* expression and surrounding muscle masses visualised with *MYOD* expression (Figures 14E-H; Supplemental figures S2E-H).

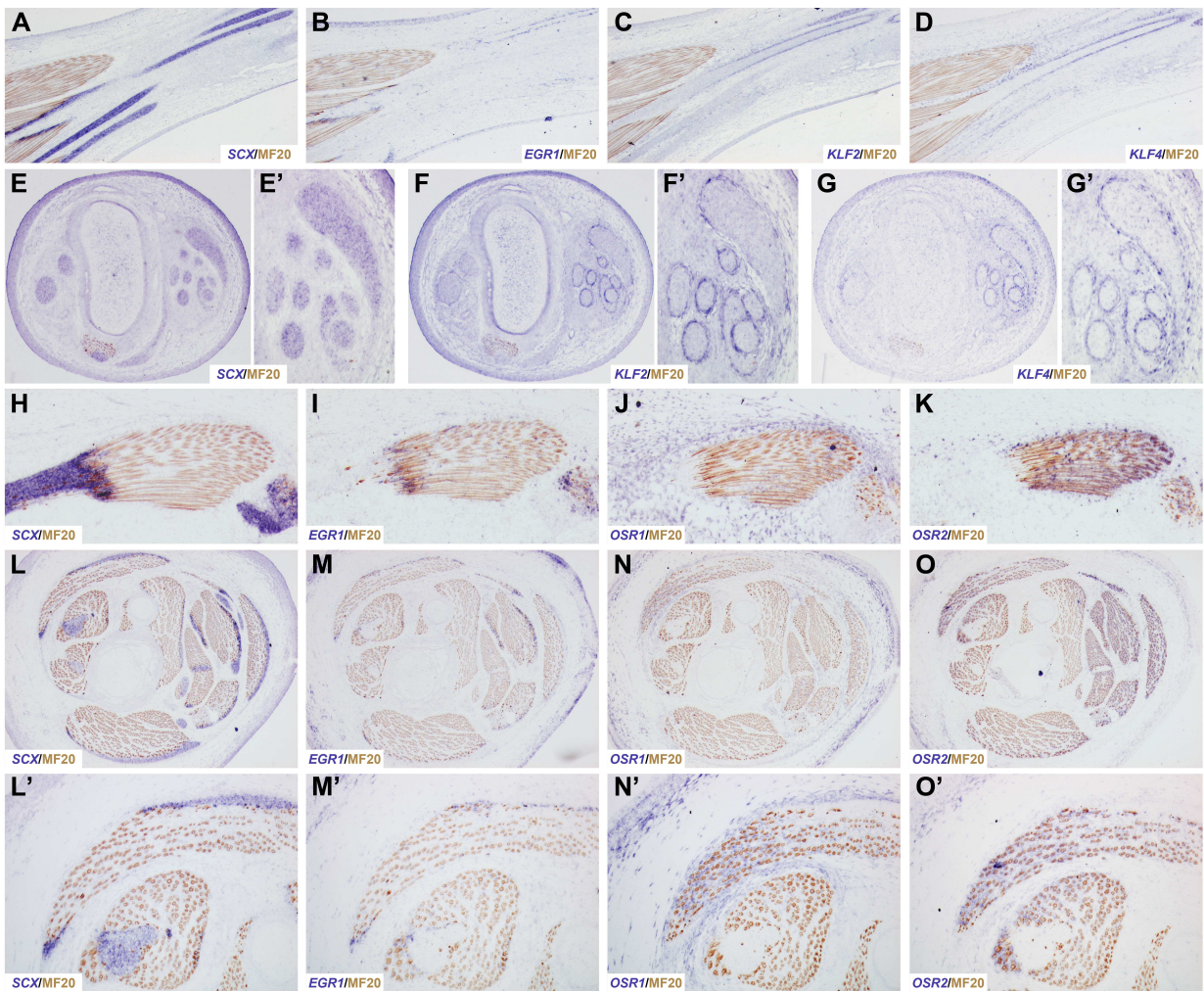


Figure 15. Endogenous expression of the transcription factors in hindlimbs of E9.5 chick embryos. ISH for *SCX* (A,E,E',H,L,L'), *EGR1* (B,I,M,M'), *KLF2* (C,F,F'), *KLF4* (D,G,G'), *OSR1* (J,N,N') and *OSR2* (K,O,O') genes in hindlimbs of E9.5 (HH35-36) chick embryos. (A-D) Adjacent and longitudinal sections were hybridized with *SCX*, *EGR1*, *KLF2* and *KLF4* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the zeugopod level. Top, anterior; left, proximal. (E-G) Adjacent and transverse sections were hybridized with *SCX*, *KLF2* and *KLF4* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the knee level. Top, posterior; left, dorsal (E'-G') Magnification 10X of ventral regions of sections (E-G). (H-K) Adjacent and longitudinal sections were hybridized with *SCX*, *EGR1*, *OSR1* and *OSR2* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X. (L-O) Adjacent and transverse sections were hybridized with *SCX*, *EGR1*, *OSR1* and *OSR2* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the stylopod level; top, posterior; left, dorsal. (L'-O') Magnification 10X of posterior-dorsal regions of sections (L-O). *SCX* is a tendon-specific marker; MF20 targets skeletal muscle myosin.

At E9.5, *EGR1* expression is observed in tendons at myotendinous junctions, in both hindlimbs (Figure 15) and forelimbs (Supplemental figure S3), similar to what was observed at E7 (Lejard et al. 2011). *KLF2* and *KLF4* are expressed in tissues delineating tendons in knee (Figures 15A-G') and wrist (Supplemental figures S3A-H') regions. The expression pattern of *KLF2* and *KLF4* seems to surround *SCX* expression domains. *OSR1* and *OSR2* are more broadly expressed in the limb (Figures 15H-O). Their transcripts are detected in muscle connective tissue (MCT), in between muscle fibres, although *OSR1* is also detected surrounding muscle masses (Figures 15L'-O'; Supplemental figures S3I-L).

OSR1 and OSR2 overexpression induces musculoskeletal defects in forelimbs

Considering their particular expression patterns in distinct compartments of the musculoskeletal system, I wondered to which extent the five TFs influence the establishment of a functional and proper limb during development. To investigate this, the coding sequence (CDS) of each TF was inserted into a retroviral vector, the RCAS-BP(A), which stands for replication-competent ASLV long terminal repeat with a splice acceptor (Hughes 2004). When transfected into avian cells, this vector allows the production of replication-competent retroviral particles that are externalized, then infect surrounding cells and finally integrate into their genome. Therefore, the overexpression of the recombinant protein is not only restricted to transfected cells and their progeny, but will also spread in dividing cells. Cells producing viral particles carrying each of the TF coding sequences were grafted into embryo forelimbs at early onset of limb development, E2.5 (HH17-18), when limb bud starts forming. Infected and contralateral wings were harvested later at E9.5 for ISH and at E11.5 for skeletal preparation (Figure 16).

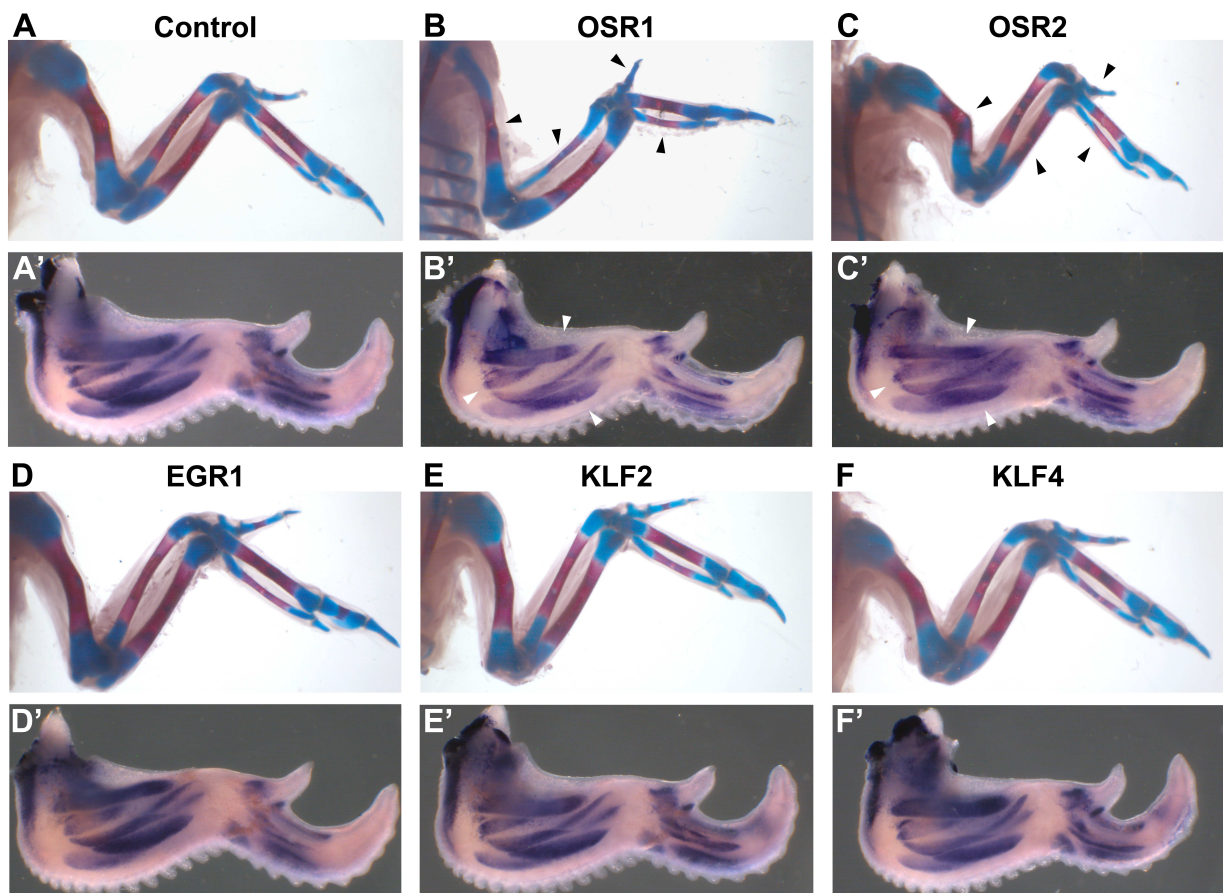


Figure 16. Overexpression of the transcription factors in chick embryo limbs. Grafts of cells producing retroviruses carrying no recombinant protein (A,A'), *OSR1* (B,B'), *OSR2* (C,C'), *EGR1* (D,D'), *KLF2* (E,E') or *KLF4* (F,F') CDS at E2.5 into chick forelimb buds. (A-F) Skeletal preparation of E11.5 forelimbs, bone is stained in red, cartilage in blue. Black arrowheads highlight skeleton defects in humerus, radius, ulna, metacarpals and phalanges. (A'-F') Whole-mount ISH at E9.5 with *MYOG* probes (purple). *MYOG* is a specific-muscle marker. White arrowheads highlight defects in forearm muscles.

Skeletal preparation at E11.5 of forelimbs grafted with cells overexpressing either OSR1, or OSR2, highlighted defects in skeleton (Figures 16B,C; black arrowheads). Although being more pronounced for OSR1 overexpression, both OSR factors induced a reduction of bone size and thickness, visualised with reduced Alizarin Red S staining and a shorter alula (thumb) as compared to the control. In addition, a bending of the humerus was visible for the wings grafted with OSR2-overexpressing cells (Figure 16C). Consistent with the skeleton phenotype, forearm muscles were overall smaller in the wings for which OSR1 and OSR2 were overexpressed (Figures 16B',C'; white arrowheads). These muscle defects are reminiscent to those observed in mutant mice for the MCT markers *Tbx4* and *Tbx5* (Hasson et al. 2010). In contrast to OSR1 and OSR2, overexpression of EGR1, KLF2 and KLF4 did not reveal any obvious phenotype in skeleton and muscle patterning (Figures 16D-F'). I found it surprising considering that ectopic expression of *Egr1* in murine mesenchymal stem cells induces tendon differentiation and prevents their commitment into osteocytes and adipocytes *in vitro* (Guerquin et al. 2013). In addition, *KLF2* and *KLF4* are expressed in chondrogenic condensation regions at E5 in chick embryo limbs, suggesting that they could be involved in cartilage formation (Antin et al. 2010).

The TFs influence cell differentiation processes in the chMM cultures

To further investigate the influence of the five TFs, the chicken micromass (chMM) model was preferred as it partly mimics differentiation processes occurring in the developing limb *in vivo* (Daniels et al. 1996). Mesenchymal cells extracted from E4.5 limb buds were plated in a high-density culture to maintain contacts between cells and kept in culture for five days. To allow overexpression of each TF, cells were mixed with retroviruses carrying each of the TF coding sequences prior to seeding. This combined system of chMM culture with RCAS-BP(A)-mediated overexpression of recombinant proteins has been shown to be particularly relevant to investigate chondrogenesis (Kuss et al. 2009; Stricker et al. 2012; Ibrahim et al. 2013).

Overall morphology of the chMM cultures for each overexpression condition was visualized by Eosin staining (Figures 17A-F). Cartilage differentiation was assessed by Alcian blue staining (Figures 17A'-F'), which displays specific affinity for glycosaminoglycans of the cartilaginous extracellular matrix (ECM). Myotube formation associated with muscle differentiation was estimated by immunohistochemistry against sarcomeric myosin heavy chains (Figures 17A''-F''). Cartilage and muscle defects observed in OSR1- and OSR2-overexpression grafts were confirmed in the chMM cultures. Both OSR factors induced a strong reduction in cartilage nodule and myotube formation after five days of culture (Figures 17B',B'',C',C''). Both differentiation processes were slightly more affected upon overexpression of OSR2 than OSR1 as compared to control cultures (Figures 17G,H; 33% vs. 42% of chondrogenic matrix

production; 53% vs. 63% of formed myotubes). Given that the overall morphology of the chMMs is similar between OSR1/2-overexpressing cultures and control cultures (Figures 17A-C), the reduction of cartilage and myotube formation is more likely due to an effect on cell differentiation rather than an increased cell death.

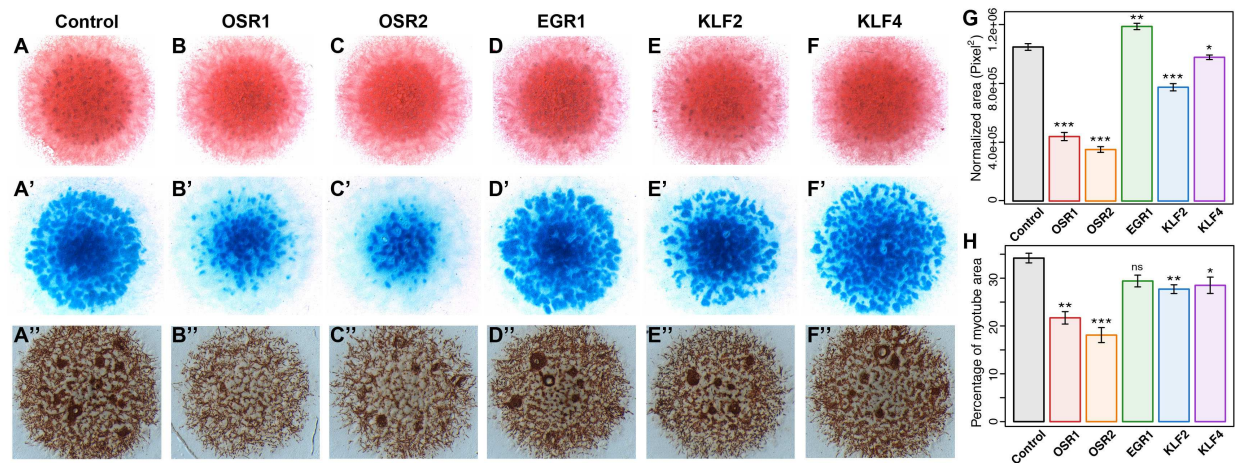


Figure 17. Cell differentiation within the chMM culture conditions. Histological staining of chMM cultures overexpressing no recombinant protein (A,A',A''), OSR1 (B,B',B''), OSR2 (C,C',C''), EGR1 (D,D',D'') KLF2 (E,E',E'') or KLF4 (F,F',F'') after 5 days of culture. (A-F) Cell cytoplasm stained with Eosin. (A'-F') Cartilage nodules stained with Alcian blue. (A''-F'') Myotubes stained by immunohistochemistry with the MF20 antibody targeting sarcomeric myosin heavy chains. (G) Quantification of chondrogenic matrix production by measuring the stained area in each culture condition as compared to the control cultures. (H) Quantification of muscle differentiation level by assessing the percentage of culture area covered by myotubes. Average \pm SEM; paired Student's *t*-test: ns, non-significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

In contrast to the *in vivo* graft experiments that revealed no obvious phenotype, effects on cartilage and muscle differentiation was observed in the chMM cultures overexpressing EGR1, KLF2 and KLF4. Similarly to OSR1 and OSR2, KLF2 induced a reduction of chondrogenic matrix production (Figure 17E') and myotube formation (Figure 17E''), but to a lower extent (Figures 17G,H; 74% and 81%, respectively). Likewise, KLF4 overexpression induced a slight reduction in both cartilage and muscle differentiation within the chMM cultures (Figures 17G,H). Interestingly, cartilage nodules appeared less dense in the center of the micromass but covered almost the entire culture (Figure 17F'). EGR1 was the only factor that increased the chondrogenic matrix production within the chMM cultures when overexpressed (Figures 17D',G; 13% of increase), although its overexpression did not induce any significant change in myotube formation (Figures 17D'',H).

To further investigate the influence of the TFs on cell differentiation processes within the chMM cultures, a gene expression profiling was performed by means of whole-transcriptome sequencing (RNA-seq; Supplemental table S2). RNA-seq data generated for each TF-overexpressing chMM culture were normalized by using DESeq2 (Love et al. 2014; Supplemental figures S1A,B). Normalized fragment counts were converted into transcripts per million (TPM) values (Wagner et al. 2012) to estimate transcript abundances. Genes belonging

to gene ontology (GO) categories related to bone, cartilage, muscle, tendon, adipose tissue and embryonic skeletal development and differentiation (see Materials and Methods section for the complete list of GO accession numbers) were retrieved. The gene list related to tendon was supplemented with the 100 genes identified as the most expressed in mouse limb *Scx*-positive cells at E14.5 (Havis et al. 2014). Transcript abundances for the selected genes were compared between each TF-overexpressing chMM culture and the control cultures (Figure 18).

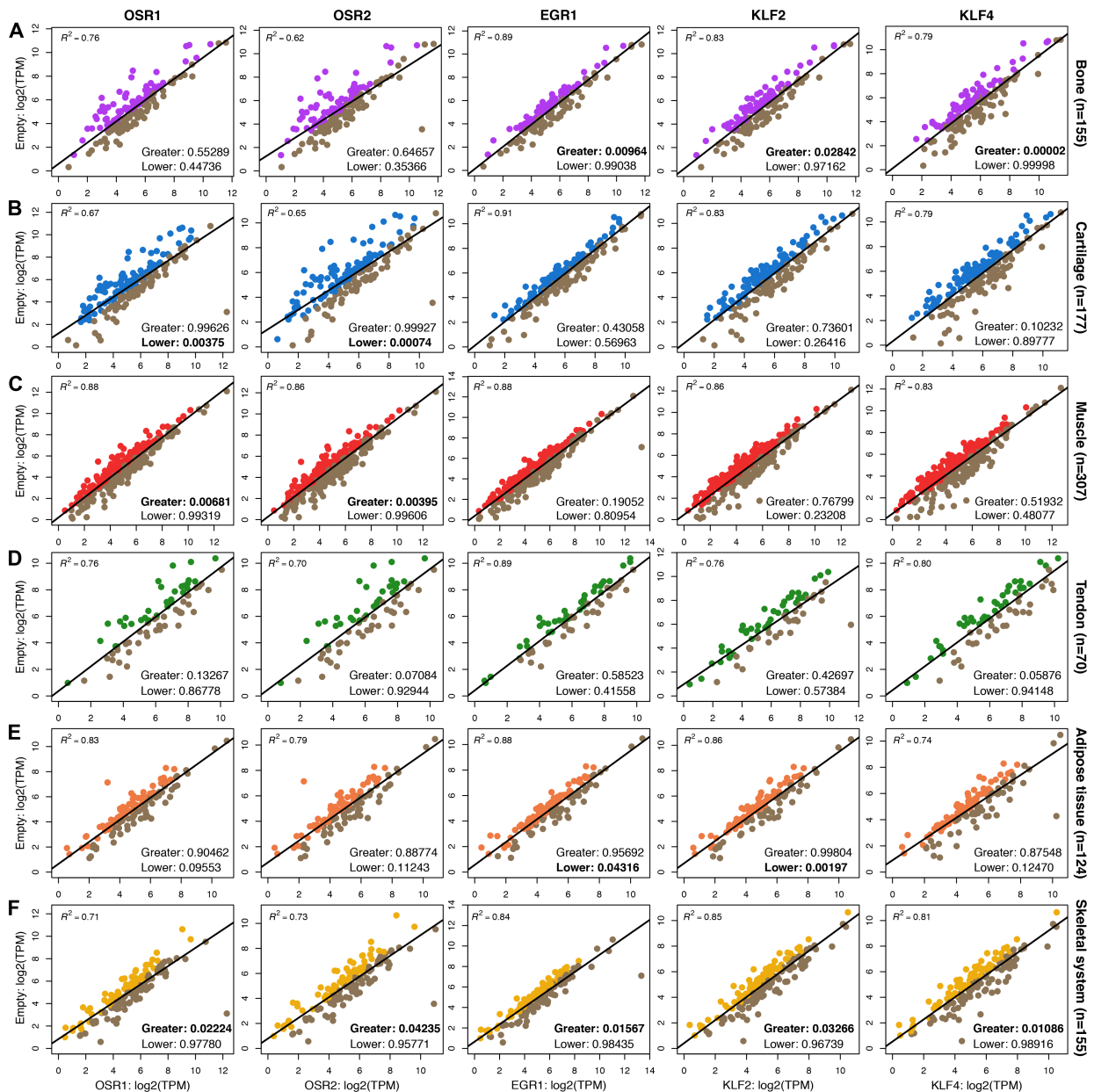


Figure 18. Transcript abundances within GO terms related to cell differentiation and development. Transcript abundances of genes belonging to GO terms related to bone (A), cartilage (B), muscle (C), tendon (D), adipose tissue (E) and embryonic skeletal system (F) differentiation and development. TPM values were averaged between both biological replicates of each chMM culture condition. Only the genes with a mean TPM value of at least 1 across all samples were compared to the controls. Rows, GO terms; columns, overexpressed TF. Genes (dots) were plotted by comparing their abundance between the TF-overexpressing chMM cultures (x-axis) and the control chMM cultures (y-axis). Genes depicted in brown are more abundant in the TF-overexpressing cultures, whereas coloured genes are more abundant in the control cultures. Paired Wilcoxon rank-sum test using both alternative hypotheses: “greater” or “lower” transcript abundances than the control cultures; significant if $P < 0.05$ (bold). TPM, transcripts per million.

Consistent with the phenotype observed in grafted wings and chMM cultures, genes belonging to cartilage GO terms were less abundant in OSR1- and OSR2-overexpressing cultures as compared to the controls (Figure 18B; Wilcoxon rank-sum test, $P < 0.05$). Surprisingly, muscle-associated genes were more abundant in both OSR chMM cultures in spite of the strong reduction of formed myotubes (Figure 18C; Wilcoxon rank-sum test, $P < 0.05$). Genes associated with bone differentiation and development were more abundant in the chMM cultures overexpressing EGR1, KLF2 and KLF4, as compared to the controls (Figure 18A; Wilcoxon rank-sum test, $P < 0.05$), while genes related to cartilage and muscle GO terms remain overall unchanged (Figures 18B,C). Although these TFs are expressed in tendons or in tissues delineating tendons, expression levels of tendon-related genes were similar in EGR1-, KLF2- and KLF4-overexpressing chMM cultures as compared to the controls (Figure 18D). Finally, genes involved in adipose tissue differentiation and development were found less abundant upon overexpression of EGR1 and KLF2 than in the control cultures (Figure 18E; Wilcoxon rank-sum test, $P < 0.05$), which is in agreement with previous observations. Indeed, multipotent murine cells ectopically expressing *Egr1* lose their ability to differentiate into adipocytes (Guerquin et al. 2013); and murine embryonic stem cells deficient for *Klf2* are able to commit into pre-adipocytes, while being unable to differentiate afterwards into adipocytes (Wu et al. 2005).

In general, overexpression of each TF increased the expression levels of genes associated with embryonic skeletal system development and morphogenesis GO terms as compared to the controls (Figure 18F; Wilcoxon rank-sum test, $P < 0.05$). Altogether, these results highlight distinct differentiation potentials of chMM cultures upon overexpression of each TF, especially for OSR1 and OSR2, which strongly reduced cartilage and muscle differentiation.

B. Gene expression profiling

In order to characterize the regulatory function orchestrated by each TF, RNA-seq data obtained for all chMM culture conditions were further examined. Given the changes observed at the cellular level, variations in gene expression levels are expected to occur depending on the overexpressed TF.

Improving the gene expression quantification

Despite many efforts led over the last past decade, the chicken genome remains incomplete and fragmented (Hillier et al. 2004). The chicken karyotype is composed of 38 autosomal chromosomes and 2 additional sexual chromosomes (Bloom et al. 1993). Out of these autosomal chromosomes, 10 are macrochromosomes, with lengths similar to those in mammals, and 28 are

microchromosomes, with lengths ranging from ~2 to less than ~25 Mb (Hillier et al. 2004). Chicken microchromosomes have a high recombination rate, contain an elevated number of repetitive elements and are GC-rich, which significantly induces bias and sequencing errors when using high-throughput technologies (Chen et al. 2013; Dohm et al. 2008). In addition, microchromosomes are gene dense and enriched in CpG islands, which is the result of shorter intronic sequences (McQueen et al. 1998; Smith et al. 2000). Consequently, the fourth version of the *Gallus gallus* genome (galGal4) released in November 2011 is still highly fragmented despite significant improvement since its first release. Out of the 40 chromosomes, 31 are sequenced and retain more than 9,000 gaps. In addition, the genome is composed of ~15,000 additional contigs that are not aligned to any chromosome or assigned with low confidence. In early 2016, the new version galGal5 of the chicken genome has been released on the NCBI, which slightly improve the previous version. Nevertheless, at the time of this thesis, this new version was not available on the UCSC browser yet. Therefore, all the following analyses were performed on the *Gallus gallus* genome version galGal4.

RNA-seq data processing and results are highly dependent on the quality of the genome sequence and the associated gene annotation model. Indeed, quantification of transcript abundances relies on the gene coverage resulting from the primary mapping step of the sequencing reads along the genome. Determination of differentially expressed (DE) genes between two conditions is then based on the variation of the read coverage. It is therefore coherent that an inaccurate definition of gene features would induce a bias in the gene expression quantification and by extension in the detection of DE genes. In order to enhance the RNA-seq data quantification, a transcript-discovery approach was first led to improve the existing galGal4 gene annotation model by using the sequencing reads generated from the control chMM cultures. Strand-specific libraries were prepared to keep the orientation of sequencing reads and therefore to know from which DNA strand genes are transcribed. This library type is particularly useful when two genes located on each strand overlap. On one hand, the transcript-discovery approach was performed by using the reference genome and its associated UCSC/Ensembl gene annotation model as guide (Figure 19A). This first approach intended to more accurately determine exon-intron junctions, to correct or complete existing annotated genes, and to identify unannotated genes. On the other hand, a complementary approach was carried out independently of the genome sequence. RNA-seq reads were assembled *de novo* and then compared to the genes identified with the first approach (Figure 19B). This second strategy was of double purpose: (i) it corrected the gene fragmentation by associating gene parts located on multiple chromosomes or contigs together; (ii) it identified gene regions or complete genes that did not belong to the genome due to the presence of gaps or uncharacterized sequence.

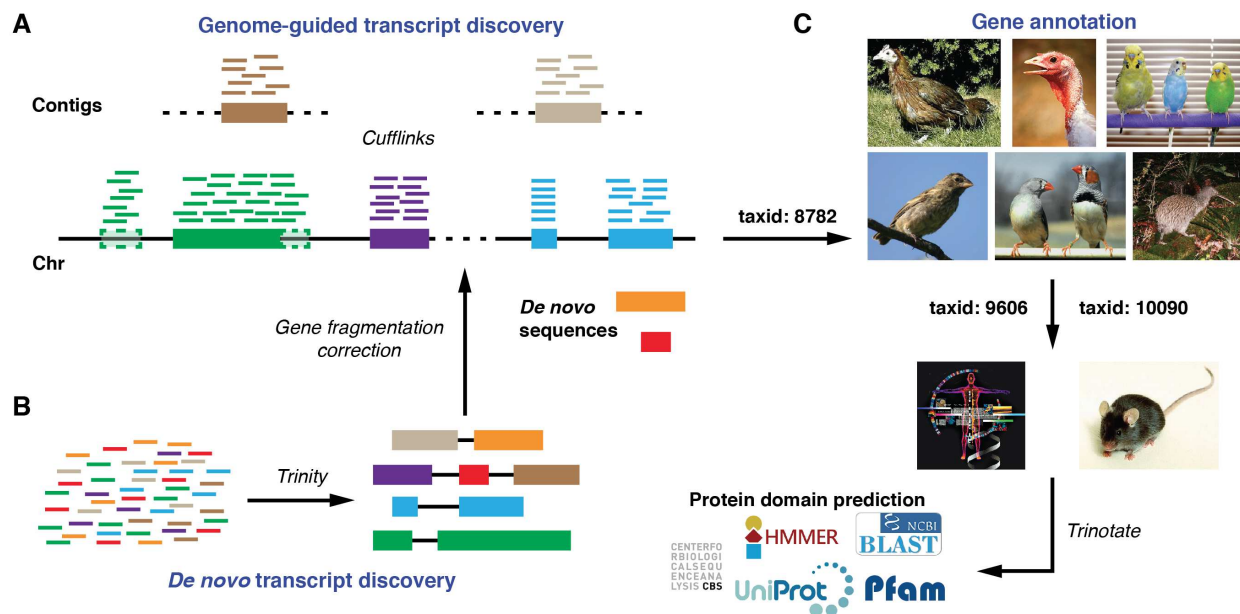


Figure 19. Workflow to design the gene annotation model. Sequencing reads obtained from both biological replicates of the control chMM cultures were used to generate a comprehensive gene annotation model. (A) Transcript-discovery approach performed by using the chicken genome galGal4 as guide to define gene features. (B) *De novo* transcript assembly computed independently of the reference genome to correct the gene fragmentation and to identify uncharacterized genes and gene regions. (C) Functional annotation of the 21,347 identified genes was carried out by comparison against Aves gene sequences (taxid: 8782), human and mouse protein database interrogation (taxid: 9606, 10090) and ORF prediction followed by protein domain identification.

Genes identified by this dual transcript-discovery approach were further annotated by database comparison and protein domain prediction (Figure 19C). Genes were first compared to bird gene sequences, taking advantage of the recent increase of available genomic data within avian species and their high DNA sequence conservation (Zhang et al. 2014). Undefined genes were then compared at the protein level against mouse and human databases. Open reading frame (ORF) and protein domain predictions were finally performed on remaining unannotated genes by using homology search against SwissProt (Bairoch et al. 2004) and Pfam (Punta et al. 2012) databases, and sequence analysis tools to identify transmembrane domains and peptide signals.

The resulting gene annotation model was composed of 21,347 genes. More than 19,000 (~91.0%) genes were found on one single chromosome or unplaced contig, whereas almost 2,000 (9.2%) genes were identified as being fragmented, including 478 (2.2%) genes that were located on multiple ordered chromosomes (Figure 20A). The computed gene annotation was mostly constituted of protein-coding genes (~16,700; ~78.0%), although 672 (3.1%) and 1,410 (6.6%) genes remain partly annotated (putative proteins having at least one protein domain detected) and unannotated (uncharacterized proteins with no protein domain identified but an ORF of at least 100 amino acids), respectively (Figure 20B). Remaining genes corresponded to miscellaneous genes (213 genes, 1.0%; such as spliceosome complex members, ribosomal RNAs and pseudogenes) and non-coding RNAs (4,418 genes, 20.7%) for which no sufficient ORF could be predicted (Figure 20B).

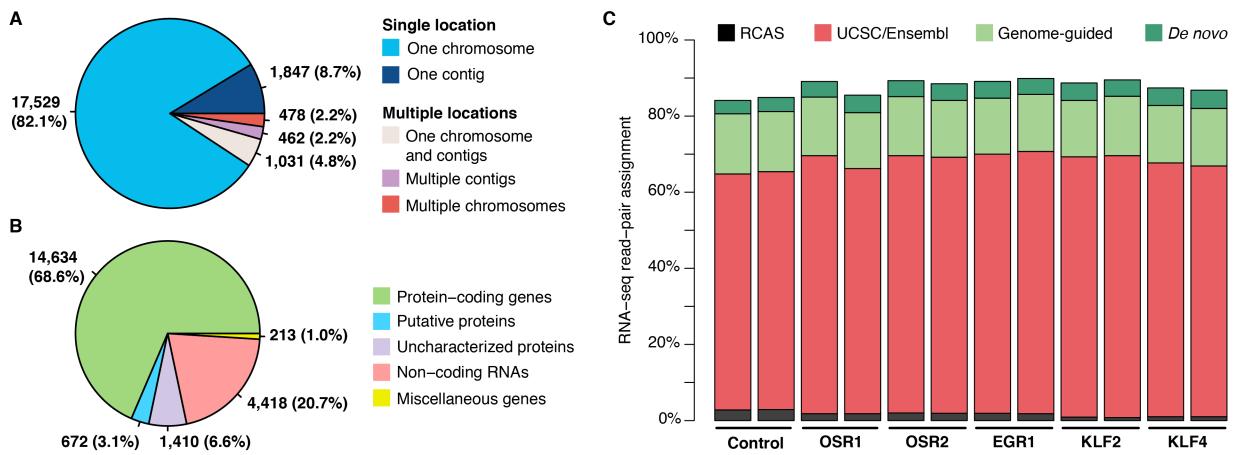


Figure 20. Characteristics of the established gene annotation model. The gene annotation model designed from the dual transcript-discovery approach was composed of 21,347 genes. **(A)** Proportion of gene location on chromosomes and contigs of the chicken reference genome galGal4. 9.2% of identified genes are fragmented due to their location on multiple chromosomes and contigs. **(B)** Annotated gene biotypes. Putative proteins correspond to genes for which at least one protein domain could be detected. Uncharacterized proteins are genes with an ORF of at least 100 amino acids without protein domain identified. Genes with no ORF predicted were classified as non-coding RNAs. **(C)** Sequencing read assignment across all samples and conditions depending on the gene annotation model used. The combined transcript-discovery approach raised the assignment rate of 20-22% as compared to the UCSC and Ensembl reference annotations.

Most importantly, the dual transcript-discovery approach significantly improved the percentage of RNA-seq read pairs assigned to gene features (Figure 20C). Indeed, between 83 and 90% of sequencing read pairs were assigned to gene features when using the newly designed gene annotation model as reference, as compared to an assignment rate of 62-69% when using both galGal4 UCSC and Ensembl reference annotations. In other words, the transcript-discovery approach enabled to retrieve 20-22% more information from the RNA-seq data across all samples and conditions.

Estimation of the levels of transcription factor overexpression

Overexpression of the TFs in the chMM cultures was performed by using the RCAS-BP(A) system. During cell infection, the RCAS-BP(A) envelope glycoprotein interacts with a specific surface receptor of the host cell (Hunter 1997). The retroviral genome is then introduced into the host cell and integrates into its genome after reverse transcription. As being replication-competent, new viral particles are produced and can then infect surrounding cells if not already infected. Indeed, the expression of the envelope glycoprotein within the host cell blocks the surface receptor and prevents infection of the cell by additional retroviruses of the same subgroup. Retroviral infection of the chMM cultures was performed by using a ratio virus/cell of 1:1, which leads to the assumption that cells contain only one copy of the RCAS-BP(A) genome. In addition to be stable, reproducible and independent of the transfection efficiency, RCAS-BP(A) overexpression of recombinant proteins is moderate as compared to other viral vectors such as the CMV and SV40. Indeed, a recent study using the same combined system of chMM

culture and RCAS-BP(A) infection determined by absolute quantification that the levels of overexpression are approximately between 3 and 5 folds (Ibrahim et al. 2013).

The transcription along the RCAS-BP(A) genome gives rise to three distinct splicing variants (Hughes 2004; Figure 21A). Although the three isoforms carry the inserted TF coding sequence, only the shorter one encodes a functional protein. To evaluate the correct overexpression level of the recombinant protein, it is therefore necessary to first distinguish the proportion of each splicing variant. Across all chMM culture conditions, the TF-encoding isoforms corresponded from one third to one fifth of all splicing variants (Figure 21B).

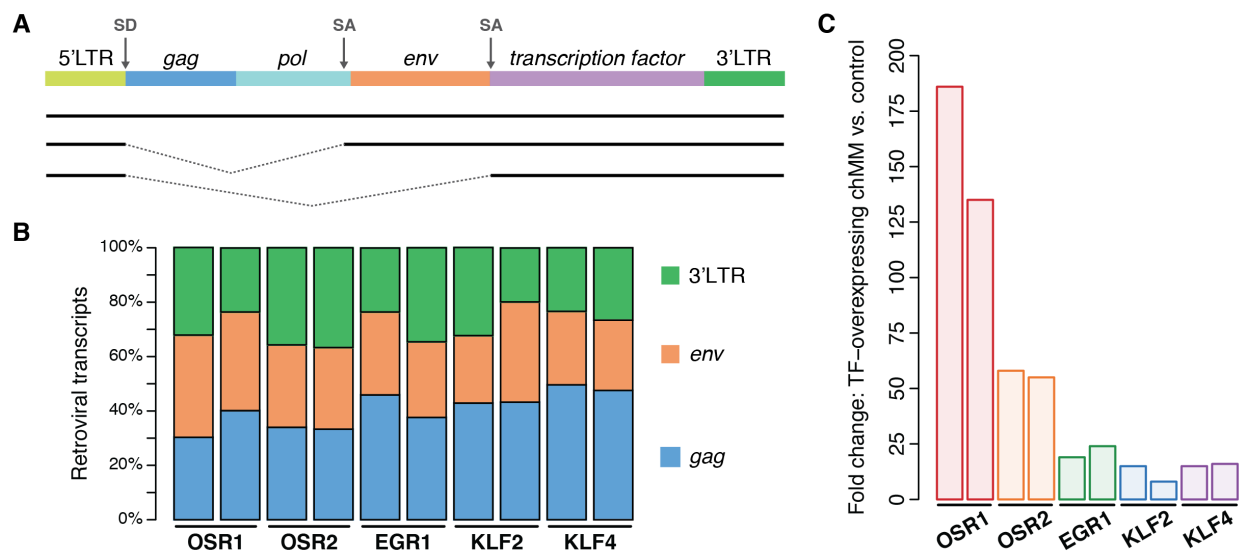


Figure 21. Organization of the viral genome. (A) RCAS-BP(A) retroviruses express three alternative transcripts. The first isoform is used for *gag* and *pol* expression; the second variant encodes the envelope glycoprotein (*env*); and the last transcript is the only variant that gives rise to the expression of a functional TF. SD, splice donor site; SA, splice acceptor site; LTR, long terminal repeat. (B) Proportion of each retroviral splicing variant across the chMM culture conditions and replicates. Transcript abundances were estimated by comparing the read coverage along three regions unique to each splicing variant (*gag*, *env* and 3'LTR). (C) Fold change of expression levels for each TF virally overexpressed as compared to its endogenous level within the control chMM cultures.

Since both endogenous and retroviral expression levels cannot be distinguished from the RNA-seq data, expression fold change of TFs was calculated as compared to their abundance within the control cultures, assuming that the endogenous expression level of each TF did not change upon retroviral overexpression. *EGR1*, *KLF2* and *KLF4* were overexpressed by 8-25 folds in their corresponding chMM cultures as compared to the control cultures (Figure 21C). Surprisingly, the overexpression level was of ~160 and ~56 folds for *OSR1* and *OSR2* in the *OSR1*- and *OSR2*-overexpressing cultures, respectively (Figure 21C), suggesting that their endogenous expression may have increased upon overexpression. One possible explanation is that *OSR1* and *OSR2* positively regulate their own expression by a positive feedback loop, so-called “autogenous regulation”. This mechanism has been mainly described in bacteria (reviewed in Wall et al. 2004), but it has been also identified in the activation of the Mos-MEK-p42 MAPK cascade during the maturation of *Xenopus* oocytes (reviewed in Ferrell 2002). A

second alternative hypothesis is related to the cell differentiation effects within the TF-overexpressing chMM cultures. Overexpression of OSR1 and OSR2 strongly reduced the differentiation potential of limb mesenchymal cells into cartilage and muscle to promote their differentiation into irregular connective tissue, as previously observed with the mouse TFs (Stricker et al. 2012). The increased expression levels of OSR1 and OSR2 may then result from the increased number of connective tissue cells, which express endogenously each factor.

The gene expression profiles in culture recapitulate the limb expression patterns

To further investigate the regulatory function of each TF, RNA-seq data obtained for each TF-overexpressing chMM culture were analysed by using DESeq2 (Love et al. 2014). Computing the Euclidean distance across all chMM culture conditions and biological replicates revealed a consistency between gene expression profiles and gene expression patterns observed by ISH (Figure 22).

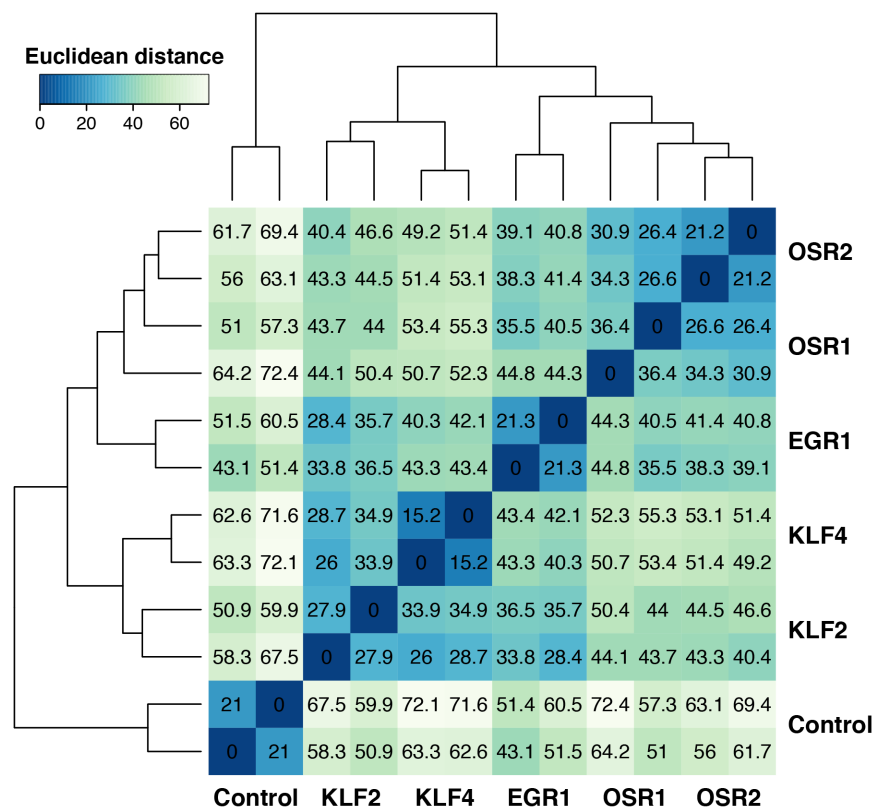


Figure 22. Sample-to-sample distance across chMM cultures. Euclidean distance was calculated across all biological replicates and conditions by using the regularized-logarithm transformed fragment counts. Gene expression profiles are clustered in accordance with the TF expression patterns observed by ISH.

In agreement with the unique phenotypes observed in the grafts and the chMM cultures, both OSR factors are closer to each other than to any other TF regarding their gene expression profiles (Figure 22). Hierarchical clustering highlighted a second cluster composed of EGR1,

KLF2 and KLF4 gene expression profiles, in which both KLF factors were gathered together (Figure 22). These two main clusters are highly consistent with the gene expression patterns observed by ISH. Indeed, OSR1 and OSR2 are expressed in the MCT, whereas EGR1, KLF2 and KLF4 expression are detected in tendons or in tissues surrounding tendons. Finally, gene expression profiles within the chMM cultures infected with retroviral particles carrying no recombinant protein were separated to all other conditions, indicating that overexpression of each TF induced changes at the molecular level. Differential expression analysis was then performed by using DESeq2 and a false-discovery rate (FDR) of 0.01. Genes with a fold change higher than 2 or lower than 0.5 and a Benjamini-Hochberg adjusted p-value (padj) below 0.01 were considered as being differentially expressed (DE). Between 1,369 and 2,907 DE genes were thus detected for each TF-overexpressing chMM culture as compared to the control cultures, resulting in a total of 10,712 DE genes across all the chMM cultures (Figure 23). Almost two thirds (6,956; 64.9%) of these genes were found upregulated.

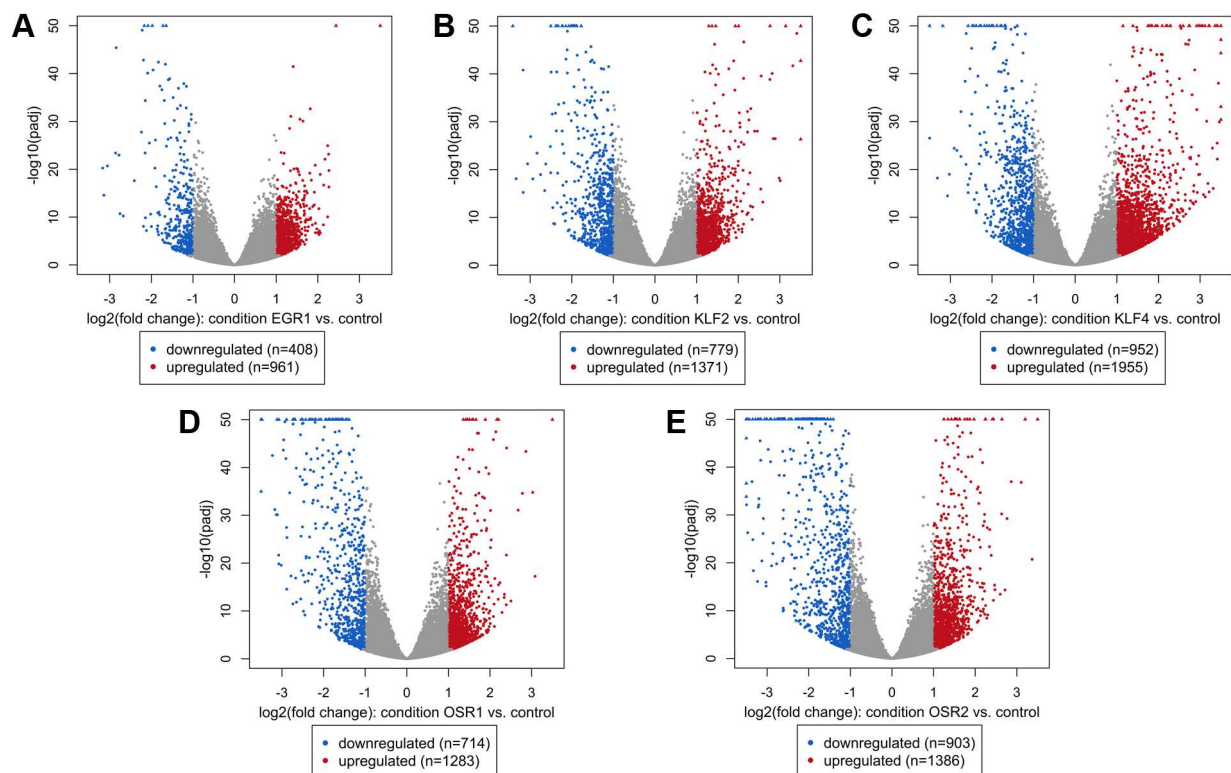


Figure 23. 10,712 differentially expressed genes in all chMM cultures. Volcano plots of the 10,712 DE genes detected in the chMM cultures overexpressing EGR1 (A), KLF2 (B), KLF4 (C), OSR1 (D) or OSR2 (E) as compared to the control cultures. (A) 1,369 DE genes detected upon EGR1 overexpression. (B) 2,150 DE genes detected upon KLF2 overexpression. (C) 2,907 DE genes detected upon KLF4 overexpression. (D) 1,997 DE genes detected upon OSR1 overexpression. (E) 2,289 DE genes detected upon OSR2 overexpression. DE genes: $\log_2(\text{fold change}) \geq 1$ or ≤ -1 ; $\text{padj} < 0.01$; FDR 0.01. Downregulated genes are depicted in blue, upregulated genes in red. padj , Benjamini-Hochberg adjusted p-value.

The TFs share common regulatory patterns

The 10,712 DE genes across all the chMM culture conditions corresponded to a list of 4,298 non-redundant genes, indicating that the TFs share common regulatory targets (Figure 24A).

143, 330, 114, 133 and 767 genes were identified as being differentially expressed specifically upon overexpression of OSR1, OSR2, EGR1, KLF2 and KLF4, respectively (Figure 24A). 2,811 (65.4%) genes were regulated by at least two TFs and 726 (16.9%) DE genes were identified in all chMM cultures (Figure 24A). Although being expressed in distinct subcompartments of the musculoskeletal system, the TFs seem to share a core of common regulatory processes.

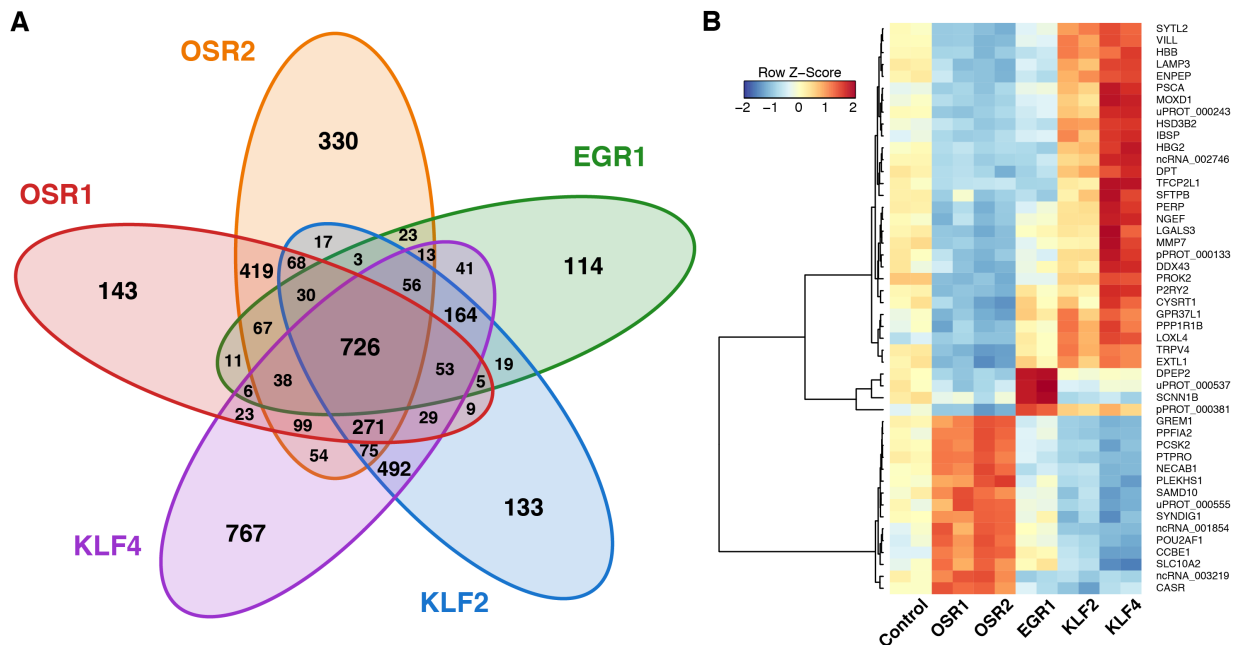


Figure 24. 4,298 unique differentially expressed genes across all chMM cultures. (A) Venn diagram of the 10,712 DE genes identified in all chMM culture conditions. Given the high number of common regulated genes, 4,298 non-redundant genes out of the 10,712 DE genes (40.1%) were found. (B) Heatmap of the 48 genes regulated in opposite directions across all chMM cultures. Hierarchical clustering was performed by using the one minus Pearson correlation.

Interestingly, fold change comparison (whether the gene is upregulated or downregulated) of the DE genes shared by at least two TFs revealed a high consistency among the TF regulatory patterns. Only 48 (1.7%) shared DE genes were identified as being regulated in opposite directions between the subset of TFs that misregulate them (Figure 24B). Among these 48 genes, *IBSP* (also known as *BSP*), a gene encoding a bone sialoprotein, was found upregulated by KLF2 and KLF4 while being downregulated by OSR1 and OSR2. *IBSP*, which is expressed by skeletal-associated cells such as chondrocytes and osteoblasts (Bianco et al. 1991), is a major component of the bone non-collagenous ECM (Fisher et al. 1990). *Ibsp* knockout mice are smaller and display impaired bone growth and repair (Malaval et al. 2009). In contrast, *GREM1* (Gremlin 1) was upregulated by OSR1 and OSR2, and repressed upon KLF2 and KLF4 overexpression. This gene encodes a member of the BMP antagonist family involved in limb bud progression by controlling the feedback loop between SHH and FGF4 (Zuniga et al. 1999). *In vivo* experiments in chick embryos showed that Gremlin regulates early limb outgrowth and represses chondrogenesis (Merino et al. 1999). In addition, *GREM1* expression is detected in

mature skeletal myofibres and interstitial muscle cells suggesting a role during myogenesis (Frank et al. 2006). Considering their distinct role in promoting and repressing chondrogenesis, the opposite regulation of *IBSP* and *GREM1* between both OSR and KLF factors is overall correlated with the previous observations in the chMM cultures and the *in vivo* graft experiments.

Given the high consistency in the TF regulatory patterns and the elevated number of shared targeted genes, a gene clustering approach was led on the 4,298 non-redundant DE genes by using *K*-means (Hartigan and Wong 1979). This method intends to partition a list of genes into a given number of clusters such that genes belonging to a same cluster are as similar as possible in term of their expression pattern, while being as distant as possible of the expression pattern of genes belonging to the other clusters. By using this approach, the 4,298 unique DE genes were partitioned into 8 clusters (Figure 25A). A gene ontology (GO) analysis was further performed in order to identify potential biological processes enriched within each cluster (Figure 25B).

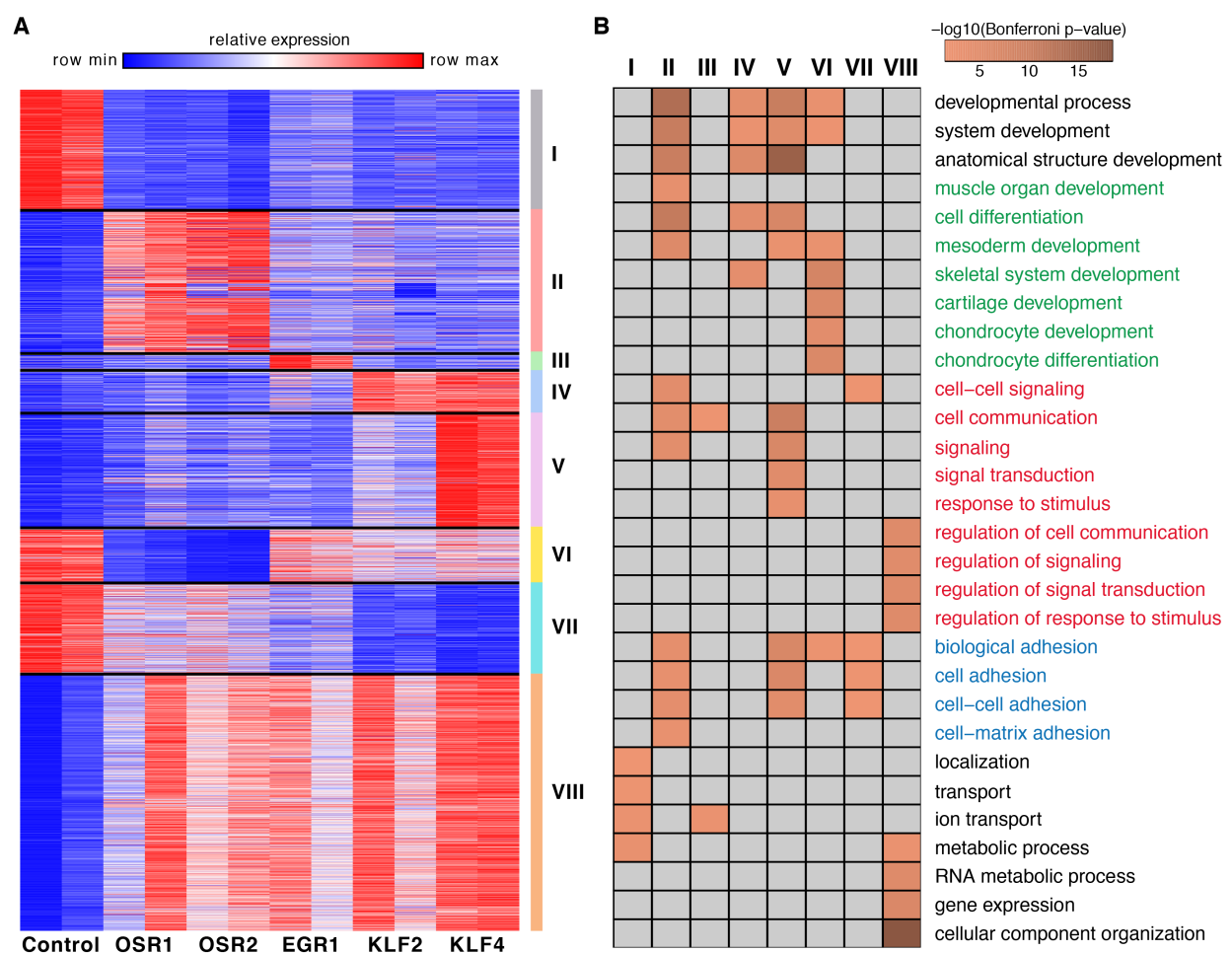


Figure 25. *K*-means gene clustering of the differentially expressed genes. (A) Gene clusters identified by *K*-means partitioning on the 4,298 non-redundant DE genes. The number of *K* clusters was fixed at 8 because lower values did not separate distinct gene clusters and higher values subdivided meaningful gene clusters. (B) GO analysis for biological processes of the DE genes belonging to each *K*-means cluster. GO terms related to cell differentiation and development are depicted in green, cell signalling and communication in red, cell adhesion in blue. Clusters having no significant enrichment for the specified GO terms are depicted in grey.

The cluster I was composed of downregulated genes across overexpression of each TF (Figure 25A), which are mainly involved in ion transport and metabolic processes (Figure 25B). By contrast to this cluster, the cluster VIII was enriched for genes being upregulated by all TFs within the chMM cultures (Figure 25A). These genes are associated with metabolism, gene expression, cellular component organization and the regulation of cell signalling and communication (Figure 25B). The clusters II, III, IV and V corresponded to genes upregulated specifically by each TF or couple of closely related TFs (Figure 25A). Genes positively regulated by the TFs are mainly enriched for cell differentiation, mesoderm development, cell signalling and biological adhesion (Figure 25B). Consistently with the previous observation that genes associated with muscle differentiation and development are more abundant in OSR1- and OSR2-overexpressing chMM cultures as compared to the native chMM cultures, Both OSR factors positively regulate genes involved in muscle organ development (Figure 25B; cluster II). The cluster VI, which depicted an opposite expression pattern to the cluster II, is composed of genes downregulated by OSR1 and OSR2 overexpression (Figure 25A). Genes belonging to this cluster are enriched for biological processes related to chondrogenesis (Figure 25B), which is in agreement with the cartilage phenotype and reduction observed in grafted wings and chMM cultures, respectively. The cluster VII, contrasting to the cluster IV, corresponded to genes being downregulated by KLF2 and KLF4 overexpression (Figure 25A) and related to cell signalling and adhesion (Figure 25B).

The TFs share common regulatory functions

The gene clustering approach revealed that a major proportion of genes upregulated by the selected connective tissue-associated TFs are involved in signal transduction and biological adhesion. To further investigate this interesting feature, signalling pathway enrichment analysis was performed on the complete set list of DE genes identified for each TF independently. Consistent with the previous observation that the TFs share common regulatory patterns, the TF-associated DE genes were related to common signalling pathways (Figure 26A). Of remarkable interest, Integrin, Cadherin and WNT signalling pathways were enriched across all chMM culture conditions. On the other hand, some signalling pathways were rather enriched for a subgroup of TFs (Figure 26A). In particular, Axon guidance regulation mediated by netrin was associated with both OSR factors; Cytoskeletal regulation by Rho GTPase was common to OSR1, OSR2 and EGR1; TGF-beta signalling pathway was enriched for OSR1, OSR2 and KLF2; whereas Notch signalling pathway was highlighted in both KLF factors.

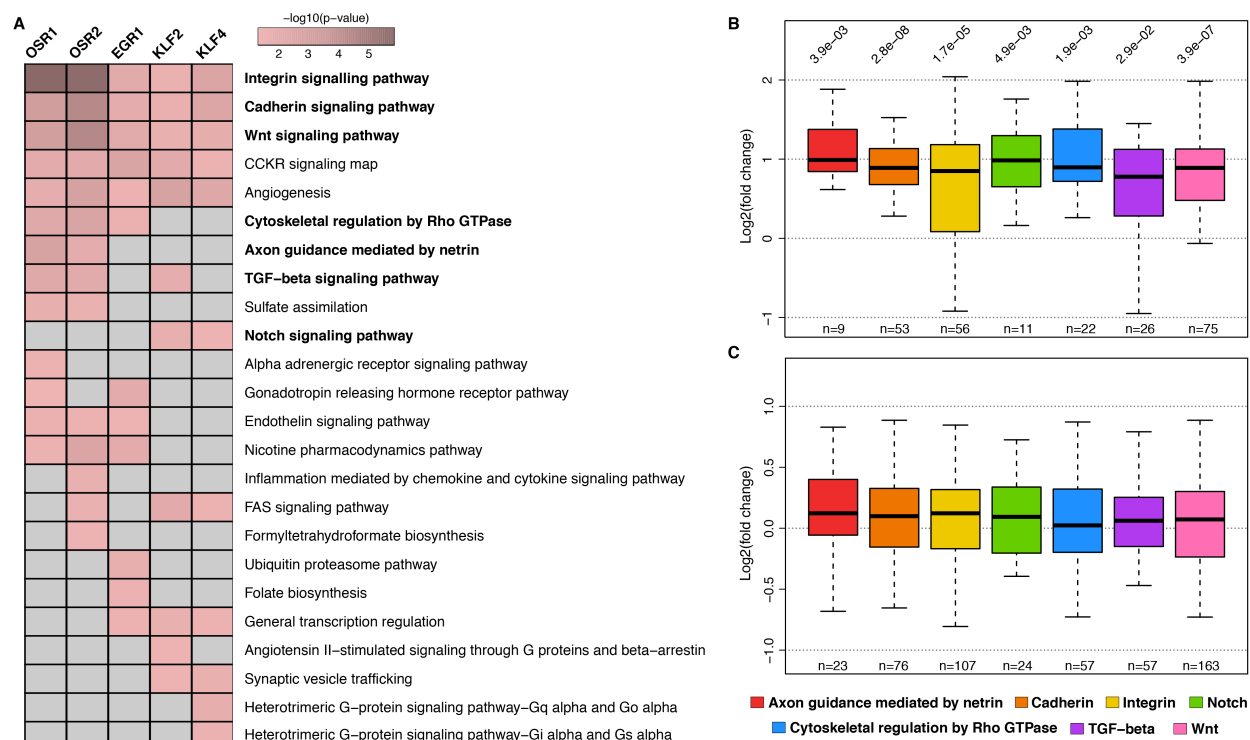


Figure 26. Signalling pathway enrichment analysis of the differentially expressed genes. (A) GO analysis on the DE genes identified upon overexpression of each TF within the chMM cultures. Signalling pathways highlighted in bold were further analysed. DE genes having no enrichment for the specified Panther pathway are depicted in grey. (B,C) Boxplots of the global expression level for the DE genes (B) and non-DE genes (C) belonging to the selected Panther pathways. Log₂ fold changes of each gene were averaged across all chMM culture conditions and replicates. Number of genes (n) is indicated on the bottom. Statistical test: paired Wilcoxon rank-sum test, p-values are indicated on the top, significant if $P < 0.05$.

DE genes associated with each aforementioned signalling pathway were retrieved and further processed. By comparing the averaged fold change across all TFs for each DE gene, it appeared that DE genes within each selected pathway were overall significantly upregulated with a median log₂ fold change close to 1 (Figure 26B; Wilcoxon rank-sum test, $P < 0.05$). This tendency was not observed for the remaining non-DE genes associated with each of the signalling pathway (Figure 26C; median log₂ fold change close to 0). Therefore, it seems that the TFs positively activate these signalling pathways by upregulating a core of targeted genes. Nevertheless, a proportion of DE genes associated with the Integrin and TGF-beta signalling pathways rather appeared downregulated (Figure 26B; lower whisker). Closer look of these two pathways highlighted a set of genes specifically repressed by both OSR factors (Supplemental figures 4A,B). Most of these downregulated genes are associated with cartilage and bone development, which is consistent with the previous observations that overexpression of OSR1 and OSR2 represses chondrogenesis in grafted forelimbs and chMM cultures.

Considering the strong enrichment for biological processes related to signal transduction, cell communication and biological adhesion, DE genes should encode proteins associated with these functions such as receptors, secreted molecules and ECM proteins. To validate this hypothesis,

overrepresentation analyses on cellular component, molecular function and protein class were carried out on the 4,298 non-redundant DE genes. As expected, DE genes identified across all TFs were enriched for proteins associated with the ECM, plasma membrane and cytoskeleton (Figure 27A). This was confirmed by the molecular function and protein class GO analyses with enrichment for receptor, ECM and cytoskeletal proteins as well as signalling and cell adhesion molecules (Figures 27B,C). DE genes encoding proteins associated with cell adhesion, cytoskeleton, ECM and signal transduction were further investigated. The averaged fold change was calculated across all TFs for each DE gene associated with these protein classes. Similarly to the signalling pathways, DE genes within each selected protein class were overall significantly upregulated (Figure 27D; median log₂ fold change close to 1; Wilcoxon rank-sum test, $P < 0.05$), as compared to the remaining non-DE genes (Figure 27E; median log₂ fold change close to 0).

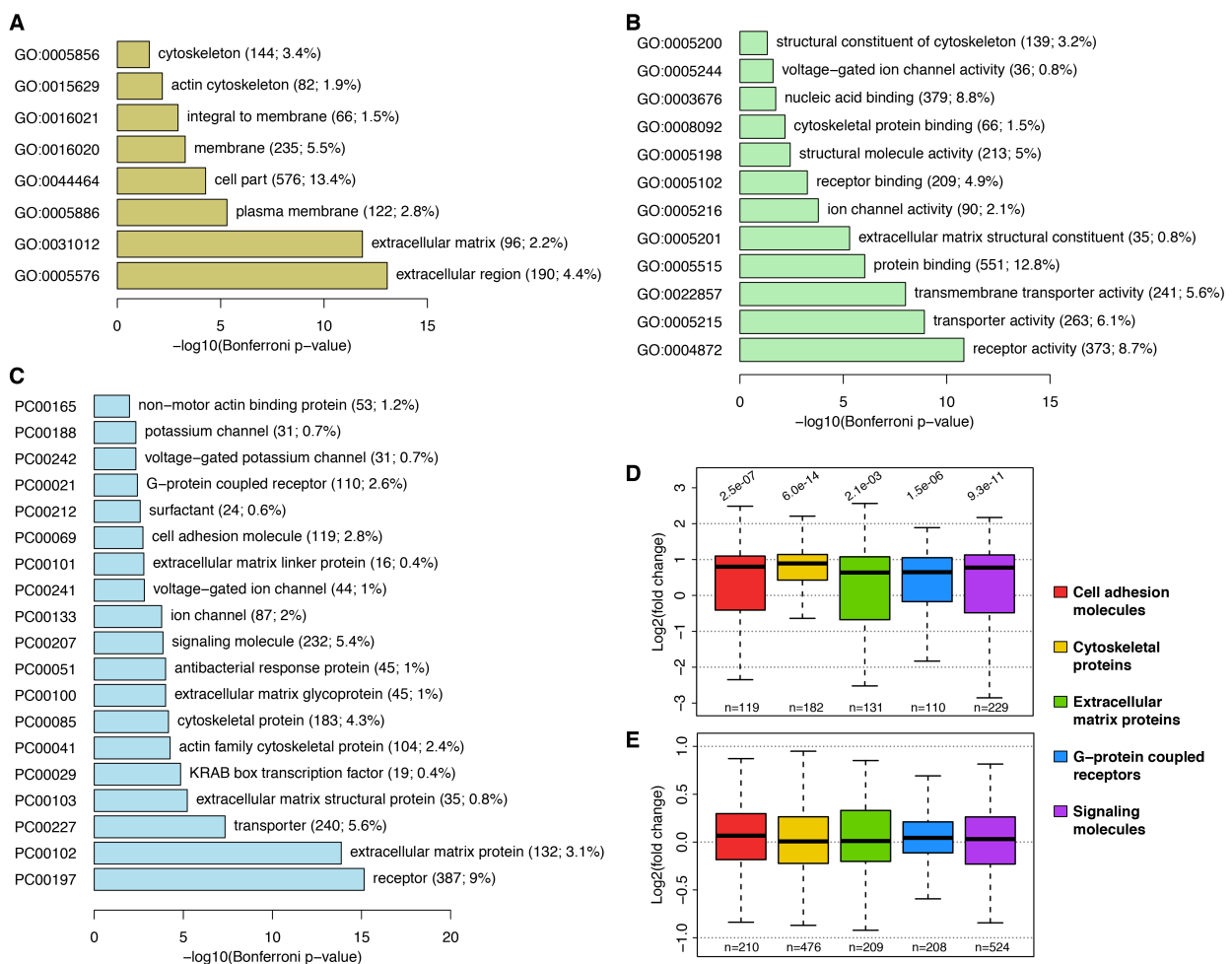


Figure 27. GO enrichment analyses on the differentially expressed genes. (A,B,C) Overrepresentation test of the 4,298 non-redundant DE genes for cellular component (A), molecular function (B) and protein class (C) GO terms. (D,E) Boxplots of the global expression level for the DE genes (D) and non-DE genes (E) belonging to the selected protein classes. Log₂ fold changes of each gene were averaged across all chMM culture conditions and replicates. Number of genes (n) is indicated on the bottom. Statistical test: paired Wilcoxon rank-sum test, p-values are indicated on the top, significant if $P < 0.05$.

Altogether, the gene expression profiling of the chMM cultures overexpressing each TF highlighted a core of common regulatory functions across all TFs. The connective tissue-associated TFs positively regulate biological processes related to cell signalling, communication and adhesion. Given that the TFs belong to distinct subcompartments of the musculoskeletal system, it is tempting to associate these biological functions more broadly. Connective tissues could be then involved in establishing an environment where cell-cell interactions and communications would take place and contribute to the development of a proper limb.

C. Chromatin landscape

The activation or the repression of a gene is highly dependent on a biological process mediating the chromatin accessibility along promoter and enhancer regulatory domains. Promoters and enhancers are associated with specific histone tail post-translational modifications, which can be assessed genome-widely by means of chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq). To further investigate the TF regulatory patterns, I performed ChIP-seq on five histone modifications frequently detected in promoters, enhancers and repressive islands (Supplemental table S3).

Defining the regulatory domains

The chromatin landscape was assessed in two independent biological replicates of 5-day chMM cultures infected with native retroviral particles carrying no TF coding sequence (Supplemental figure S5). Five histone modifications were investigated by ChIP-seq: mono-, bi- and trimethylation of lysine 4 of the histone protein H3 (H3K4me1, H3K4me2 and H3K4me3), acetylation and trimethylation of lysine 27 of the histone protein H3 (H3K27ac and H3K27me3). Histone mark coverage profiles for each biological replicate were generated by normalizing the ChIP-seq enrichment signal against the corresponding input controls (sonicated DNA sequenced to measure chromatin accessibility). Similarity across all histone modifications and replicates was assessed by computing the Pearson correlation. Biological replicates for each histone modification were highly consistent and clustered together (Figure 28A). H3K4 bi- and trimethylated were strongly correlated and enriched at the transcriptional start site (TSS) of annotated transcripts (Figures 28A,B). H3K27ac was more correlated with H3K4me3 than with any other histone mark and detected surrounding the TSS (Figures 28A,B), as expected since these both modifications are associated with transcriptional activation (Santos-Rosa et al. 2002; Creyghton et al. 2010). In contrast, H3K27me3 was weakly detected at TSSs and H3K4me1 was not globally enriched in the vicinity of TSSs (Figure 28B).

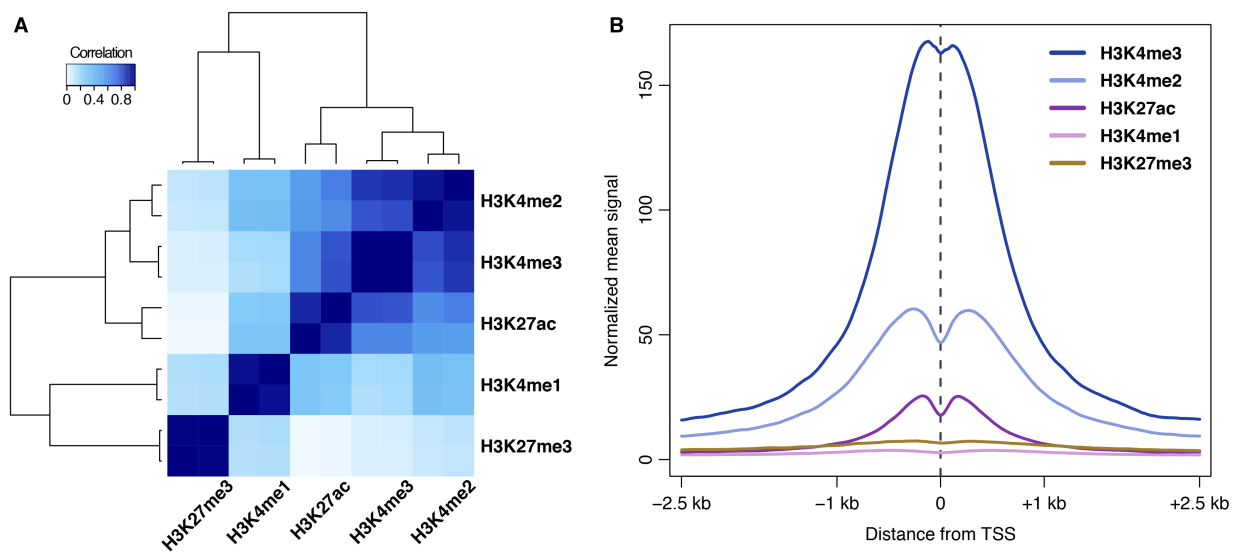


Figure 28. Histone modification signal profiles. (A) Similarity across histone modification enrichment profiles and biological replicates. Signal coverage was normalized against the input controls by using the negative log₁₀ of the Poisson p-value. Pearson correlation across all histone modifications and biological replicates was computed in 500-bp non-overlapping windows along the whole genome. (B) Normalized mean histone ChIP-seq signal surrounding the TSS of all transcripts. Signal was normalized by using the negative log₁₀ of the Poisson p-value. TSS, transcriptional start site.

Peak calling of the different ChIP-seq profiles was performed by following the ENCODE and Roadmap Epigenomics consortiums' recommendations (Kellis et al. 2014; Roadmap Epigenomics Consortium et al. 2015). Only the peaks that were identified in both biological replicates for each histone mark were further processed. Regulatory domains were determined by combining the different histone modification profiles. Due to the chromosome fragmentation and therefore gene fragmentation of the chicken genome, promoters and enhancers were not defined according to the TSS position, as it is usually performed. H3K4me3 is a mark that is frequently found at promoter regions (Barski et al. 2007; Heintzman et al. 2007). Although it can also be observed at enhancer regions (Pekowska et al. 2011), the presence of H3K4me3 signal was used to identify promoter regions, since it is strongly enriched surrounding the TSS (Figure 28B). Consistent with its association to enhancer regions (Heintzman et al. 2007) and low enrichment at the TSS (Figure 28B), H3K4me1 signal was used to define enhancers. H3K4me2 is a mark that is found at both promoter and enhancer regions (He et al. 2010; Heintzman et al. 2007; Kaikkonen et al. 2013). Remaining H3K4me2 regions, which were devoid of H3K4me1 and H3K4me3 signals, were classified as promoters. In total, 20,427 promoters and 55,597 enhancer regions were thus defined (Figures 29A,B). H3K27me3, a Polycomb group-associated repressive mark (Cao et al. 2002; Czermin et al. 2002; Kuzmichev et al. 2002; Müller et al. 2002), is a signature of facultative heterochromatin. Regions enriched for this histone modification and devoid of any other mark were defined as repression islands, which accounted for 71,664 domains (Figure 29C).

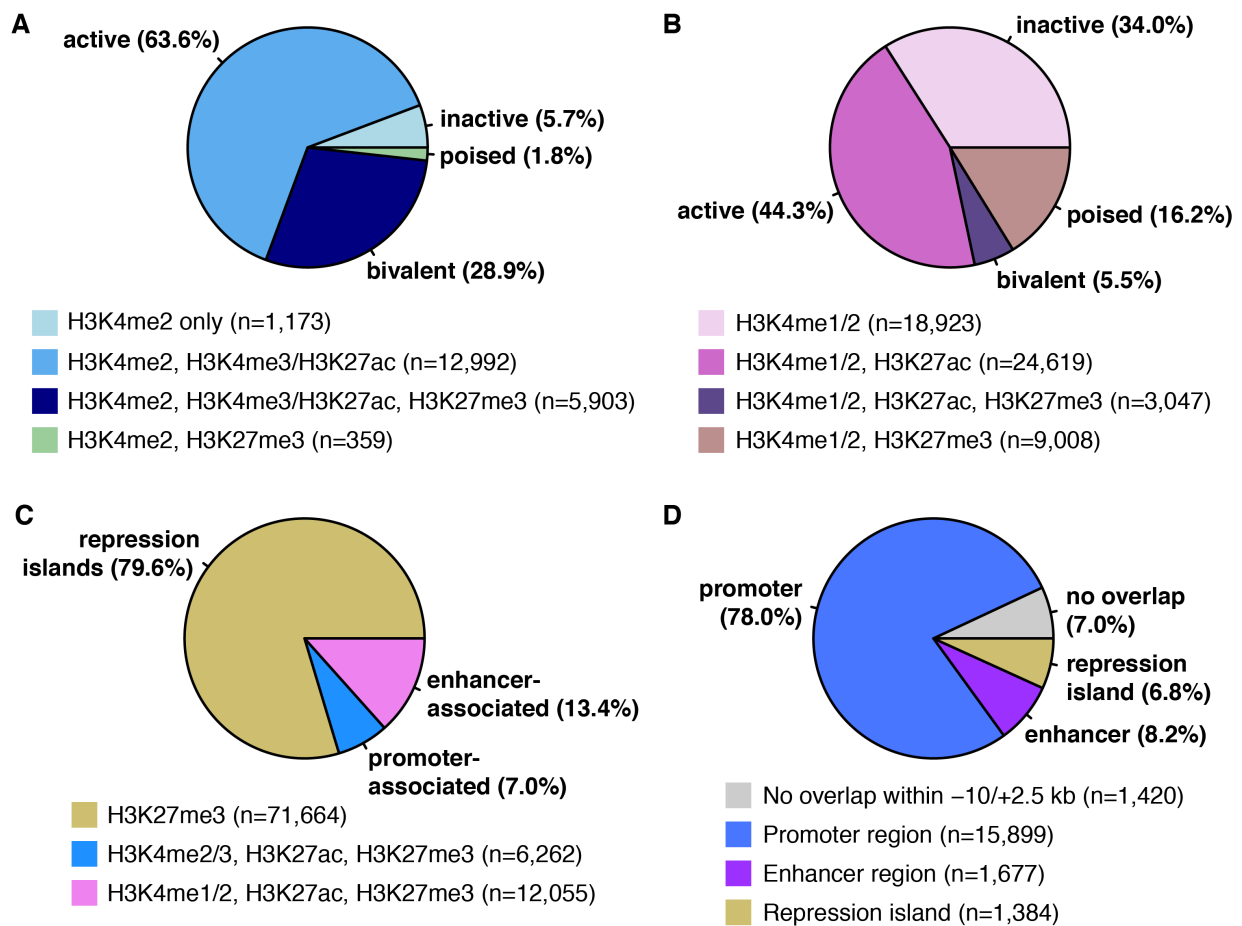


Figure 29. Regulatory domains. (A) 20,427 identified promoters. (B) 55,597 identified enhancers. (C) 89,981 regions associated with the H3K27me3 repressive mark. (D) Association of regulatory domains with gene TSSs. Chromatin domains were searched in the region surrounding TSS positions ranging from 10-kb upstream to 2.5-kb downstream. The regulatory domain located the closest to the TSS position was associated with the corresponding gene. Number of regions (n) for each category is indicated.

Promoter and enhancer regulatory domains were then further divided into four chromatin states: (i) inactive; (ii) poised; (iii) active; and (iv) bivalent. The distinction of the four states was based on the presence or absence of H3K4me3 and H3K27ac, which are both marks enriched in the vicinity of actively transcribed regions, as well as on H3K27me3, which is associated with facultative heterochromatin. H3K27ac is a modification detected in both active promoter and enhancer (Creighton et al. 2010; Heintzman et al. 2009), whereas H3K27me3 is a signature of poised enhancer (Rada-Iglesias et al. 2011) and bivalent promoter (Azuara et al. 2006; Bernstein et al. 2006). Inactive regulatory domains corresponded to promoters and enhancers only enriched for H3K4me2 and H3K4me1/2, respectively (Figures 29A,B; Table 1). Poised domains were characterized by the absence of active mark (H3K4me3 and/or H3K27ac) and the presence of the repressive mark H3K27me3 (Figures 29A,B; Table 1). Active domains were defined by the unique presence of an active mark (promoters: H3K4me3 and/or H3K27ac; enhancers: H3K27ac), whereas bivalent domains were composed of both active and H3K27me3 repressive

marks (Figures 29A,B; Table 1). Overall, the repression-associated H3K27me3 domains were identified in 6,262 (7.0%) promoters and 12,055 (13.4%) enhancers (Figure 29C).

Table 1. Promoter and enhancer regulatory domains. Red, active marks; brown, repressive mark.

State	Promoters (H3K4me2 only; H3K4me2/3)	Enhancers (H3K4me1; H3K4me1/2)
Inactive	H3K4me2 only	H3K4me1 only; H3K4me1/2
Poised	H3K4me2 H3K27me3	H3K4me1; H3K4me1/2 H3K27me3
Active	H3K4me2 H3K4me3, H3K27ac	H3K4me1; H3K4me1/2 H3K27ac
Bivalent	H3K4me2 H3K4me3, H3K27ac H3K27me3	H3K4me1; H3K4me1/2 H3K27ac H3K27me3

To evaluate the consistency between the regulatory domains thus defined and the genes identified by the transcript-discovery approach, TSS position of each transcript was retrieved. Genes belonging to the *de novo* assembled chromosome were ignored due to the lack of information regarding their surrounding genome sequence and arrangement. For the remaining genes, since the 5'-untranslated regions (UTRs) were the least well-defined by the transcript-discovery approach, an extended region surrounding TSS positions was investigated as compared to the usual -2.5/+2.5-kb window. Region from 10-kb upstream to 2.5-kb downstream around TSS positions were first scanned for the presence of promoter domains. The closest promoter identified was then associated with the corresponding transcript. Regions lacking of promoter domain were further investigated for the presence of enhancer domains and repression islands. The domain detected in the closest vicinity of each TSS was then linked to the corresponding transcript. Out of the 20,380 genes with at least one transcript located on the chicken genome sequence, 15,899 (78.0%) genes had their TSS associated with a promoter domain (Figure 29D). Remaining genes had their TSS in a close vicinity of an enhancer domain (1,677; 8.2%), a repression island (1,384; 6.8%) or not associated with any regulatory domain (1,420; 7%) in the investigated coverage area (Figure 29D).

DE genes are associated with bivalent promoters

In total, 13,254 promoters were detected near gene TSSs (Figure 30A). Genes were mostly associated with an active (9,589; 72.3%) or bivalent (3,355; 25.3%) chromatin state. Interestingly, the proportion of bivalent promoters was higher near the TSS of DE genes (36.9%) than for all genes (25.3%) or a set of randomly selected genes (19.5%) with similar expression levels (Figure 30).

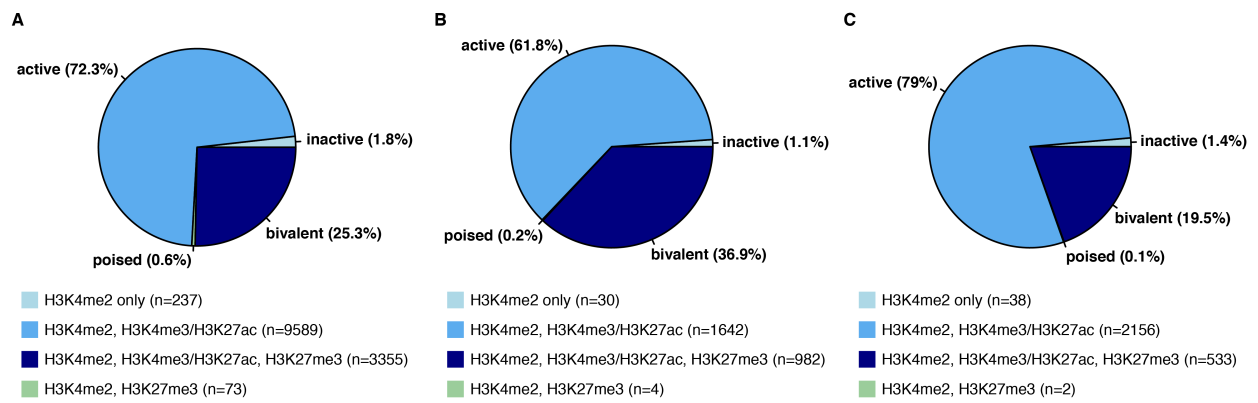


Figure 30. Promoter states associated with genes. Promoter regulatory domains identified in the vicinity of all genes (A), DE genes (B) and randomly selected genes (C). Regions surrounding TSS positions ranging from 10-kb upstream to 2.5-kb downstream were covered. Number of regions (n) for each category is indicated.

This tendency was confirmed by assessing the averaged signal for each histone modification at TSS positions. The H3K27me3 signal was higher in the TSS vicinity of DE genes (Figure 31A), as compared to the random set of genes (Figure 31B). Consistent with the increased repressive signal, the active marks H3K4me3 and H3K37ac had a weaker signal at the TSS of DE genes than that of randomly selected genes (Figures 31A,B; y-axis scale).

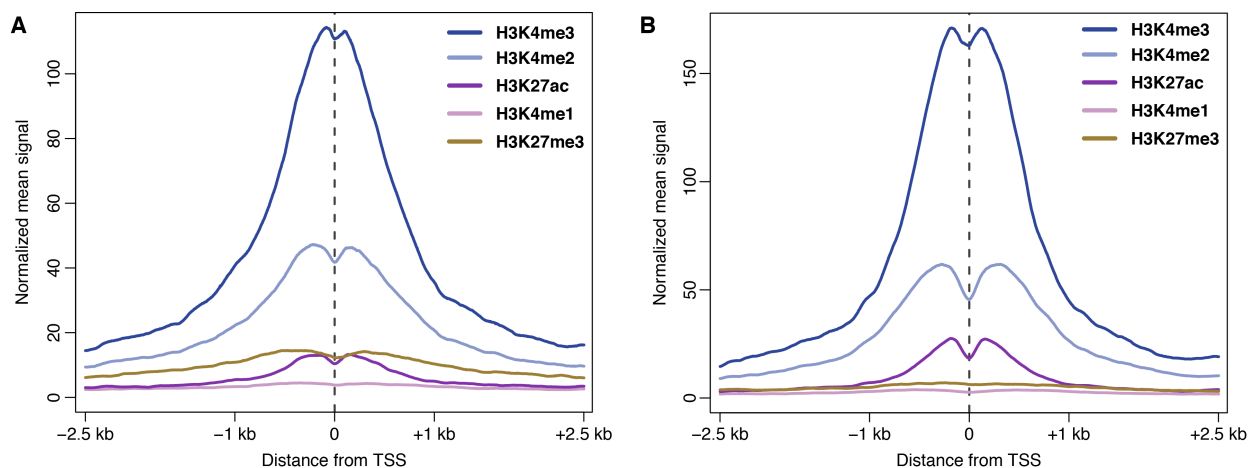


Figure 31. Histone mark coverage profiles at the TSS of differentially expressed genes. Normalized histone ChIP-seq signal surrounding the TSS of DE genes (A) and randomly selected genes (B). Signal was normalized by using the negative log₁₀ of the Poisson p-value. TSS, transcriptional start site.

To investigate this dual association of active and repressive marks, the 4,298 DE genes were filtered based on three criteria: (i) gene located on one single chromosome with a minimum size of 10 kb; (ii) gene body length of at least 1 kb; and (iii) -10/+2.5-kb regions around TSS within the chromosome borders. The resulting list was composed of 3,070 genes. The same criteria were applied to the randomly selected genes giving rise to a set of 3,080 random genes. 10-kb regions surrounding each TSS were retrieved and split into 100 intervals of 200 bp. For the genes having multiple transcripts with distinct TSS positions, the most upstream TSS was selected. Regulatory domains contained in each 200-bp interval were recovered in order to identify the most dominant domain per interval and plotted (Figure 32). Consistent with the

previous observations, DE genes were more associated with bivalent promoter domains than it would occur by chance (Figures 32A,D). The higher distribution of bivalent promoters was independent of the gene expression levels given the similarity in the range of gene expression of both gene sets (Figures 32A,D; white curves). The H3K4me3 active signal was less enriched and spread in the vicinity of DE gene TSSs (Figure 32B), as compared to the randomly selected genes (Figure 32E). In contrast, the Polycomb group-associated repressive histone modification H3K27me3 was much more abundant at the TSS of DE genes (Figures 32C,F). Bivalent promoter domains have been identified for the first time in embryonic stem cells (Bernstein et al. 2006; Mikkelsen et al. 2007). They are associated with genes that are dynamically regulated such as those involved during development and cell fate commitment. Bivalent domains maintain genes in an unstable silent state, which can be reverted for rapid gene activation and expression. Recently, bivalent domains have also been identified in adult skeletal muscle stem cells (Liu et al. 2013), suggesting a role as an epigenetic state memory for cell reprogramming.

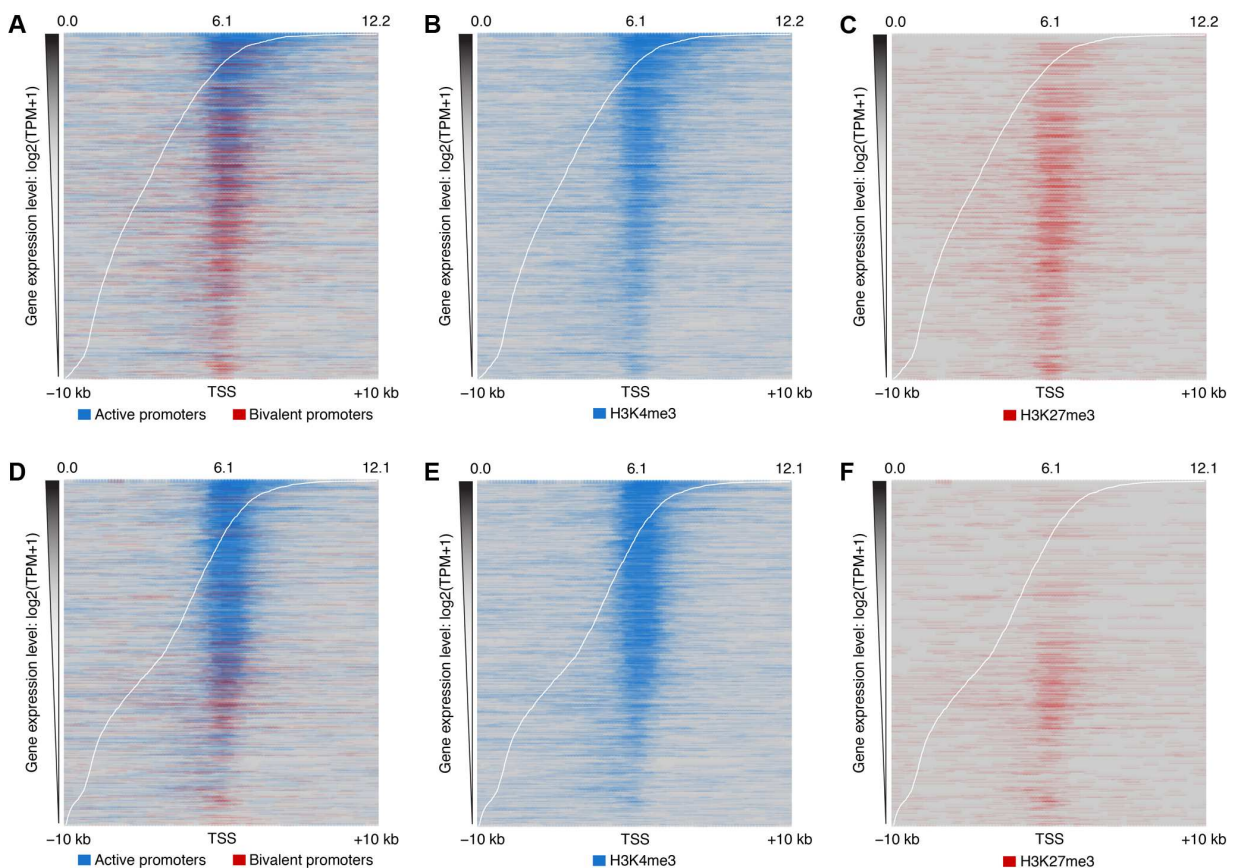


Figure 32. Promoter state surrounding the TSS of differentially expressed genes. Distribution maps of active and repressive histone marks at the TSS of 3,070 DE genes (A,B,C) and 3,080 random genes (D,E,F). (A) Distribution of active (blue) and bivalent (red) promoter domains at the TSS of DE genes. (B,C) Enrichment of H3K4me3 (B, blue) and H3K27me3 (C, red) signal at the TSS of DE genes. (D) Distribution of active (blue) and bivalent (red) promoter domains at the TSS of randomly selected genes. (E,F) Enrichment of H3K4me3 (E, blue) and H3K27me3 (F, red) signal at the TSS of randomly selected genes. Intervals with a main regulatory domain being different from active and bivalent promoter are depicted in grey. Genes were ordered according to their expression level (white curve). TPM, transcripts per million; TSS, transcriptional start site.

Nevertheless, given the multiple cell types within the chMM cultures, the bivalent promoter domains observed in the vicinity of the DE genes could simply reflect cell-type specific genes. Indeed, these genes could be activated ($H3K4me3^+/H3K27me3^-$) and therefore expressed in a subpopulation of cells, while being repressed ($H3K4me3^-/H3K27me3^+$) in another subpopulation of cells. However, the combination of both active and repressive histone marks, whether they are located at the same loci or in distinct cell populations, provides the relevant feature that these genes are dynamically regulated, as opposed to housekeeping and ubiquitous genes that would be active and expressed in most of the cells.

D. Transcription factor-binding sites

After detection of DE genes upon overexpression of OSR1, OSR2, EGR1, KLF2 and KLF4, and definition of the chromatin landscape within the chMM cultures, I next analysed which genes are directly regulated by each TF. To address this question, I performed a second series of ChIP-seq experiments targeting the individual TFs (Supplemental table S4). This approach intended to identify the TF-binding sites (TFBS) in order to distinguish the directly regulated genes from those that are indirectly regulated.

Quality control and peak calling

Due to the absence of specific and sufficient antibodies targeting each of the selected chicken TFs, a triple-FLAG (3F) tag was fused C-terminally to the CDS of each TF prior to its insertion into the RCAS-BP(A) vector. All the TFs could then be targeted by using a single antibody directed against this tag. This experimental setup allowed a high reproducibility and a direct comparison of each TFBS set, but also prevented a bias in peak detection in regards to different antibody specificities. In addition, the moderate RCAS-BP(A)-mediated TF overexpression level limits the identification of binding site artefacts that would result if the TFs would be express at a level far from their physiological level (Ibrahim et al. 2013). ChIP-seq was performed on two independent biological replicates of 5-day chMM cultures overexpressing each of the selected TFs. Prior to ChIP, retroviral infection and overexpression of the tagged TFs were assessed by immunohistochemistry (Figure 33A) and Western blot analysis (Figure 33B) against the 3F tag.

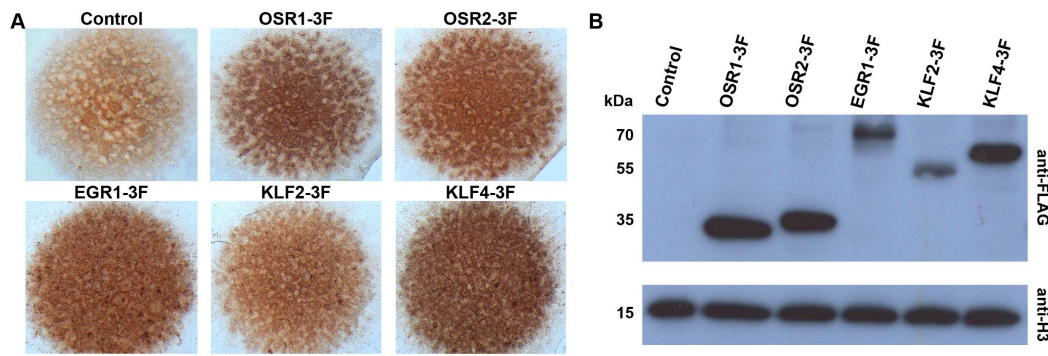


Figure 33. Detection of the 3F-tagged transcription factors within the chMM cultures. (A) Immunohistochemistry with an anti-FLAG antibody to detect the infection of retroviral RCAS-BP(A) particles carrying each of the TF CDS fused at their 3'-end with the triple-FLAG tag. (B) Western blot analysis of the 3F-tagged recombinant proteins overexpressed in chMM cultures. TFs were detected by using an anti-FLAG antibody; protein amount in each sample loaded was controlled by using an anti-H3 antibody recognizing histone proteins H3. Control corresponded to a chMM culture infected with native retroviral particles carrying no recombinant protein and no triple-FLAG CDS.

ChIP-seq coverage profiles obtained for each biological replicate and TF were normalized against the corresponding input controls by using the negative log₁₀ Poisson p-value. Similarity across all normalized signal profiles was assessed genome-wide in 500-bp non-overlapping windows by principal components analysis (PCA). Although the two first principal components accounted for 84.2% of the global variance across ChIP-seq profiles, they did not allow a sufficient separation of all signal profiles, except for EGR1 (Figure 34A). However, comparison of the second and third principal components partitioned the TF signal profiles into three distinct quadrants (Figure 34B). Both biological replicates of each TF signal profile were clustered together, as for TF-related couple profiles. EGR1 enrichment profiles were located in one quadrant, both OSR1 and OSR2 profiles were gathered together in a same quadrant, as for both KLF2 and KLF4 replicates in a third quadrant. Therefore the PCA analysis revealed two features: (i) both biological replicates of each TF are consistent; (ii) both factors of a same subgroup (OSR1 and OSR2; KLF2 and KLF4) have a similar distribution across the genome.

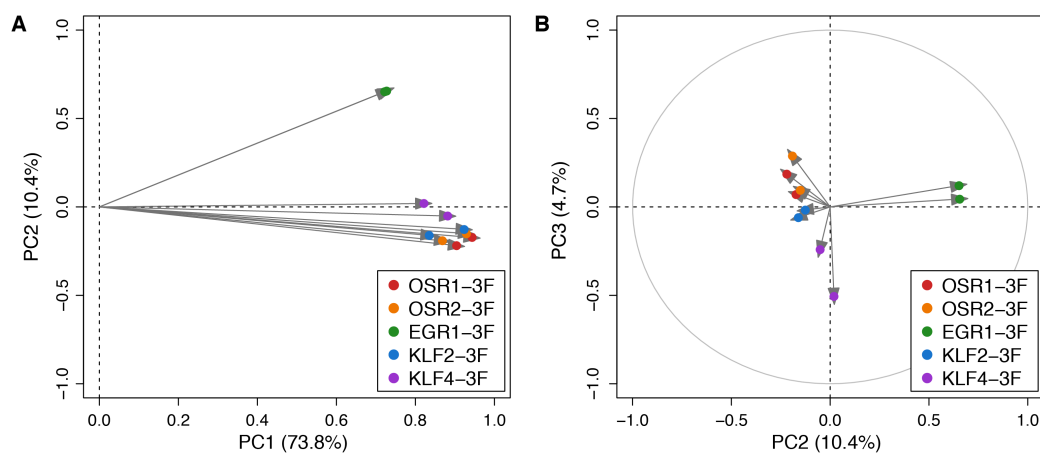


Figure 34. Principal components analysis of the transcription factor-signal profiles. PCA analysis of the normalized ChIP-seq signal profiles of all TFs and biological replicates. (A) Biplot of PC1 vs. PC2. (B) Biplot of PC2 vs. PC3. ChIP-seq signal was normalized by using the negative log₁₀ of the Poisson p-value. PC, principal component.

At the time of the first release of publicly available data in September 2012, the ENCODE consortium defined a set of good practices and metrics to assess the quality of TF-related ChIP-seq experiments (Landt et al. 2012). The guidelines thus provided intended to standardize the analysis of TF ChIP-seq data prior to biological interpretation. ChIP-seq data generated for each TF were then analysed according to the ENCODE recommendations (Supplemental table S4). Strand cross-correlation analysis was first performed to evaluate the enrichment signal in each ChIP-seq data set. This analysis is based on the accumulation of sequencing reads on each DNA strand towards the binding location of the protein of interest. It is therefore dependent of the fragment size distribution generated during DNA sonication. This analysis offers a relevant assessment of the signal-to-noise ratio to distinguish between true binding sites with a sufficient enrichment signal and binding artefacts with a weak enrichment signal. The cross-correlation analysis is accompanied of two quality metrics to evaluate the signal-to-noise ratio: (i) the normalized strand correlation (NSC; ≥ 1.05); and (ii) the relative strand correlation (RSC; ≥ 0.8). Both metrics assess the ratio of the ChIP signal over the background signal and the ratio of the ChIP peak over the phantom peak (read-length peak). Overall, the TF cross-correlation analyses passed both thresholds fixed by the ENCODE consortium (Figure 35; Supplemental figure S6; Supplemental table S4). One replicate of the KLF2 ChIP-seq was suboptimal with a NSC value below 1.05, indicative of a weak enrichment; however, its RSC value was above 0.8 (Figure 35B). In addition, one of the KLF4 ChIP-seq replicates had a moderate enrichment (NSC of 1.037) but its RSC value was at 1.263 (Supplemental figure S6G).

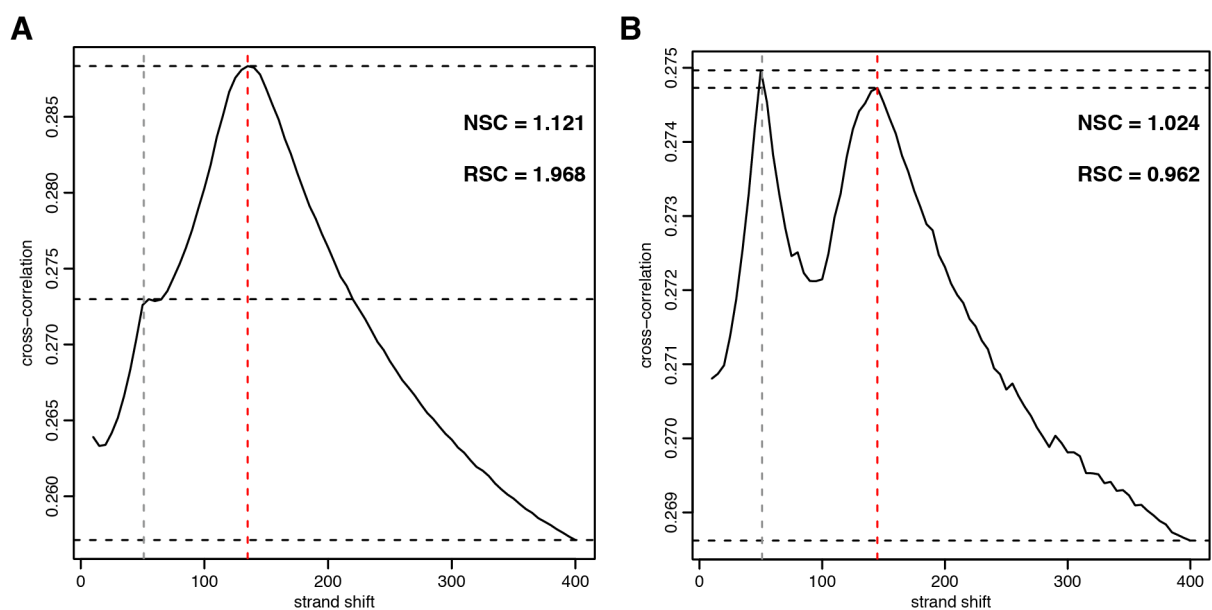


Figure 35. Cross-correlation analysis of ChIP-seq enrichment. Strand cross-correlation plot of EGR1-3F RepB (A) and KLF2-3F RepA (B) to evaluate the signal-to-noise ratio of the ChIP-seq enrichment profiles. NSC, normalized strand correlation (≥ 1.05); RSC, relative strand correlation (≥ 0.8). The first peak corresponds to the phantom peak identified at the sequencing-read length; the second peak corresponds to the ChIP peak identified at the half of the fragmentation length.

Peaks for each TF were then called by using MACS2 (Zhang et al. 2008) and low-stringency thresholds. Indeed, the ENCODE consortium recommends to perform a consistency analysis by assessing the irreproducible discovery rate (IDR) to identify TFBS rather than using a fixed threshold during peak calling (Landt et al. 2012; Li et al. 2011a). The IDR analysis assumes that the most enriched peaks and by extension the most significant ones are more likely to be consistent between biological replicates than noise peaks. IDR analysis was first performed on the peak sets detected for both replicates of each TF using a consistency threshold of 0.01 (Figure 36; Supplemental figure S7; Supplemental table S4). The highest consistency between both biological replicates was obtained for the ChIP-seq against EGR1 (Figure 36A). By contrast and in agreement with the cross-correlation analysis, both replicates of KLF2 ChIP-seq were the least consistent (Figure 36B). Due to the weaker enrichment signal obtained for the RepA, the number of significant peaks between both replicates was low. Indeed, the IDR analysis is dominated by the replicate with the weakest quality, which results in the loss of the significant peaks identified in the replicate of higher quality due to an inconsistency between both replicates.

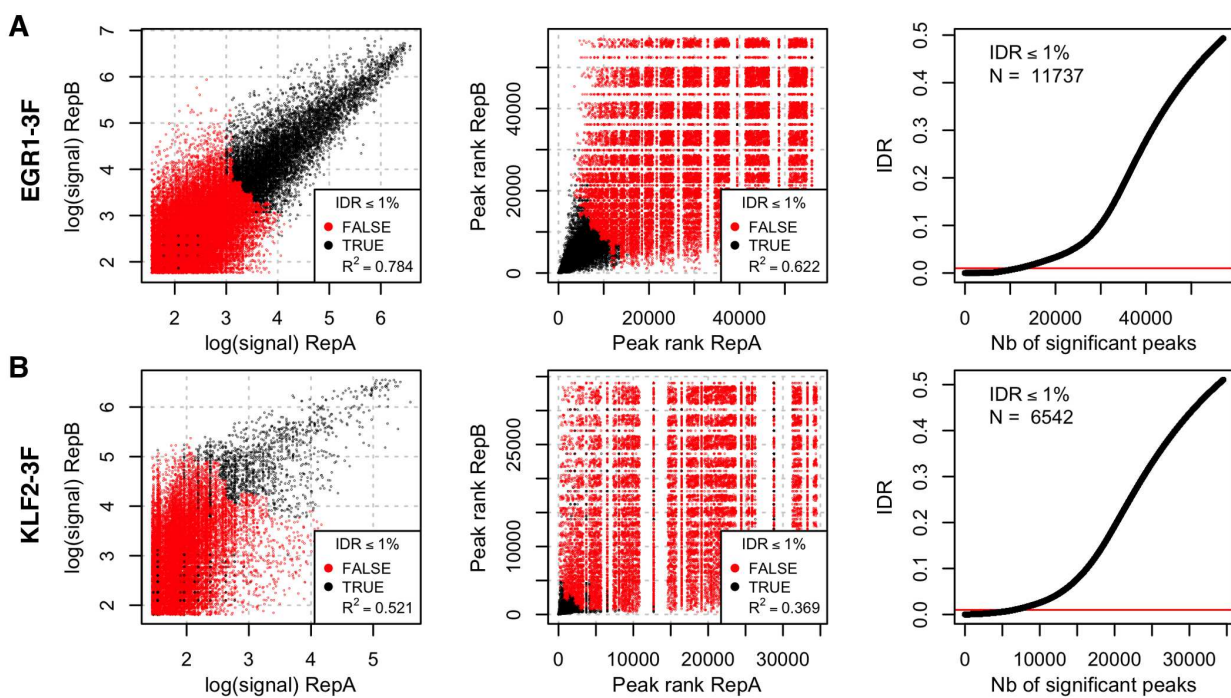


Figure 36. Replicate consistency IDR analysis. IDR analysis on the peaks identified for both biological replicates of EGR1-3F (A) and KLF2-3F (B) ChIP-seq. Peaks (N) passing an IDR threshold of 0.01 were considered as consistent between both replicates.

To obtain the final list of peaks for a given TF, the ENCODE consortium recommends to perform an IDR analysis on the pooled biological replicates. Signal profiles of both replicates for each TF ChIP-seq were merged and randomly partitioned into two pseudo-replicates. Peaks were called with low stringency thresholds and compared by IDR analysis (Figure 37; Supplemental figure S8; Supplemental table S4). The number of peaks that passed an IDR threshold of 0.01

was used to identify the final set of peaks for each TF. Peaks were called on the pooled enrichment profiles of both biological replicates for each TF and sorted according to their p-value. The number of top peaks retrieved corresponded to the number of consistent peaks identified by the pooled-replicate IDR analysis for each TF (Figure 37; Supplemental figure S8; Supplemental table S4). As a result, the numbers of binding sites selected for each TF were as follows: OSR1, 20,983; OSR2, 22,403; EGR1, 16,627; KLF2, 21,352; KLF4, 14,519. To validate this final set of peaks, especially for KLF2 that showed the lowest consistency between its both replicates, an additional quality metric was measured, the fraction of sequencing reads that are located in the final set of selected peaks (FRiP, ≥ 0.01). Each biological replicate had a FRiP value between 1.2 and 3.3% (Supplemental table S4). KLF2 RepA and KLF4 RepA had a FRiP value of 1.7 and 1.2%, respectively. Although their signal-to-noise ratio was suboptimal to moderate according to the cross-correlation analysis (Figure 35B; Supplemental figure S6G), the FRiP value above 1% indicated that their ChIP-seq enrichment profile was composed of true signal. In addition for KLF4, the ratio between replicate consistency and pooled-replicate consistency IDR analyses was below the threshold of 2 required by the ENCODE consortium (Supplemental figures S7C, S8C). By contrast, the ratio of both IDR analyses for KLF2 was over 3 (Figures 36B, 37B). Therefore, using the output of the replicate consistency IDR analysis would have strongly underestimated the number of binding sites for KLF2.

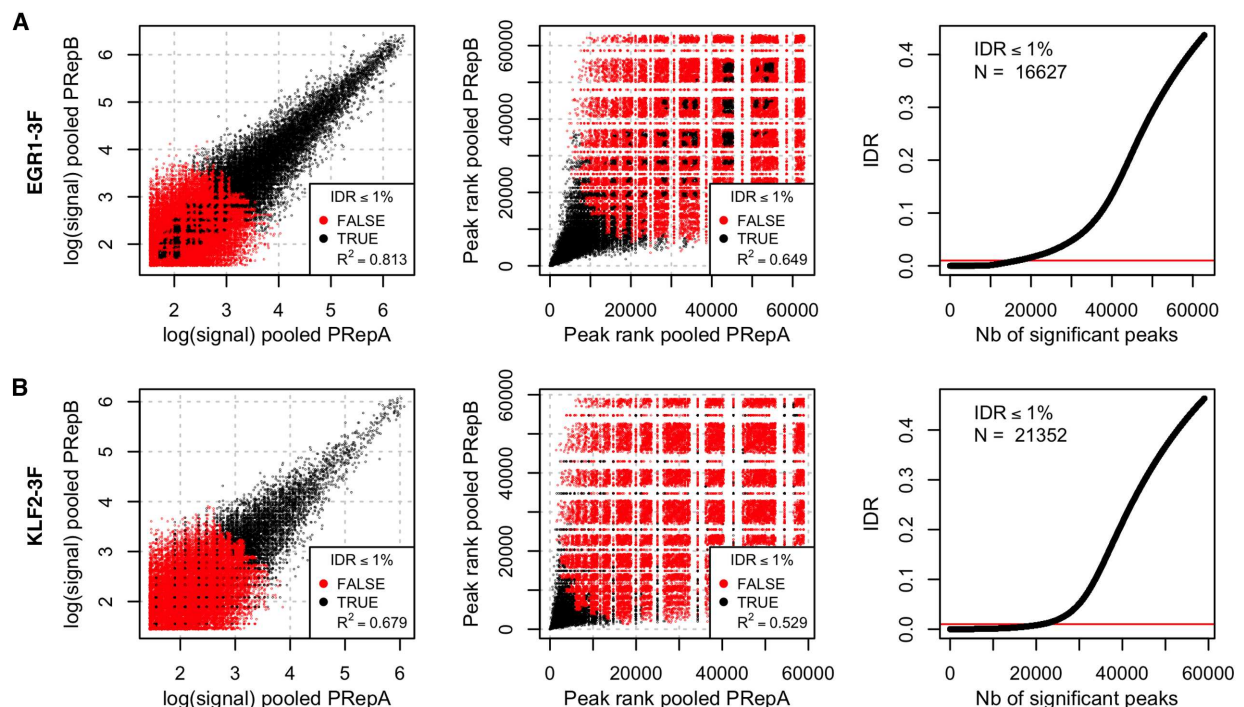


Figure 37. Pooled-replicate consistency IDR analysis. IDR analysis on the peaks identified for the pooled biological replicates of EGR1-3F (A) and KLF2-3F (B) ChIP-seq. Both biological replicates of each TF ChIP-seq were merged and randomly partitioned into two pseudo-replicates. Peaks (N) passing an IDR threshold of 0.01 were considered as consistent between both pseudo-replicates.

Functionality of TFBS

In order to assess their functionality, binding sites identified for each TF were compared with the regulatory domains and the gene annotation model (Figure 38A). The TFs depicted a variable occupancy of regulatory domains and gene features. From 10.3% to 41.6% of TFBS were found in promoters, whereas between 7.1% and 18.9% of TFBS were contained in enhancers. More than half (55.9%) of EGR1 binding sites were located within promoters and enhancers, while KLF2 had only 17.9% of occupancy in these regulatory domains. Intriguingly, between 8.6% and 14.2% of TFBS overlapped with repression islands, consisting in regions of facultative heterochromatin. Given that the chromatin landscape was solely assessed in the control chMM cultures, it is possible that these regions become accessible upon TF overexpression. In total, from 32.7% to 64.4% of TF-binding events occurred in chromatin domains. The remaining TFBS located aside from chromatin domains were mainly found in intergenic regions (20.6%-37.0%) or introns (16.3%-24.2), while no more than 6.0% of binding sites overlapped with exons (Figure 38A).

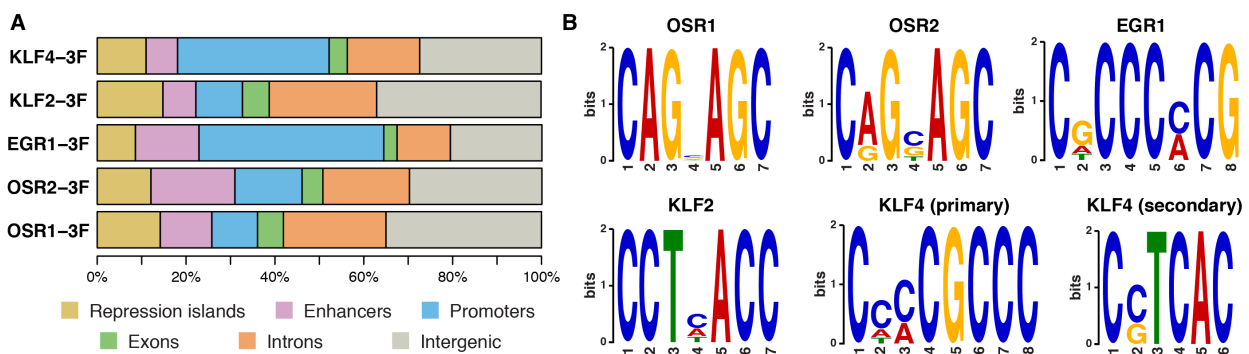


Figure 38. Transcription factor-binding events within DNA regulatory domains. (A) Proportion of TFBS in chromatin domains and gene features. (B) TF recognition motifs. Position weight matrices (PWMs) were assessed from the 1,000 most significant peaks located within promoters and enhancers for each TF (summits \pm 75 bp).

TF-binding events located in promoters and enhancers are likely to be functional and therefore to contribute to the regulation of gene expression. The 3,819-9,291 (17.9%-55.9%) binding sites for each TF identified in these domains were retrieved for further analysis. TF-binding specificity was assessed by investigating their recognition motif. *De novo* motif analysis was performed on the 1,000 most significant binding sites for each TF (Figure 38B). Recognition motifs identified for OSR1 and OSR2 are highly consistent with their known binding sites in *Drosophila* and the mouse (Badis et al. 2009; Meng et al. 2005). Both factors recognize a very similar binding motif although OSR1 motif at position 2 is more stringent than OSR2 motif (A vs. [A/G]). In agreement with previous reports, EGR1 and KLF4 binding motifs are enriched in cytosine/guanine (Badis et al. 2009; Chen et al. 2008; Jolma et al. 2013). In contrast to the other TFs, KLF2 binding motif is not known. However, its recognition motif is highly consistent with the core binding sequence of the KLF protein family (Sunadome et al. 2011). Besides for the

matrix positions 2 and 4 (C vs. [C/G]; H vs. C), the KLF2 recognition motif is similar to the secondary motif identified for KLF4. Both binding motifs identified for KLF4 may contribute to its regulatory pattern observed in the chMM cultures, considering the 767 DE genes specifically identified for KLF4 and the 1,866 DE genes shared with KLF2 (Figures 24A, 25A).

E. Characterization of candidate target genes

To validate the DE genes identified by RNA-seq and the TFBS detected by ChIP-seq, a few candidate genes were selected. All the DE genes upregulated by at least one TF were first retrieved. Then, regions spanning from 10-kb upstream of the TSS to 2.5-kb downstream of the 3'UTR were investigated for the presence of TFBS located within a promoter or enhancer regulatory domain. The resulting list of potential direct target genes was then manually parsed to identify interesting candidate genes. Six genes were thus selected and further characterized by ISH at two chick developmental stages: E5.5 (HH27-28) and E8 (HH34). By analogy with the chMM cultures, E5.5 corresponds to one day following the developmental stage at which cultures were prepared, and E8 is similar to 5-day cultures when muscle fibres are formed.

Selected candidate genes

Given their similar regulatory profiles, a common target of OSR1 and OSR2 was selected, namely *WNT11*. Both OSR factors bind at the same location within an intronic region detected as an active enhancer in the chMM cultures (Figures 39A,A'). *WNT11* is a secreted component of the non-canonical WNT planar cell polarity pathway. It is involved in cell adhesion and migration by acting on cadherins and focal adhesion proteins, as well as in the differentiation of cardiac, hematopoietic and bone marrow-derived cells (reviewed in Uysal-Onganer and Kypta 2012). In chick embryos, *WNT11* has been shown to regulate muscle fibre type and orientation during myogenesis (Anakwe et al. 2003; Gros et al. 2009). In addition, *WNT11* is expressed in joints and hypertrophic chondrocytes during chick limb development, suggesting a role in cartilage formation (Witte et al. 2009).

Additionally to *WNT11*, *ADGRG2* (also known as *GPR64*) is upregulated upon overexpression of OSR1 and OSR2 in the chMM cultures, although only a binding site for the latter was found in its vicinity (Figures 39B,B'). *ADGRG2* encodes an adhesion G protein-coupled receptor, which is expressed in epididymis and crucial to maintain male fertility (Davies et al. 2004; Kirchhoff et al. 2006). Recent studies in the mouse have suggested that *ADGRG2* is involved in adipogenesis and osteoblast differentiation since its expression is regulated by the TFs Prdm16 and Runx2, respectively (Harms et al. 2014; Teplyuk et al. 2008).

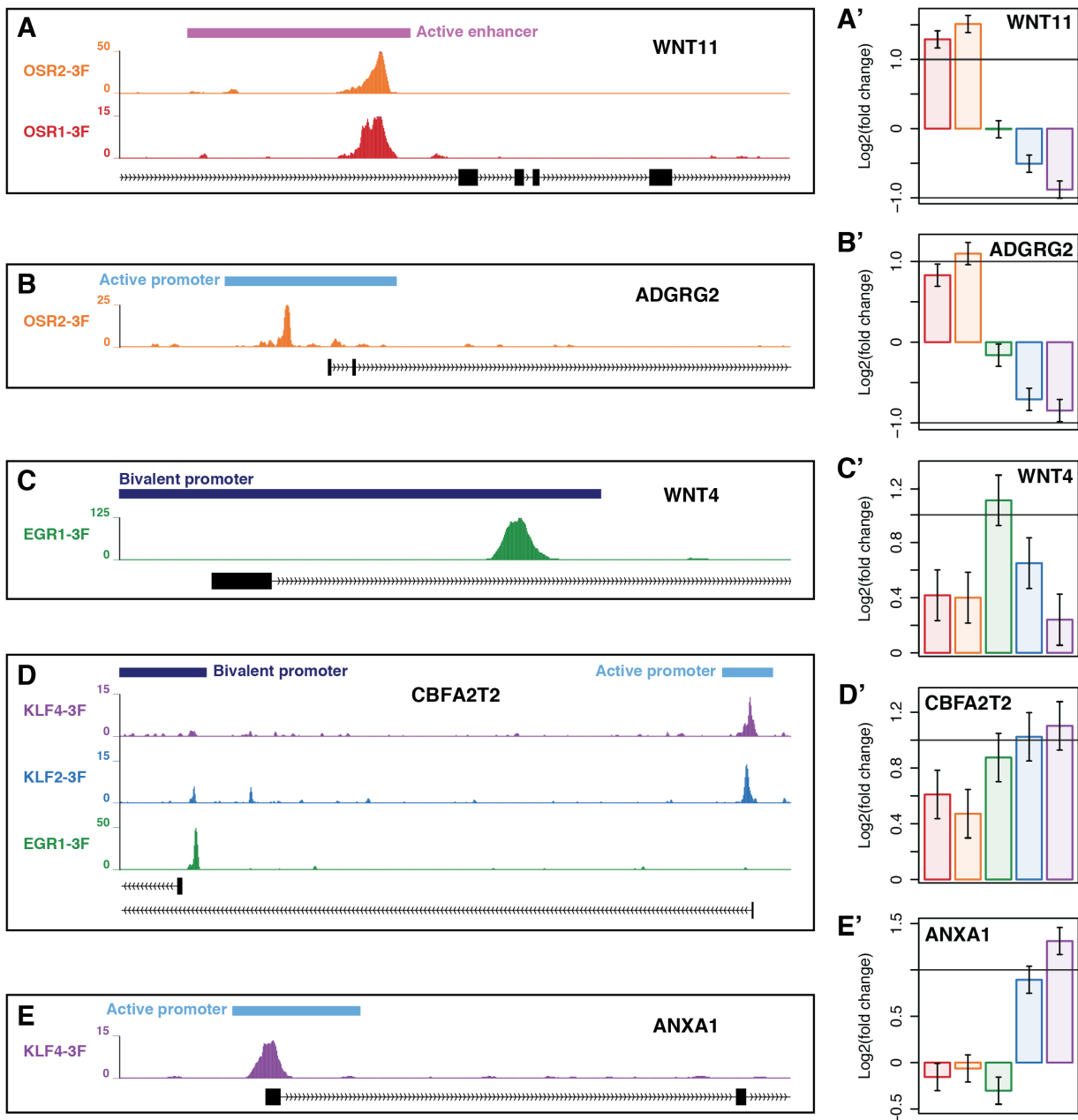


Figure 39. Candidate target genes of the transcription factors. (A-E) UCSC browser screenshots depicting the TFBS detected in regulatory domains associated with the selected candidate genes by ChIP-seq. (A'-E') Expression levels of the selected target genes within each TF-overexpressing chMM cultures determined by RNA-seq. Red, OSR1-3F; orange, OSR2-3F; green, EGR1-3F; blue, KLF2-3F; purple, KLF4-3F.

WNT4 is another member of the non-canonical WNT signalling pathway with a protein sequence highly similar to WNT11. In the chMM cultures, *WNT4* was detected as being upregulated upon EGR1 overexpression and was associated with an EGR1 binding site in its promoter region (Figures 39C,C'), which is consistent with previous findings in the uterine endometrium during mouse pregnancy (Liang et al. 2014). Besides its role as promoter of ovarian development while repressing male sexual development (reviewed in Bignon-Laubert and Chaboissier 2015), WNT4 is involved in joint development, bone formation and chondrocyte maturation (Hartmann and Tabin 2000; Loganathan et al. 2005; Später et al. 2006). In addition, *WNT4* overexpression in chick limbs upregulates *PAX7* and *MYOD*, increases muscle masses

and enhances fast-type myofibre formation, indicating a role during myogenesis (Takata et al. 2007).

CBFA2T2 (also known as MTGR1) is a transcriptional co-repressor that mediates pluripotency and germline specification (Tu et al. 2016). In the chMM cultures, *CBFA2T2* is upregulated upon EGR1, KLF2 and KLF4 overexpression, which bind to two distinct promoter regions, each being associated with an alternative isoform (Figures 39D,D'). Although CBFA2T2 has been shown to repress Notch signalling as well as to contribute to intestinal and hematopoietic cell differentiation (Ajore et al. 2012; Parang et al. 2015), its role during embryo development remains to be determined.

ANXA1 (also known as *LPCI*) encodes a calcium-dependent phospholipid-binding protein and is upregulated by both KLF factors in the chMM cultures, although only KLF4 binds to its promoter region (Figures 39E,E'). ANXA1 has been associated with an anti-inflammatory activity and seems to control adiposity (reviewed in Akasheh et al. 2013). In addition, ANXA1 is involved in cell migration, cytoskeleton organization and is thought to mediate developmental and regenerative myogenesis (reviewed in Bizzarro et al. 2012; Leikina et al. 2015).

Expression pattern of the TFs and their target genes in E5.5 chick limbs

The expression of the selected target genes was first characterized by whole-mount ISH at E5.5 (Figure 40), which corresponds to a developmental stage where most of the limb cells have reached their final localization and actively initiate their differentiation process.

WNT11 is expressed in forelimb mesenchyme along the proximodistal axis highly consistent with the expression pattern of *OSR2*, as well as the posterior *OSR1* expression domain (Figures 40A-C; white arrowheads). *ADGRG2* expression is restricted to the proximal-posterior level of the forelimb, overlapping with *OSR2* expression (Figures 40C,D). *EGR1* and *WNT4* display similar expression domains at the proximal-posterior part of the forelimb (Figures 40E,F) although *EGR1* expression extends more distally. *EGR1*, *KLF2* and *KLF4* are all expressed in limb distal mesenchyme underneath the AER, which is in agreement with the expression pattern of their direct target gene *CBFA2T2* (Figures 40E,G-I; white arrowheads). *ANXA1* and *KLF4* are both expressed in the stylopod, adjacent to the trunk and at the proximal-anterior level, as well as posteriorly in the zeugopod (Figures 40I,J).

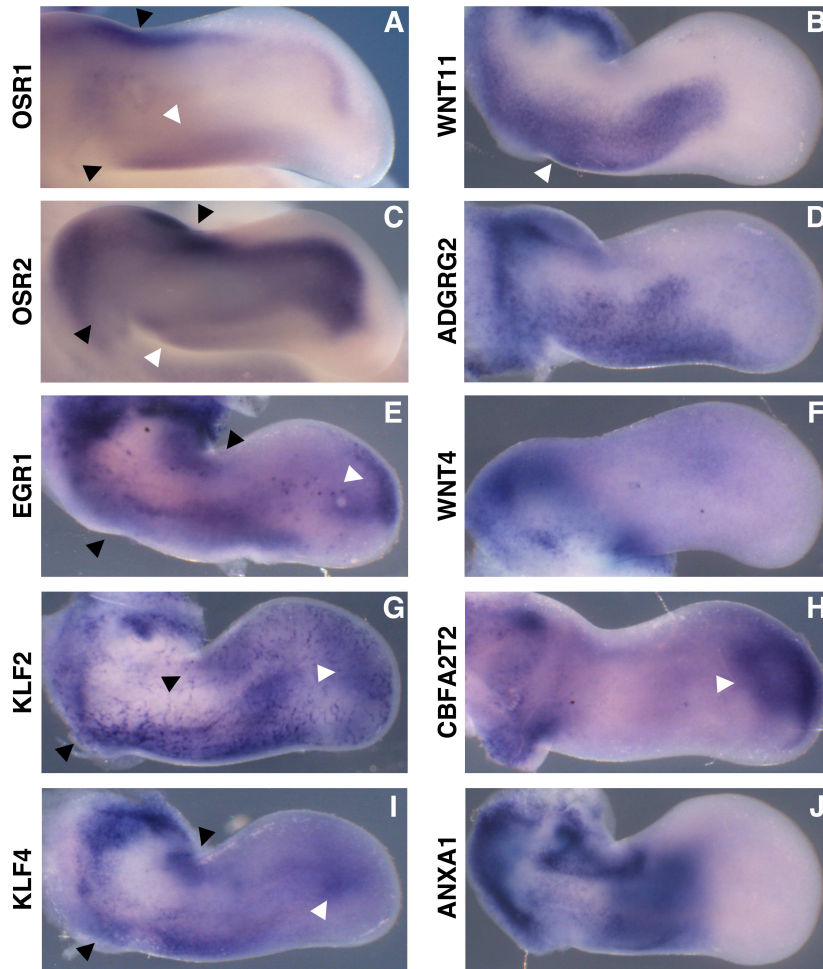


Figure 40. Expression of selected target genes in chick forelimbs at E5.5. Whole-mount ISH for *OSR1* (A), *WNT11* (B), *OSR2* (C), *ADGRG2* (D), *EGR1* (E), *WNT4* (F), *KLF2* (G), *CBFA2T2* (H), *KLF4* (I) and *ANXA1* (J) genes in forelimbs of E5.5 (HH27-28) chick embryos. Forelimbs were hybridized with the corresponding probes (purple).

Given the high consistency between the regulatory profiles of all TFs, an additional target gene was selected, *NTN1*, which is upregulated in the chMM cultures upon overexpression of each TF and for which all the TFs bind at the same location within an intronic enhancer (Figures 41A,B). *NTN1* encodes a laminin-related secreted protein, which is involved in axon guidance and biological adhesion through interactions with multiple receptors, such as integrins, during nervous system development (Serafini et al. 1996; Srinivasan et al. 2003; Yebra et al. 2003). Although this gene has been mostly investigated for its role during neural development, a recent study has highlighted that *NTN1* promotes osteoclast differentiation by inducing cytoskeletal rearrangements (Mediero et al. 2015). In E5.5 chick embryos, *NTN1* is expressed in the stylopod of both forelimbs and hindlimbs as well as in the hindlimb zeugopod, and display overlapping expression domains with those of each TF (forelimbs: Figures 40A,C,E,G,I and 41C black arrowheads; hindlimbs: Figures 41D-I white arrowheads).

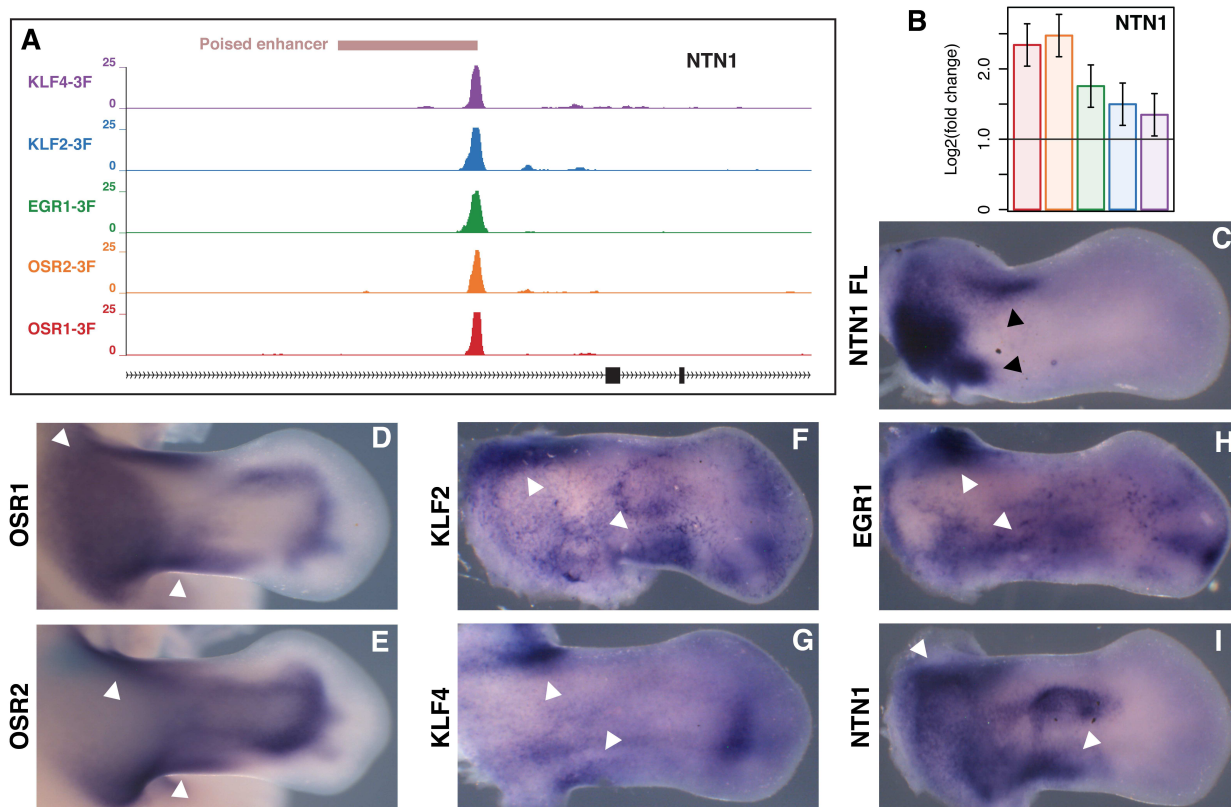


Figure 41. *NTN1* is directly regulated by all transcription factors within the chMM cultures. (A) UCSC browser screenshot depicting the TFBS detected in an intronic enhancer of *NTN1* by ChIP-seq. (B) Expression levels of *NTN1* within each TF-overexpressing chMM culture determined by RNA-seq. Red, OSR1-3F; orange, OSR2-3F; green, EGR1-3F; blue, KLF2-3F; purple, KLF4-3F. (C) Whole-mount ISH with *NTN1* probes (purple) in forelimbs of E5.5 (HH27-28) chick embryo. (D-I) Whole-mount ISH for *OSR1* (D), *OSR2* (E), *KLF2* (F), *KLF4* (G), *EGR1* (H) and *NTN1* (I) genes in hindlimbs of E5.5 (HH27-28) chick embryos. Hindlimbs were hybridized with the corresponding probes (purple). FL, forelimb.

The TFs and their target genes display overlapping expression domains in limb tissues at E8

To correlate the expression patterns of the TFs and their target genes with the components of the musculoskeletal system, ISH were performed on forelimb tissue sections at E8, when the final pattern of muscle, tendon and bone is established in limbs. *WNT11*, a common target gene of both OSR factors, is expressed anteriorly in the wrist, in a region depicting *OSR1* expression (Figures 42A,A',B,B'; black arrowheads). In addition, *WNT11* expression is detected surrounding a ventral muscle, consistent with *OSR2* expression (Figures 42B,B'',C,C'; white arrowheads). The specific target gene of OSR2, *ADGRG2*, is broadly expressed surrounding bones and muscles, overlapping with *OSR2* expression pattern but not *OSR1* (Figures 42C,C',D,D'; white arrowheads).

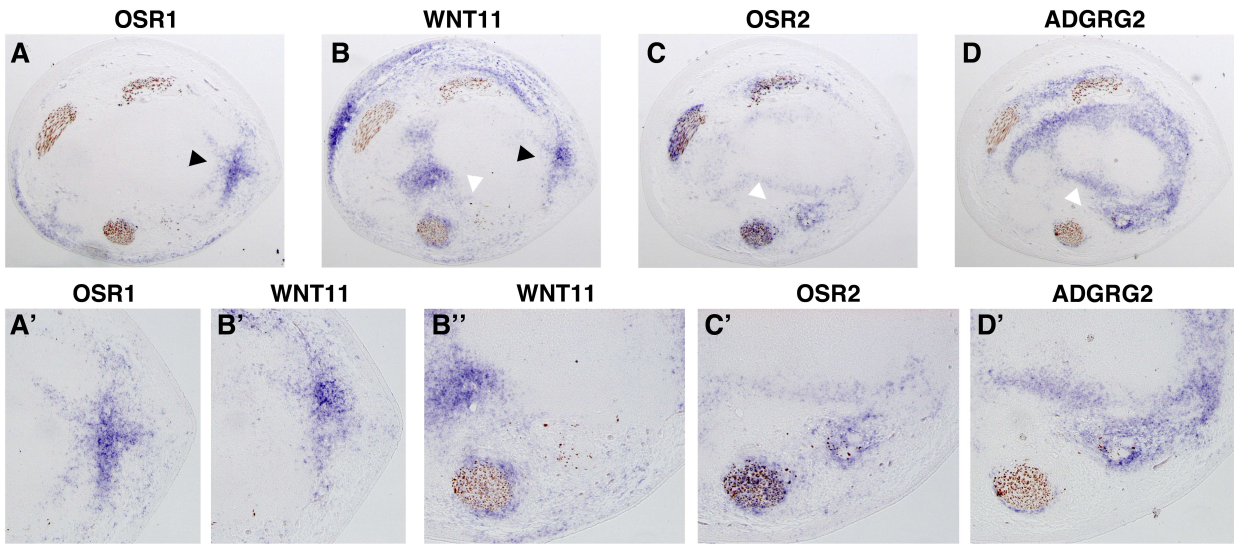


Figure 42. Endogenous expression of OSR1 and OSR2 target genes in forelimbs of E8 chick embryos. ISH for *OSR1* (A,A'), *WNT11* (B,B',B''), *OSR2* (C,C') and *ADGRG2* (D,D') genes in forelimbs of E8 (HH34) chick embryos. (A-D) Adjacent and transverse sections were hybridized with *OSR1*, *WNT11*, *OSR2* and *ADGRG2* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the wrist level. (A'-D') Magnification 10X of ventral regions marked with black and white arrowheads of sections (A-D). MF20 targets skeletal muscle myosin. Top, ventral; left, posterior.

ANXA1, which is directly upregulated by KLF4 in the chMM cultures, is expressed in regions delineating tendons similarly to KLF4 (Figures 43B,B',C,C'; black arrowheads). As expected considering the similar expression pattern between both KLF factors in limbs of E9.5 chick embryos, *ANXA1* expression also overlaps with that of *KLF2* at E8 (Figures 43A,A',C,C'; black arrowheads). *ANXA1* was detected as being upregulated in the chMM cultures overexpressing KLF2 as well (Figures 39E'). However, given the absence of a binding site in the vicinity of *ANXA1* gene body, *ANXA1* is more likely to be an indirect target gene of KLF2.

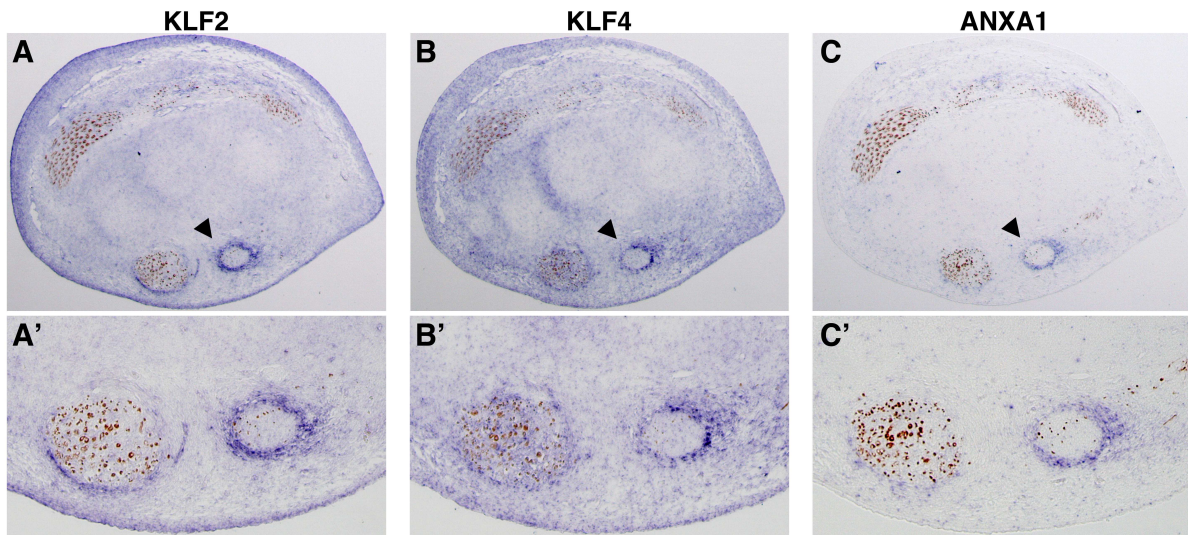


Figure 43. Endogenous expression of ANXA1, a target gene of KLF4, in forelimbs of E8 chick embryos. ISH for *KLF2* (A,A'), *KLF4* (B,B') and *ANXA1* (C,C') genes in forelimbs of E8 (HH34) chick embryos. (A-C) Adjacent and transverse sections were hybridized with *KLF2*, *KLF4* and *ANXA1* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the wrist level. (A'-C') Magnification 10X of ventral regions marked with black arrowheads of sections (A-C). MF20 targets skeletal muscle myosin. Top, dorsal; left, posterior.

Development of the musculoskeletal system requires spatial, temporal, molecular and mechanical interactions between its components for proper morphogenesis and functionality. Throughout limb development, muscle and connective tissue cells are intimately associated suggesting that they differentiate in a coordinate manner and interact with each other. What are these interactions and how do they take place remain two fundamental unanswered questions that need to be addressed. In contrast to skeletal muscle and skeleton elements, connective tissues remain poorly characterized although emerging evidence have highlighted that they are essential throughout limb embryogenesis. The PhD thesis presented here intended to decipher the molecular mechanisms taking place during connective tissue differentiation that could influence skeletal muscle development. By analysing several layers of gene regulation, such as gene expression profiles, chromatin landscape and transcription factor (TF) occupancy, I established the regulatory patterns of connective tissue-associated TFs in chick limb cells.

A. Designing a three-level transcriptional network

Development of high-throughput sequencing technologies over the last past two decades has greatly contributed to decipher the complexity underlying gene expression in eukaryotic cells. Nowadays, various strategies can be applied to investigate genome-widely a particular gene regulation process depending on the biological question to address. Large-scale technologies offer remarkable advantage by providing the global repertoire of e.g. gene expression levels or TF occupancy, which are compensated by the complex computational analysis that they require.

Improving the gene annotation of the chicken genome

Since its first release in 2004 and despite significant improvements over the last past decade, the *Gallus gallus* genome is presently not complete and highly fragmented (Hillier et al. 2004). One of the main challenges is raised with the intrinsic fragmentation of the chicken karyotype and its 28 microchromosomes. Since January 2016, the fifth version of the chicken genome has been released. As compared to the fourth version, this release includes ~200 more Mb and ~5,000 additional annotated genes but retains 10% of its sequence as unassigned. Improvement of the chicken genome is an on-going project and a sixth version will be released in a near future (Wesley Warren, USA; personal communication). At the time of this thesis, this fifth version was not available. Therefore, all the analyses have been carried out on the fourth version of the *Gallus gallus* genome released in 2011 (galGal4).

In contrast to ChIP-seq experiments, the wide range of different applications renders the establishment of standard procedures to analysis RNA-seq data difficult. From the experimental design to the detection of differentially expressed (DE) genes, generation and computational processing of gene expression profiles differ depending on the biological questions to address. Read mapping is one of the critical steps that will further influence sample normalization, gene quantification and the identification of relevant target genes. Besides the choice of the appropriate mapper, the quality of the reference genome or transcriptome along which RNA-seq reads are aligned is of pivotal importance since quantification is computed according to read coverage. In the absence of an appropriate reference genome, *de novo* assembly relies on read overlaps to build contigs as long as possible. This strategy has been widely used to generate new genome sequences but it can also be applied on RNA-seq data to build transcript sequences. Prior to the analysis of RNA-seq data *per se*, sequencing reads obtained from both chicken micromass (chMM) culture replicates infected with RCAS-BP(A) retroviruses carrying no recombinant protein were used to improve the gene annotation model. RNA-seq libraries were prepared to maintain strand specificity and to sequence both extremities of transcript fragments (paired-end sequencing). A first approach was led by using the galGal4 reference genome as guide. Sequencing reads were mapped along the genome followed by a transcript-discovery strategy that computed read coverage and exon-intron junctions from gapped alignments and distance between both reads of each pair. This strategy intended to correct existing annotated genes and to identify missing genes from the galGal4 reference annotation. A second transcript-discovery approach was also carried out independently of the reference genome. Sequencing reads were *de novo* assembled to reconstruct full-length transcripts. Transcripts thus generated were then compared with the genes identified by the first approach to correct gene fragmentation. Indeed, 2,771 (15.6%) genes were identified among the ~15,000 unplaced contigs associated with the galGal4 genome, including 1,031 (4.8%) genes that are located on one main chromosome and additional contigs due to the presence of numerous gaps of unknown sequence (Figure 20A). Although the new version of the chicken genome succeeded to assign most of these unplaced contigs to a main chromosome, their exact localization remains unclear. The resulting gene annotation model was composed of 21,347 genes accounting for 4,029 additional genes as compared to the UCSC and Ensembl reference annotations associated with the galGal4 genome version (Figure 20B). Therefore, combination of both transcript-discovery approaches enabled to correct existing genes, to annotate new genes and to gather gene parts that were split together. Most importantly, the designed gene annotation model increased the assignment rate of sequencing reads by 20-22% as compared to when using both galGal4 reference annotations, thus contributing to a more accurate estimation of transcript abundances (Figure 20C).

The recent sequencing of a wide range of avian species have provided extensive insights into evolutionary and adaptive traits within birds (Zhang et al. 2014). DNA conservation of protein-coding genes greatly facilitated the annotation of the 21,347 genes identified by the dual transcript-discovery approach. By combining DNA sequence comparison against avian genes with protein sequence comparison against mammal species and protein domain prediction, 14,847 (69.5%) genes could be assigned and 672 (3.1%) putative protein-coding genes could be identified (Figure 20B). The 5,828 remaining genes were divided between uncharacterized proteins and non-coding RNAs (ncRNAs) based on the detection of an open reading frame (ORF) or not, respectively (Figure 20B). However, genes encoding uncharacterized proteins could be also potentially non-coding since none of the protein domains investigated were detected within their putative ORF. Nevertheless, ncRNAs remain challenging to annotate according to a recent study comparing an extensive repertoire of long multi-exonic ncRNAs across 11 tetrapods separated by up to 370 million years, ranging from *Xenopus* to human and including chicken and platypus (Necsulea et al. 2014). Besides their overall weak conservation as compared to protein-coding sequences, long ncRNAs (lncRNAs) depict high tissue specificity and rapidly diverge through evolution, which render their annotation difficult by comparing with other species.

Distinguishing true ChIP-seq signal

Assessing gene expression profiles in the chMM cultures upon overexpression of each TF highlighted overall regulatory processes influenced by the TFs and connective tissue cells, but it does not provide the exact molecular mechanisms driven by the TFs. Indeed, changes in gene expression levels arose not only from TF occupancy, but also from downstream regulatory events following direct binding of each selected TF. Determination of TF-binding sites (TFBS) is therefore critical to distinguish between directly and indirectly targeted genes. Development of the ChIP-seq technology greatly enhances the exploration of TF occupancy by enabling a genome-wide resolution of protein-DNA interactions (Barski et al. 2007; Johnson et al. 2007; Mikkelsen et al. 2007; Robertson et al. 2007). Since then, ChIP-seq has been widely used to become a standard technique to assess various DNA-binding proteins such as TFs, polymerases, histones, chaperones and chromatin-modifying enzymes. Therefore, efforts have been made to standardize ChIP-seq data analysis prior to biological interpretation. Although the analysis of histone occupancy profiles remain challenging due to their broad domains and diffused signal, the ENCODE consortium established a list of quality metrics and methods to validate TF ChIP-seq data (Landt et al. 2012). One of the main fundamental questions regarding ChIP-seq data is related to the binding signal: do the peaks visualized after ChIP-seq read mapping correspond to

true signal resulting from TF binding, or background signal originating from unspecific binding or amplification artefact? Evaluation of signal-to-noise ratio is therefore a first prerequisite to validate ChIP-seq data. When considering a protein bound at a certain location, DNA fragmentation by sonication occurs on each side while the binding site remains protected. DNA fragments encompassing this location are purified during the ChIP procedure and one can assume that sequencing arises from both ends. Therefore, the center of the binding site is marked by an accumulation of sequencing reads originating from both forward and reverse strands. By contrast, noise signal lacks this shift stranded read density. Cross-correlation analysis can thus measure the ratio between true and noise signal and determine if the enrichment is sufficient enough to distinguish binding peaks from potential artefacts.

A second issue related to ChIP-seq experiments is to ensure the reproducibility between binding profiles of biological replicates. Reproducibility can be assessed by measuring the Pearson correlation coefficient of read coverage along the genome (Bardet et al. 2012). Nevertheless, Pearson correlation can be dominated by regions located near centromeres, telomeres and repeats, that usually depict strong ChIP-seq signal (Bailey et al. 2013). Such regions have been blacklisted by the ENCODE consortium and the 1000 Genomes project, which recommend to remove them prior to data analysis. However, these regions remain undefined for the chicken genome. Another approach to measure ChIP-seq reproducibility is the irreproducible discovery rate (IDR) analysis (Landt et al. 2012). ChIP-seq peaks are characterized by their significance level based on the read coverage, signal-to-noise ratio and p-value or q-value. When comparing two sets of peaks for the same TF, the most significant peaks are more likely to be consistent between both replicates, whereas noise peaks are more prone to display a weaker consistency. Such approach appears to be more powerful to identify TFBS as compared to significance metrics such as false-discovery rate (FDR), p-value and q-value, since it takes into account both replicates instead of treating them independently (Landt et al. 2012). Cross-correlation and IDR analyses were used to assess the quality of the TF occupancy profiles generated from the chMM cultures. Both replicates of OSR1, OSR2 and EGR1 ChIP-seq profiles passed all the ENCODE thresholds, whereas one replicate for KLF2 and KLF4 displayed a suboptimal signal-to-noise ratio, albeit sufficient for further analysis (Figures 35, 36, 37; Supplemental figures S6, S7, S8; Supplemental table S4). The number of peaks identified for both KLF factors may be therefore underestimated due to the loss of low-affinity binding sites, which display a weaker signal as compared to high-affinity binding regions. Due to the broad domains obtained with ChIP-seq on histone proteins, cross-correlation and IDR analyses cannot be applied and quality metrics still remain to be developed. Histone domains were identified according to a significance threshold fixed by the peak caller and following the ENCODE and Roadmap Epigenomics consortiums'

recommendations (Kellis et al. 2014; Roadmap Epigenomics Consortium et al. 2015). Consistency between histone domains of both replicates was evaluated independently of their significance.

Addressing the functionality of transcription factor-binding events

Despite the advances in the development of genome-wide strategies and the increasing knowledge of gene regulation processes, defining the functionality of TF-binding events remains challenging. Between 14,519 and 22,403 binding sites were identified per TF in the chMM cultures, which contrasts with the 1,369-2,907 DE genes detected in the same cultures. Although some binding events may correspond to false-positives due to the expression of the TFs in cells that do not normally express them, the use of RCAS-BP(A)-mediated TF overexpression limits such artefacts. Indeed, the absolute overexpression level from retroviral particles has been estimated between 3 and 5 folds per cell (Ibrahim et al. 2013). Consistent with a previous study correlating TFBS and DE genes upon TF knockdown, most of the TF-binding events identified in the chMM cultures would be non-functional, in terms of regulating gene expression (Cusanovich et al. 2014). While studies have suggested that TFBS simply reflect chromatin accessibility (John et al. 2011; Li et al. 2011b), other reports have proposed that TFBS may contribute to chromatin remodelling and nucleosome positioning to influence gene expression at later stages or in response to signals (Cao et al. 2010a; Buck and Lieb 2006; Iwafuchi-Doi et al. 2016). On the other hand, conventional ChIP methods measure steady-state TF occupancy, whereas competitive ChIP experiments in the yeast highlighted that TFs depict a high binding turnover (Lickwar et al. 2012). Therefore, assessing binding kinetics rather than Boolean occupancy may be a more relevant measure of TF-binding functionality (Chen et al. 2014). In addition, TF specificity often relies on the cooperativity with other TFs and co-factors. For example, preventing Gata1 to interact with its non-DNA-binding partner Fog1 alters its occupancy profile while enhancing its binding at other locations (Chlon et al. 2012). Clustering of multiple TF-binding events at high-occupancy target (HOT) regions seem to be a better predictor of functional *cis*-regulatory modules (Foley and Sidow 2013; Heidari et al. 2014; Junion et al. 2012; Kvon et al. 2012). Therefore, overexpression of a single TF in the chMM cultures may reflect its binding sites that contain its recognition motif, but locations that require cooperative recruitment may be missed.

Several approaches have been developed to more accurately distinguish between functional and non-functional TF-binding events from an initial ChIP-seq experiment. While some studies have investigated evolutionary constraints of TF occupancy between distant species or organisms (e.g. Schmidt et al. 2010; Ballester et al. 2014), others have focused on co-occupancy by combining

multiple ChIP-seq data sets of functionally related TFs (e.g. Junion et al. 2012), or by associating TF binding with chromatin states and domains (e.g. Heintzman et al. 2007; Ernst and Kellis 2010, 2013). The latter approach was chosen for this thesis by assessing different histone tail post-translational modifications associated with promoters and enhancers, as well as transcriptional activity and repression (Supplemental figure S5). In addition to link TFBS and DNA regulatory domains, this approach provides valuable insights into chromatin states complementary to gene expression levels. 20,427 promoters and 55,597 enhancers were thus identified in the chMM cultures (Figures 29A,B). Assignment of regulatory domains with their target genes remains a difficult task and has often relied on the closest proximity between both elements. This strategy is rather relevant for promoter regions, which are generally located in the vicinity of gene transcriptional start sites (TSSs). Indeed, 15,899 (78.0%) genes had their TSS associated with a promoter, whereas 1,677 (8.2%) gene TSSs were closely related to an enhancer (Figure 29D). In contrast to promoters, enhancers can be located distally to their target genes. Therefore, an enhancer may not necessarily regulate the closest gene but another gene located several hundreds of kilobases upstream or downstream. Additional genome-wide techniques have been developed in regards to identify chromatin interactions by chromosome conformation captures (Fullwood et al. 2009; Lieberman-Aiden et al. 2009). Continuous improvements of such techniques have greatly enhanced their resolution. It is now feasible to visualize chromatin contacts spanning within 1 kb (Rao et al. 2014). Comparison of TFBS and regulatory domains identified between 17.9% and 55.9% of TF occupancy in promoter and enhancer regions (Figure 38A). Assessing the contribution of TF-binding events to regulate gene expression requires additional functional assays such as luciferase reporter assay or electrophoretic mobility shift assay (EMSA). However, *in vitro* methods do not obviously reflect the context-dependent contribution of a TF. *In vivo* assays are therefore more reliable to validate TF functionality on gene expression. Given the absence of good chicken antibodies available for these TFs, I chose to compare the expression domains of a few interesting candidate genes with their corresponding TF in chick limbs by *in situ* hybridization experiments. Although this approach does not prove the direct regulation of the TF with its target gene, it provides at least preliminary insights into the relationship that may exist between them in regards to their endogenous expression.

B. Effects of the transcription factors on cell differentiation

The five zinc-finger TFs, OSR1, OSR2, EGR1, KLF2 and KLF4, were selected based on their expression within different compartments of the musculoskeletal system and their potential involvement in mediating connective tissue differentiation. To assess their influence on cell

differentiation, I chose to combine chMM cultures of limb mesenchymal cells with retroviral-mediated overexpression of each TF. The chMM culture system offers a versatile experimental model to investigate differentiation processes occurring in the developing limb. Limb mesenchymal cells of E4.5 chick embryos contain different cell types that are mostly undifferentiated at this stage of development. The chMM culture model has been widely used to study chondrogenesis (e.g. DeLise et al. 2000; Stricker et al. 2012), but it seems also adapted to assess myogenesis since limb mesenchymal cells differentiate spontaneously into muscle cells in chMM cultures (Duprez et al. 1996a). Although adipogenesis and tenogenesis do not occur spontaneously in chMM cultures, levels of genes associated with these differentiation processes were altered depending on the overexpressed TF. In contrast to other cell culture systems, limb mesenchymal cells in chMM cultures maintain interactions between cell types and undergo actively various differentiation processes. It is therefore a good experimental alternative to investigate the influence of TFs on connective tissue cell differentiation, while maintaining their interactions with myogenic cells. I believe that these chMM cultures reflect cell interactions and differentiation processes that occur during the formation of the musculoskeletal system during limb development.

Odd-skipped related 1 and 2 (OSR1, OSR2)

During chick limb development, *OSR1* and *OSR2* are expressed in muscle connective tissue (MCT), surrounding muscle masses at E4.5 and myofibres at E9.5 (Figures 14G,H, 15J,K,N-O'; Supplemental figures S2G,H, S3K,L), which is consistent with previous observations (Stricker et al. 2006, 2012). Interestingly, some muscles were associated with expression of both *OSR1* and *OSR2*, whereas other muscles were accompanied with expression domains of either *OSR1* or *OSR2* only, or none of them. It indicates that both OSR factors are expressed in distinct subpopulations of limb MCT. This observation is consistent with a recent study in mice showing that only a fraction of *Osr1*⁺ cells express the MCT marker *Tcf4* (Vallecillo García et al., in revision). In agreement with previous observations (Stricker et al. 2012), Overexpression of *OSR1* or *OSR2* in chick forelimbs induced a size reduction of skeletal elements (Figures 16B,C). Accordingly in chMM cultures, the formation of cartilage nodules and the expression of cartilage-associated genes were decreased (Figures 17B',C',G, 18B, 25). Similarly, limb *Osr1*⁺ cells isolated from mouse embryos fail to differentiate into chondrocytes when cultured under chondrogenic conditions (Vallecillo García et al., in revision). Although no significant change was observed in *OSR1*- and *OSR2*-overexpressing chMM cultures, homozygous deletion of *Osr1* induces an upregulation of tendon-associated genes such as *Scx* and *Tnmd* in *Osr1*⁺ cells indicating that *Osr1* prevent tendon differentiation (Vallecillo García et al., in revision).

Altogether, these results indicate that OSR1 and OSR2 direct the differentiation of mesenchymal cells into connective tissue by repressing cartilage and tendon differentiation. Consistently, MCT-associated genes *COL3A1*, *COL6A5* and *FNI* (fibronectin 1) were significantly upregulated upon overexpression of OSR1 and OSR2 in chMM cultures (Supplemental figure S9A).

Although *Osr1* is not expressed in myogenic cells in chick and mouse embryos, *Osr1*-deficient mice display muscle patterning defects (Stricker et al. 2012; Vallecillo García et al., in revision). Consistently, retroviral overexpression of OSR1 or OSR2 led to muscle size reduction in chick limbs and decreased myotube formation in chMM cultures (Figures 16B',C', 17B'',C'',H). Surprisingly, genes associated with muscle differentiation were significantly more abundant in OSR1- and OSR2-overexpressing chMM cultures as compared to control cultures (Figure 18C, 25). 87.3% (268/307) of these genes were found to be expressed (transcript abundances ≥ 1) as well in *Osr1*⁺ cells isolated by FACS from E13.5 mouse limbs (Vallecillo García et al., in revision). Developmental limb *Osr1*⁺ cells give rise to fibro-adipogenic (FAP) cells (Vallecillo García et al., in revision). In addition, *Osr1* and *Osr2* are expressed in a subpopulation of FAPs, the SCA1⁺/PW1⁺ interstitial cells (Stumm 2016). At early stages of disease, FAPs activate a pro-myogenic program by chromatin remodelling in dystrophic muscles, which would contribute to enhance the regenerative capacity of the muscle satellite cells (Saccone et al. 2014). Therefore, it is possible that both OSR factors activate a similar pro-myogenic program or maintain the expression of muscle-associated genes within the chMM cultures, while preventing the cells to differentiate toward a muscle lineage.

Early growth response 1 (EGR1)

EGR1 is expressed in tendons close to muscle attachments in limbs of E7 and E9.5 chick embryos (Lejard et al. 2011; Figures 15B,I,M,M'; Supplemental figures S3B,B',F,F',J). Although it is not clear whether EGR1 regulates the formation of myotendinous junctions, *Egr1*-deficient mice display disorganized and less robust tendon collagen fibrils (Guerquin et al. 2013). Retroviral overexpression of EGR1 in chick forelimbs did not reveal an obvious defect in muscle and skeleton organization (Figures 16D,D'). Although it increased the production of chondrogenic matrix, EGR1 overexpression in chMM cultures did not alter the expression of cartilage- or muscle-related genes (Figures 17D',G, 18B,C). By contrast, the *SCX* and *COL1A1* tendon markers were significantly upregulated upon EGR1 overexpression (Supplemental figure S9B). In addition, EGR1 negatively regulated genes involved in adipogenesis in chMM cultures (Figure 18E). Altogether, these results are consistent with previous observations in mice. EGR1 is sufficient to direct tendon differentiation in murine mesenchymal stem cells by promoting the

expression of tendon-associated genes and collagens (Guerquin et al. 2013). In addition, ectopic expression of *Egr1* in these cells prevents their differentiation into adipocytes while they retain their ability to undergo chondrogenesis when cultured under adipogenic or chondrogenic conditions, respectively (Guerquin et al. 2013). Likewise, upregulation of *Egr1* in murine pre-adipocytes inhibits their differentiation into adipocytes, and conversely, downregulation of *Egr1* enhances their differentiation (Boyle et al. 2009). On the other hand, overexpression of EGR1 in chMM cultures increased the expression of bone-associated genes (Figure 18A), whereas activation of *Egr1* in mouse mesenchymal stem cells prevents their differentiation into osteocytes when cultured under osteogenic conditions (Guerquin et al. 2013).

Krüppel-like factors 2 and 4 (KLF2, KLF4)

KLF2 and *KLF4* are expressed in tissues delineating tendons of the knee and the wrist of E8 and E9.5 chick embryos (Figures 15C,D,F-G', 43A-B'; Supplemental figures S3C-D',G-H'). Similarly to EGR1, overexpression of KLF2 and KLF4 in chick forelimbs did not lead to a skeleton defect but upregulated the expression of bone-related genes in chMM cultures (Figures 16E,F, 18A). Conversely to EGR1, both KLF factors reduced the production of chondrogenic matrix when overexpressed in chMM cultures, while not affecting the expression of cartilage-related genes (Figures 17E',F',G, 18B). In agreement with previous observations in the mouse (Sen Banerjee et al. 2003; Wu et al. 2005), overexpression of KLF2 in chMM cultures repressed adipogenesis-related gene expression (Figure 18E). Although KLF2 and KLF4 overexpression in chMM cultures did not significantly affect the global expression of tendon-associated genes, they were sufficient to upregulate the tendon markers *SCX* and *COL1A1* (Supplemental figure S9B). Therefore, it will be interesting to perform *in situ* hybridization on tissue sections against tendon markers such as *SCX* and *TNMD* to investigate tendon formation upon overexpression of KLF2 or KLF4 in chick limbs.

Overexpression of KLF2 and KLF4 in chick forelimbs did not reveal an obvious defect in muscle patterning (Figures 16E',F'). By contrast, overexpression of KLF2 and KLF4 in chMM cultures displayed a significant, albeit moderate, reduction of culture area covered by myotubes as compared to control cultures (Figures 17E'',F'',H). In the mouse, *Klf2* and *Klf4* have been highlighted as being involved in myoblast fusion but not in their differentiation (Sunadome et al. 2011). Indeed, misexpression of either *Klf2* or *Klf4* does not alter the expression of muscle differentiation genes. In addition, ectopic expression of each factor is not sufficient alone to trigger the fusion of mesenchymal stem cells, but requires the concomitant induction of *Myod*. Consistently, genes associated with muscle differentiation were overall not significantly affected by overexpression of both KLF factors (Figure 18C).

In summary, based on expression pattern in chick limbs, histological staining and gene expression profiles upon retroviral overexpression, it appears that the TFs differentially influence cell differentiation processes in chMM cultures (Figure 44). OSR1 and OSR2 drive the formation of MCT at the expense of cartilage and muscle differentiation. EGR1 promotes the expression of the tendon marker *SCX*, while contributing positively to chondrogenic matrix production and the expression of bone-related genes. KLF2 and KLF4 upregulate the expression of *SCX* and bone-associated genes but represses the production of chondrogenic matrix and myotube formation.

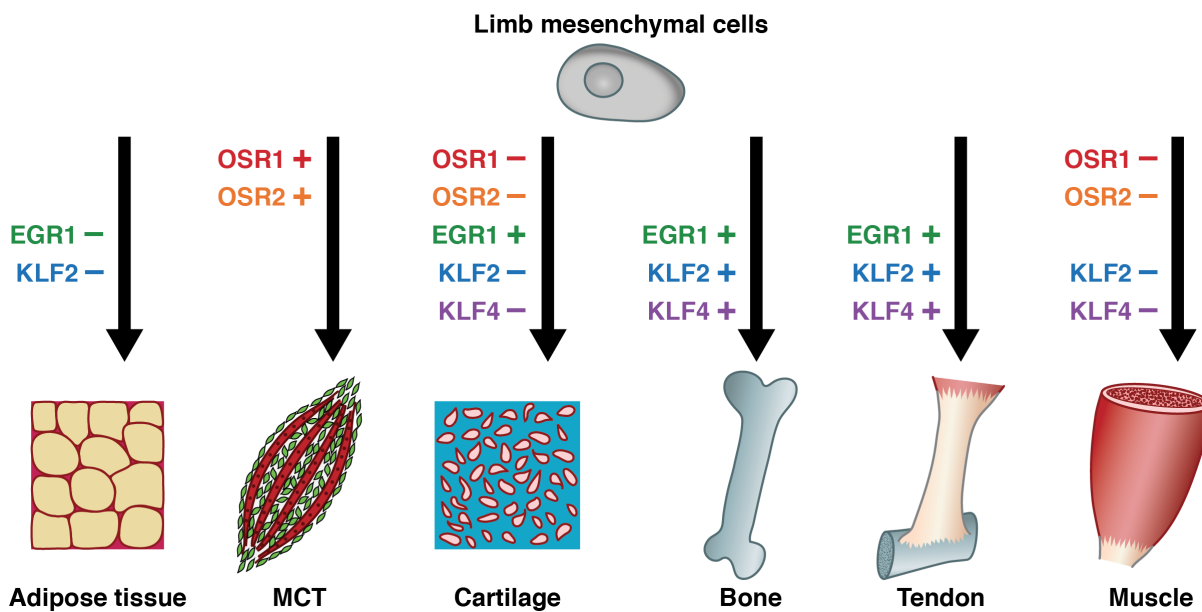


Figure 44. Overview of cell differentiation processes in chMM cultures. Cell type prediction was based on histological staining and changes in gene expression levels. When overexpressed in chMM cultures, the TFs positively (+) or negatively (-) influence the differentiation of limb mesenchymal cells into adipose tissue, MCT, cartilage, bone, tendon or muscle.

C. Regulatory profiles of the transcription factors

In addition to their distinct expression domains within the musculoskeletal system, the five TFs influence cell differentiation processes in chMM cultures. To further elucidate their molecular role, three levels of gene regulation were investigated by means of genome-wide strategies: (i) gene expression profiling to identify DE genes; (ii) chromatin signatures associated with promoters, enhancers and facultative heterochromatin to assess DNA regulatory domains; and (iii) TF-binding events to characterize their targeted genes.

Biological processes and signalling pathways regulated by connective tissue cells

The 10,712 DE genes detected across all TF-overexpressing chMM cultures corresponded to a list of 4,298 non-redundant genes (Figure 24A). 726 (16.9%) DE genes were identified in all TF-

overexpressing chMM cultures, indicating that the five TFs share a core of common regulatory processes despite their expression domain within different subcompartments of the musculoskeletal system. On the other hand, the TFs displayed specific regulatory patterns. KLF4 and OSR2 exhibited the highest specificity with 767 (17.8%) and 330 (7.7%) genes detected as differentially expressed only in each condition, respectively. By contrast, less than 150 (2.6-3.3%) DE genes were specifically detected upon overexpression of OSR1, EGR1 or KLF2. In addition, 2,811 (65.4%) genes were differentially expressed upon overexpression of at least two different TFs, including 419 (9.7%) and 492 (11.4%) specific genes common to both OSR and KLF factors, respectively. However, *K*-means clustering of all DE genes did not highlight an OSR2-specific cluster as compared to KLF4 (Figure 25A). Rather, DE genes identified upon overexpression of both OSR factors tended to cluster together, suggesting that they may share more targets than identified, although these genes did not pass the fixed thresholds.

Gene ontology (GO) analysis on the DE genes highlighted that all TFs share involvement in biological processes related to cell communication and migration, as well as biological adhesion (Figure 25B). Consistently, genes encoding cytoskeletal proteins, receptors, cell adhesion and signalling molecules as well as components of the extracellular matrix (ECM) were positively regulated by the TFs (Figure 27). The ECM is a three-dimensional insoluble and non-cellular network composed of secreted macromolecules, providing temporal, positional and physical information to influence developmental processes and cell migration (reviewed in Charras and Sahai 2014; Rozario and DeSimone 2010). ECM functionality relies on its composition between fibrous proteins (collagens, elastin) and glycoproteins (fibronectin, proteoglycans, laminin) that mainly contribute to its structural architecture and organization (reviewed in Mecham 2012). Further analysis of DE genes identified for each TF revealed that the TFs positively regulate components of the Integrin and Cadherin signalling pathways, as well as cytoskeletal organization controlled by Rho GTPase and axon guidance mediated by netrin (Figure 26). In addition, TFs positively regulate members of the WNT, TGF- β and Notch signalling pathways (Figure 26). The ECM acts as a source of developmental signals by sequestering and diffusing paracrine factors. For example, heparin sulphate proteoglycans can bind to WNT, FGF and TGF- β signalling components as well as *Ihh* to regulate chondrogenesis (reviewed in Kirn-Safran et al. 2004). If the ECM can accumulate and disseminate secreted molecules from one cell, it must also transmit these signals to surrounding cells. Cell-cell and cell-ECM interactions are mediated by receptors spanning from the cytoplasm to the extracellular region, namely the integrins (reviewed in Humphrey et al. 2014). Cadherins are cell surface glycoproteins involved in calcium-dependent cell adhesion and favour intercellular interactions (reviewed in Halbleib and Nelson 2006). On the other hand, the intracellular domain of the cadherins can alter cell

morphology and adhesion by interacting with the actin cytoskeleton. Cadherin-dependent remodelling of the actin cytoskeleton is mediated by Rho GTPases, which converts soluble actin into actin filaments that bind at the cadherins (reviewed in Etienne-Manneville and Hall 2002). Therefore ECM constitutes an important mediator during cell differentiation, migration and adhesion by providing an extensive scaffold for signal transduction between cell surface receptors and secreted factors. These results are consistent with the notion that connective tissue cells provide extrinsic signals for the patterning of surrounding tissues such as muscles.

Characterization of genes directly regulated by the TFs

The gene expression profiles reflect the global function of connective tissue cells, whereas distinguishing between directly and indirectly regulated genes highlight the regulatory programs specifically driven by the TFs. In order to assess the TF functionality, six putative target genes (*NTN1*, *WNT11*, *ADGRG2*, *WNT4*, *CBFA2T2* and *ANXA1*) were selected by associating their expression fold change upon TF overexpression in the chMM cultures with the presence of TFBS within regulatory domains located in their vicinity (Figures 39, 41A,B). *NTN1* is a laminin-related secreted protein, which is involved in axon guidance, biological adhesion via interactions with integrins, cytoskeletal organization and osteoclast differentiation (Mediero et al. 2015; Serafini et al. 1996; Srinivasan et al. 2003; Yebra et al. 2003). *WNT4* and *WNT11* are secreted components of the non-canonical planar cell polarity pathway, which act on cell adhesion and migration by interacting with cadherins and focal adhesion proteins. Both proteins seem to be involved in joint development and muscle fibre organization (Anakwe et al. 2003; Gros et al. 2009; Hartmann and Tabin 2000; Loganathan et al. 2005; Später et al. 2006; Takata et al. 2007; Witte et al. 2009). *ADGRG2* is an adhesion G protein-coupled receptor that is regulated by the TFs *Prdm16* and *Runx2* and thus may mediate adipocyte and osteoblast differentiation (Harms et al. 2014; Teplyuk et al. 2008). *CBFA2T2* encodes a transcriptional co-repressor acting on Notch signalling and recently characterized as a mediator of pluripotency and germline specification (Tu et al. 2016). *ANXA1* is a calcium-dependent phospholipid-binding protein that regulates cell migration and organization of the cytoskeleton, while being thought to mediate muscle development and regeneration (reviewed in Bizzarro et al. 2012; Leikina et al. 2015).

The six selected candidate genes displayed overlapping expression domains with their corresponding TF in limbs of E5.5 chick embryos (Figures 40, 41C-I). Further exploration of *WNT11*, *ADGRG2* and *ANXA1* expression in forelimbs at E8 confirmed the spatial and temporal link between the TFs and their respective target genes, while being seemingly not expressed in muscle fibres (Figures 42, 43). By contrast, *NTN1* and *WNT4* did not depict overlapping expression domains with their corresponding TFs in forelimbs of E8 chick embryos.

Nevertheless, they were detected in joints outside of muscle area (Supplemental figure S10). These candidate genes thus primarily validate the *in vitro* data obtained in the chMM cultures. It would be therefore interesting to further investigate genome-widely the TF occupancy profiles and the DE genes. Identifying the specific ECM components and signalling pathway members that each TF directly regulate would bring insights regarding their specific molecular mechanisms.

DE genes are enriched for bivalent promoter domains

Interestingly, exploring histone tail modifications related to transcriptional activity and repression (H3K4me3, H3K27me3) revealed that DE genes were significantly associated with bivalent promoter domains (Figures 30, 31, 32). Bivalent domains were originally identified in embryonic stem cells and enriched in lineage-regulatory genes (Azuara et al. 2006; Bernstein et al. 2006; Ku et al. 2008; Mikkelsen et al. 2007; Pan et al. 2007; Zhao et al. 2007). Based on their reversibility, two molecular mechanisms have been proposed. On one hand, H3K27me3 may contribute to maintain pluripotency by poising lineage-regulatory genes and prevent their aberrant transcription upon differentiation by tuning their activation. On the other hand, H3K4me3 may preserve lineage-regulatory genes from permanent silencing by impeding DNA methylation or the recruitment of transcriptional repressors, while allowing their rapid and synchronous transcriptional induction upon differentiation (Boettiger and Levine 2009; Fouse et al. 2008). Although a sequential ChIP on H3K4me3 and H3K27me3 marks would be necessary to validate the bivalent promoter structure of DE genes (Bernstein et al. 2006; Pan et al. 2007), co-localization of both histone modifications may simply reflect the cell heterogeneity in the chMM cultures (Hong et al. 2011). Indeed, DE genes could be expressed in a subpopulation of cells (H3K4me3), while being repressed in another subset of cells (H3K27me3). Both signals observed in the ChIP-seq data would therefore result from the average of each cell population. In both cases, bivalent domains indicate that DE genes are overall dynamically regulated in contrast to housekeeping and ubiquitous genes, active in most of the cells, and permanently repressed genes. This indicates that DE genes are differentially regulated in distinct cell populations and therefore likely to play a role in the differentiation and/or function of these populations.

For this study, the chromatin landscape was performed only in the chMM cultures infected with retroviral particles carrying no recombinant protein. It would be interesting to extend the investigation of the chosen histone modifications in the different TF-overexpressing chMM cultures. Indeed, a recent report has shown that chromatin profiles are highly dependent on the microenvironment and the ontogeny (Lavin et al. 2014). Tissue-resident macrophages display distinct enhancer signatures that correlate well with their different gene expression profiles and

embryonic origins. More intriguingly, transplantation of either adult bone-marrow macrophages or differentiated tissue-resident macrophages into another environment was sufficient to switch their chromatin landscape. Therefore, assessing chromatin dynamics in the chMM cultures upon TF overexpression could bring insights on critical enhancers that contribute to the differentiation of connective tissue cells. In addition, two recent reports have demonstrated that *Egr1* can mediate nucleosome positioning in murine hepatocytes and that *Klf4* can act as a pioneer factor and bind to compacted chromatin during human fibroblast reprogramming (Riffo-Campos et al. 2015; Soufi et al. 2012). This suggests that both factors may be involved in regulating chromatin accessibility during connective tissue cell differentiation. Extending the chromatin landscape to each TF-overexpressing chMM culture could also bring insights into the 8.6%-14.2% of TF-binding events that were located in regions of facultative heterochromatin (Figure 38A).

Identification of misregulated non-coding RNAs

Although this thesis has only focused on the DE genes with a known biological function, 550 (12.8%) genes encoding putative ncRNAs were identified as being differentially expressed in TF-overexpressing chMM cultures. Assessing the functionality of ncRNAs remains elusive due to their weak evolutionary conservation (Necsulea et al. 2014). While some ncRNAs arise from transcribed enhancers (eRNAs), others, referred as PROMPTs, originate from promoters in the reverse orientation of the coding genes (Core et al. 2008; De Santa et al. 2010; Kim et al. 2010; Preker et al. 2008; Seila et al. 2008). Although their role is presently unclear, PROMPTs and eRNAs could be distinguished from other ncRNAs by comparing their location with the presence of promoters and enhancers identified in the chMM cultures. Characterization of eRNA-producing enhancers is of particular interest given that eRNA synthesis is the first transcriptional event arising during the differentiation of developmental progenitor cells (Arner et al. 2015). On the other hand, long non-coding RNAs (lncRNAs) appears to be critical regulators during embryogenesis such as limb development (Carlson et al. 2015; Wang et al. 2011b). Recently, a study in the bat has identified several lncRNAs being differentially expressed between forelimbs and hindlimbs (Eckalbar et al. 2016). Two of these lncRNAs are expressed in interdigit tissue in the developing limb, suggesting that they could be involved in digit patterning. In addition, ncRNAs have been shown to regulate skeletal muscle formation (reviewed in Simionescu-Bankston and Kumar 2016). For instance, the lncRNA H19 promotes myoblast differentiation by repressing *Smad1*, *Smad5* and *Cdc6*, as well as mediates the proliferation and the number of satellite cells in the adult muscle (Dey et al. 2014; Martinet et al. 2016).

D. Future prospects

The transcriptional network presented here brings new insights into the molecular mechanisms orchestrated by the connective tissue-related TFs. Although additional connective tissue-associated TFs could be implemented, this regulatory network offers valuable resources and opens new investigation roads to better understand the influence of connective tissues during limb development. Of particular interest, it would be interesting to perform TF recognition motif analyses within the regulatory domains of directly and indirectly DE genes. Such TF-binding prediction could highlight additional co-factors and TFs that may extend the list of effector proteins involved in connective tissue patterning. It is noteworthy that besides *CBFA2T2*, whose expression was not detected in limbs of E8 chick embryos, the five remaining target genes selected were all expressed in connective tissues at this stage. Although supplementary stages are necessary to fully define their expression pattern, selected candidate genes in addition to the other DE genes should increase the limited number of connective tissue markers currently available.

The chMM culture appears to be a relevant *in vitro* model to investigate molecular and cellular interactions between the different cell types of the musculoskeletal system. The ability of limb mesenchymal cells to differentiate into distinct cell types in response to developmental signals offers multiple applications when *in vivo* testing is limiting. Prediction of cell differentiation relies so far on the expression of specific differentiation markers or histological staining. Given the increasing number of available data sets and continuous progress in genome-wide technologies, I believe that it will be feasible in a near future to predict more accurately the propensity of given cells to differentiate towards certain lineages within a multicellular population. Combining several layers of gene regulation, such as gene expression profiles, chromatin landscape and TF occupancy, should enable to decipher the developmental programs triggered by a progenitor cell during its commitment. Similar strategies have been applied in metagenomics by incorporating whole-genome sequencing and polymorphism signals to identify microbial species within a given environment or tissue (Human Microbiome Project Consortium 2012). Although the complexity within a multicellular organism is far beyond its genetic information, the recent development of single-cell transcriptome profiling seems to provide remarkable progress towards the determination of cell identity (Grün et al. 2016).

Tendon repair following injury often results in formation of scar tissue with reduced functional and mechanical properties as compared to native tendons. To date, there is no method that enables to restore normal tendon tissue in response to injury. Deciphering the molecular aspects of tendon development and differentiation would enhance the ability to provide better treatments

for tendon healing. Similarly, muscle regeneration following injury or in degenerative conditions remains not fully elucidated. In addition to the central role of satellite cells in muscle regeneration, emerging evidence have highlighted the importance of signals coming from surrounding tissues promoting muscle repair. MCT cells appear to be important mediators of regenerative mechanisms in the adult (Murphy et al. 2011; Stumm 2016). To a similar extent with tendons, elucidating the molecular interplay between MCT and muscles during development would bring valuable insights to better understand muscle repair.

REFERENCES

- Adelman K, Lis JT. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**: 720–31.
- Agger K, Cloos PAC, Christensen J, Pasini D, Rose S, Rappsilber J, Issaeva I, Canaani E, Salcini AE, Helin K. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**: 731–4.
- Ajore R, Kumar P, Dhanda RS, Gullberg U, Olsson I. 2012. The leukemia associated nuclear corepressor ETO homologue genes MTG16 and MTGR1 are regulated differently in hematopoietic cells. *BMC Mol Biol* **13**: 11.
- Akashah RT, Pang J, York JM, Fantuzzi G. 2013. New pathways to control inflammatory responses in adipose tissue. *Curr Opin Pharmacol* **13**: 613–7.
- Akita K. 1996. The effect of the ectoderm on the dorsoventral pattern of epidermis, muscles and joints in the developing chick leg: a new model. *Anat Embryol (Berl)* **193**: 377–86.
- Akiyama H, Lyons JP, Mori-Akiyama Y, Yang X, Zhang R, Zhang Z, Deng JM, Taketo MM, Nakamura T, Behringer RR, et al. 2004. Interactions between Sox9 and beta-catenin control chondrocyte differentiation. *Genes Dev* **18**: 1072–87.
- Alberton P, Popov C, Prägert M, Kohler J, Shukunami C, Schieker M, Docheva D. 2012. Conversion of human bone marrow-derived mesenchymal stem cells into tendon progenitor cells by ectopic expression of scleraxis. *Stem Cells Dev* **21**: 846–58.
- Allen BL, Taatjes DJ. 2015. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* **16**: 155–166.
- Amthor H, Christ B, Patel K. 1999. A molecular mechanism enabling continuous embryonic muscle growth - a balance between proliferation and differentiation. *Development* **126**: 1041–53.
- Amthor H, Christ B, Weil M, Patel K. 1998. The importance of timing differentiation during limb muscle development. *Curr Biol* **8**: 642–52.
- Anakwe K, Robson L, Hadley J, Buxton P, Church V, Allen S, Hartmann C, Harfe B, Nohno T, Brown AMC, et al. 2003. Wnt signalling regulates myogenic differentiation in the developing avian wing. *Development* **130**: 3503–14.
- Anderson DM, Arredondo J, Hahn K, Valente G, Martin JF, Wilson-Rawls J, Rawls A. 2006. Mohawk is a novel homeobox gene expressed in the developing mouse embryo. *Dev Dyn* **235**: 792–801.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* **507**: 455–61.
- Andrés V, Walsh K. 1996. Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. *J Cell Biol* **132**: 657–66.
- Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**: 1234167.
- Antin PB, Pier M, Sesepasara T, Yatskievych T a, Darnell DK. 2010. Embryonic expression of the chicken Krüppel-like (KLF) transcription factor gene family. *Dev Dyn* **239**: 1879–87.
- Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Ronnerblad M, Hrydzusko O, Vitezic M, et al. 2015. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**: 1010–4.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–9.
- Auradé F, Pinset C, Chafey P, Gros F, Montarras D. 1994. Myf5, MyoD, myogenin and MRF4 myogenic derivatives of the embryonic mesenchymal cell line C3H10T1/2 exhibit the same adult muscle phenotype. *Differentiation* **55**: 185–92.
- Ausio J, Dong F, van Holde KE. 1989. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone “tails” in the stabilization of the nucleosome. *J Mol Biol* **206**: 451–63.
- Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, et al. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**: 532–8.
- Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science* **324**: 1720–3.
- Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, Liu T, Madrigal P, Taslim C, Zhang J. 2013. Practical

- guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Comput Biol* **9**: 5–12.
- Bailey TL. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**: 1653–9.
- Bairoch A, Boeckmann B, Ferro S, Gasteiger E. 2004. Swiss-Prot: juggling between evolution and stability. *Brief Bioinform* **5**: 39–55.
- Ballester B, Medina-Rivera A, Schmidt D, González-Porta M, Carlucci M, Chen X, Chessman K, Faure AJ, Funnell APW, Gonçalves A, et al. 2014. Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *Elife* **3**: e02626.
- Banerji J, Olson L, Schaffner W. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**: 729–40.
- Banerji J, Rusconi S, Schaffner W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308.
- Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* **21**: 381–95.
- Bannister AJ, Schneider R, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**: 17732–6.
- Bardet AF, He Q, Zeitlinger J, Stark A. 2012. A computational pipeline for comparative ChIP-seq analyses. *Nat Protoc* **7**: 45–61.
- Barrera LO, Li Z, Smith AD, Arden KC, Cavenee WK, Zhang MQ, Green RD, Ren B. 2008. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res* **18**: 46–59.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–37.
- Benazet J, Bischofberger M, Tiecke E, Gonçalves A, Martin J, Zuniga A, Naef F, Zeller R. 2009. A self-regulatory system of interlinked signaling feedback loops controls mouse limb patterning. *Science* **323**: 1050–3.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57**: 289–300.
- Bergstrom DA, Tapscott SJ. 2001. Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family. *Mol Cell Biol* **21**: 2404–12.
- Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. 2004. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol Cell* **14**: 465–77.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315–26.
- Bettini M, Xi H, Milbrandt J, Kersh GJ. 2002. Thymocyte development in early growth response gene 1-deficient mice. *J Immunol* **169**: 1713–20.
- Bhattacharyya S, Wu M, Fang F, Tourtellotte W, Feghali-Bostwick C, Varga J. 2011. Early growth response transcription factors: key mediators of fibrosis and novel targets for anti-fibrotic therapy. *Matrix Biol* **30**: 235–42.
- Bianco P, Fisher LW, Young MF, Termine JD, Robey PG. 1991. Expression of bone sialoprotein (BSP) in developing human tissues. *Calcif Tissue Int* **49**: 421–6.
- Biason-Lauber A, Chaboissier M. 2015. Ovarian development and disease: the known and the unexpected. *Semin Cell Dev Biol* **45**: 59–67.
- Biau S, Bayle S, de Santa Barbara P, Roig B. 2007. The chick embryo: an animal model for detection of the effects of hormonal compounds. *Anal Bioanal Chem* **387**: 1397–403.
- Bienz M, Pelham HR. 1986. Heat shock regulatory elements function as an inducible enhancer in the *Xenopus* hsp70 gene and when linked to a heterologous promoter. *Cell* **45**: 753–60.
- Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**: 3045–6.
- Birsoy K, Chen Z, Friedman J. 2008. Transcriptional Regulation of Adipogenesis by KLF4. *Cell Metab* **7**: 339–47.
- Bizzarro V, Petrella A, Parente L. 2012. Annexin A1: novel roles in skeletal muscle biology. *J Cell Physiol* **227**: 3007–15.
- Bladt F, Riethmacher D, Isenmann S, Aguzzi A, Birchmeier C. 1995. Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature* **376**: 768–71.
- Blitz E, Sharir A, Akiyama H, Zelzer E. 2013. Tendon-bone attachment unit is formed modularly by a distinct pool of Scx- and Sox9-positive progenitors. *Development* **140**: 2680–90.
- Bloom S, Delany M, Muscarella D. 1993. Constant and variable features of avian chromosomes. In *Manipulation of*

- the avian genome*. (eds. A. Gibbins and R. Etches), pp. 39–59.
- Boettiger AN, Levine M. 2009. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325**: 471–3.
- Bohmann D, Keller W, Dale T, Schöler HR, Tebb G, Mattaj IW. A transcription factor which binds to the enhancers of SV40, immunoglobulin heavy chain and U2 snRNA genes. *Nature* **325**: 268–72.
- Bökel C, Brown NH. 2002. Integrins in development: moving on, responding to, and sticking to the extracellular matrix. *Dev Cell* **3**: 311–21.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–20.
- Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EE. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**: 148–56.
- Bonnin M-A, Laclef C, Blaise R, Eloy-Trinquet S, Relaix F, Maire P, Duprez D. 2005. Six1 is not involved in limb tendon development, but is expressed in limb connective tissue under Shh regulation. *Mech Dev* **122**: 573–85.
- Bourgeois A, Esteves de Lima J, Charvet B, Kawakami K, Stricker S, Duprez D. 2015. Stable and bicistronic expression of two genes in somite- and lateral plate-derived tissues to study chick limb development. *BMC Dev Biol* **15**: 39.
- Boyle KB, Hadaschik D, Virtue S, Cawthorn WP, Ridley SH, O’Rahilly S, Siddle K. 2009. The transcription factors Egr1 and Egr2 have opposing influences on adipocyte differentiation. *Cell Death Differ* **16**: 782–9.
- Bradley RK, Li X, Trapnell C, Davidson S, Pachter L, Chu HC, Tonkin LA, Biggin MD, Eisen MB. 2010. Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. *PLoS Biol* **8**: e1000343.
- Braghetta P, Ferrari A, Fabbro C, Bizzotto D, Volpin D, Bonaldo P, Bressan GM. 2008. An enhancer required for transcription of the Col6a1 gene in muscle connective tissue is induced by signals released from muscle cells. *Exp Cell Res* **314**: 3508–18.
- Bren-Mattison Y, Hausburg M, Olwin BB. 2011. Growth of limb muscle is dependent on skeletal-derived Indian hedgehog. *Dev Biol* **356**: 486–95.
- Brent AE, Braun T, Tabin CJ. 2005. Genetic analysis of interactions between the somitic muscle, cartilage and tendon cell lineages during mouse development. *Development* **132**: 515–28.
- Brohmann H, Jagla K, Birchmeier C. 2000. The role of Lbx1 in migration of muscle precursor cells. *Development* **127**: 437–45.
- Buck MJ, Lieb JD. 2006. A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nat Genet* **38**: 1446–51.
- Buckingham M, Bajard L, Chang T, Daubas P, Hadchouel J, Meilhac S, Montarras D, Rocancourt D, Relaix F. 2003. The formation of skeletal muscle: from soma to limb. *J Anat* **202**: 59–68.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Cameron TL, Belluoccio D, Farlie PG, Brachvogel B, Bateman JF. 2009. Global comparative transcriptome analysis of cartilage formation in vivo. *BMC Dev Biol* **9**: 20.
- Cantor AB, Iwasaki H, Arinobu Y, Moran TB, Shigematsu H, Sullivan MR, Akashi K, Orkin SH. 2008. Antagonism of FOG-1 and GATA factors in fate choice for the mast cell lineage. *J Exp Med* **205**: 611–24.
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**: 1039–43.
- Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, et al. 2010a. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* **18**: 662–74.
- Cao Z, Sun X, Icli B, Wara AK, Feinberg MW. 2010b. Role of Kruppel-like factors in leukocyte development, function, and disease. *Blood* **116**: 4404–14.
- Carlson HL, Quinn JJ, Yang YW, Thornburg CK, Chang HY, Stadler HS. 2015. LncRNA-HIT functions as an epigenetic regulator of chondrogenesis through its recruitment of p100/CBP complexes. *PLoS Genet* **11**: e1005680.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**: 626–35.
- Carvajal JJ, Cox D, Summerbell D, Rigby PW. 2001. A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. *Development* **128**: 1857–68.

- Cerdá-Esteban N, Spagnoli FM. 2014. Glimpse into Hox and tale regulation of cell differentiation and reprogramming. *Dev Dyn* **243**: 76–87.
- Charras G, Sahai E. 2014. Physical influences of the extracellular environment on cell migration. *Nat Rev Mol Cell Biol* **15**: 813–24.
- Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, et al. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**: 1274–85.
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106–17.
- Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. 2013. Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* **8**: e62856.
- Chevallier A, Kieny M, Mauger A. 1977. Limb-somite relationship: origin of the limb musculature. *J Embryol Exp Morphol* **41**: 245–58.
- Chlon TM, Doré LC, Crispino JD. 2012. Cofactor-Mediated Restriction of GATA-1 Chromatin Occupancy Coordinates Lineage-Specific Gene Expression. *Mol Cell* **47**: 608–21.
- Choi J, Costa ML, Mermelstein CS, Chagas C, Holtzer S, Holtzer H. 1990. MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc Natl Acad Sci U S A* **87**: 7988–92.
- Christ B, Jacob HJ, Jacob M. 1977. Experimental analysis of the origin of the wing musculature in avian embryos. *Anat Embryol (Berl)* **150**: 171–86.
- Church V, Nohno T, Linker C, Marcelle C, Francis-West P. 2002. Wnt regulation of chondrocyte differentiation. *J Cell Sci* **115**: 4809–18.
- Clements A, Poux AN, Lo W-S, Pillus L, Berger SL, Marmorstein R. 2003. Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase. *Mol Cell* **12**: 461–73.
- Collins CA, Gnocchi VF, White RB, Boldrin L, Perez-Ruiz A, Relaix F, Morgan JE, Zammit PS. 2009. Integrated functions of Pax3 and Pax7 in the regulation of proliferation, cell size and myogenic differentiation. *PLoS One* **4**: e4475.
- Cooper KL, Sears KE, Uygur A, Maier J, Baczkowski K, Brosnahan M, Antczak D, Skidmore JA, Tabin CJ. 2014. Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511**: 41–5.
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–20.
- Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845–8.
- Cosgrove MS, Boeke JD, Wolberger C. 2004. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* **11**: 1037–43.
- Coumailleau P, Duprez D. 2009. Sim1 and Sim2 expression during chick and mouse limb development. *Int J Dev Biol* **53**: 149–57.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**: 21931–6.
- Cusanovich DA, Pavlovic B, Pritchard JK, Gilad Y. 2014. The functional consequences of variation in transcription factor binding. *PLoS Genet* **10**: e1004226.
- Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. 2002. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**: 185–96.
- Daniels K, Reiter R, Solursh M. 1996. Micromass cultures of limb and other mesenchyme. *Methods Cell Biol* **51**: 237–47.
- Daston G, Lamar E, Olivier M, Goulding M. 1996. Pax-3 is necessary for migration but not differentiation of limb muscle precursors in the mouse. *Development* **122**: 1017–27.
- Davies B, Baumann C, Kirchhoff C, Ivell R, Nubbemeyer R, Habenicht U-F, Theuring F, Gottwald U. 2004. Targeted deletion of the epididymal receptor HE6 results in fluid dysregulation and male infertility. *Mol Cell Biol* **24**: 8642–8.
- Davis RL, Weintraub H, Lassar AB. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987–1000.
- de la Serna IL, Ohkawa Y, Berkes CA, Bergstrom DA, Dacwag CS, Tapscott SJ, Imbalzano AN. 2005. MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol Cell Biol* **25**: 3997–4009.

- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei C-L, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**: e1000384.
- de Wit E, Bouwman BAM, Zhu Y, Klous P, Splinter E, Verstegen MJAM, Krijger PHL, Festuccia N, Nora EP, Welling M, et al. 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* **501**: 227–31.
- Dealy CN, Roth A, Ferrari D, Brown AM, Kosher RA. 1993. Wnt-5a and Wnt-7a are expressed in the developing chick limb bud in a manner suggesting roles in pattern formation along the proximodistal and dorsoventral axes. *Mech Dev* **43**: 175–86.
- Delfini M-C, Duprez D. 2004. Ectopic Myf5 or MyoD prevents the neuronal differentiation program in addition to inducing skeletal muscle differentiation, in the chick neural tube. *Development* **131**: 713–23.
- Delfini MC, Hirsinger E, Pourquié O, Duprez D. 2000. Delta 1-activated notch inhibits muscle differentiation without affecting Myf5 and Pax3 expression in chick limb myogenesis. *Development* **127**: 5213–24.
- DeLise AM, Stringa E, Woodward WA, Mello MA, Tuan RS. 2000. Embryonic limb mesenchyme micromass culture as an in vitro model for chondrogenesis and cartilage maturation. *Methods Mol Biol* **137**: 359–75.
- Dex S, Lin D, Shukunami C, Docheva D. 2016. Tenogenic modulating insider factor: systematic assessment on the functions of tenomodulin gene. *Gene* **587**: 1–17.
- Dey BK, Pfeifer K, Dutta A. 2014. The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev* **28**: 491–501.
- Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, et al. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**: 671–83.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–80.
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101–08.
- Docheva D, Hunziker EB, Fässler R, Brandau O. 2005. Tenomodulin is necessary for tenocyte proliferation and tendon maturation. *Mol Cell Biol* **25**: 699–705.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**: e105.
- Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, Gingeras TR, Gerstein M, Guigó R, Birney E, et al. 2012. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol* **13**: R53.
- Dunn OJ. 1961. Multiple comparisons among means. *J Am Stat Assoc* **56**: 52–64.
- Duprez D. 2002. Signals regulating muscle formation in the limb during embryonic development. *Int J Dev Biol* **46**: 915–25.
- Duprez D, Fournier-Thibault C, Le Douarin N. 1998. Sonic Hedgehog induces proliferation of committed skeletal muscle cells in the chick limb. *Development* **125**: 495–505.
- Duprez D, Lapointe F, Edom-Vovard F, Kostakopoulou K, Robson L. 1999. Sonic Hedgehog (SHH) specifies muscle pattern at tissue and cellular chick level, in the chick limb bud. *Mech Dev* **82**: 151–63.
- Duprez DM, Coltey M, Amthor H, Brickell PM, Tickle C. 1996a. Bone morphogenetic protein-2 (BMP-2) inhibits muscle development and promotes cartilage formation in chick limb bud cultures. *Dev Biol* **174**: 448–452.
- Duprez DM, Kostakopoulou K, Francis-West PH, Tickle C, Brickell PM. 1996b. Activation of Fgf-4 and HoxD gene expression by BMP-2 expressing cells in the developing chick limb. *Development* **122**: 1821–8.
- Duttke SHC, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U. 2015. Human promoters are intrinsically directional. *Mol Cell* **57**: 674–84.
- Eckalbar WL, Schlebusch SA, Mason MK, Gill Z, Parker A V, Booker BM, Nishizaki S, Muswamba-Nday C, Terhune E, Nevonen KA, et al. 2016. Transcriptomic and epigenomic characterization of the developing bat wing. *Nat Genet* **48**: 528–36.
- Edom-Vovard F, Bonnin MA, Duprez D. 2001. Misexpression of Fgf-4 in the chick limb inhibits myogenesis by down-regulating Frk expression. *Dev Biol* **233**: 56–71.
- Edom-Vovard F, Schuler B, Bonnin MA, Teillet MA, Duprez D. 2002. Fgf4 positively regulates scleraxis and tenascin expression in chick limb tendons. *Dev Biol* **247**: 351–66.
- Eloy-Trinquet S, Wang H, Edom-Vovard F, Duprez D. 2009. Fgf signaling components are associated with muscles and tendons during limb development. *Dev Dyn* **238**: 1195–206.
- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

- Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**: 817–25.
- Ernst J, Kellis M. 2013. Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. *Genome Res* **23**: 1142–54.
- Etienne-Manneville S, Hall A. 2002. Rho GTPases in cell biology. *Nature* **420**: 629–35.
- FANTOM Consortium, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy I V., Lizio M, et al. 2014. A promoter-level mammalian expression atlas. *Nature* **507**: 462–70.
- Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS. 2015. Suboptimization of developmental enhancers. *Science* **350**: 325–8.
- Ferrell JE. 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* **14**: 140–8.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–37.
- Fischer L, Boland G, Tuan RS. 2002. Wnt-3A enhances bone morphogenetic protein-2-mediated chondrogenesis of murine C3H10T1/2 mesenchymal cells. *J Biol Chem* **277**: 30870–8.
- Fisher LW, McBride OW, Termine JD, Young MF. 1990. Human bone sialoprotein. Deduced protein sequence and chromosomal localization. *J Biol Chem* **265**: 2347–51.
- Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB, et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proc Natl Acad Sci USA* **109**: 21330–5.
- Foley JW, Sidow A. 2013. Transcription-factor occupancy at HOT regions quantitatively predicts RNA polymerase recruitment in five human cell lines. *BMC Genomics* **14**: 720.
- Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**: 160–69.
- Frank NY, Kho AT, Schatton T, Murphy GF, Molloy MJ, Zhan Q, Ramoni MF, Frank MH, Kohane IS, Gussoni E. 2006. Regulation of myogenic progenitor proliferation in human fetal skeletal muscle by BMP4 and its antagonist Gremlin. *J Cell Biol* **175**: 99–110.
- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed Y Bin, Orlov YL, Velkov S, Ho A, Mei PH, et al. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58–64.
- Gao Y, Lan Y, Liu H, Jiang R. 2011. The zinc finger transcription factors Osr1 and Osr2 control synovial joint formation. *Dev Biol* **352**: 83–91.
- Garber M, Yosef N, Goren A, Raychowdhury R, Thielke A, Guttman M, Robinson J, Minie B, Chevrier N, Itzhaki Z, et al. 2012. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol Cell* **47**: 810–22.
- Gaut L, Duprez D. 2016. Tendon development and diseases. *Wiley Interdiscip Rev Dev Biol* **5**: 5–23.
- Gavard J, Marthiens V, Monnet C, Lambert M, Mège RM. 2004. N-cadherin activation substitutes for the cell contact control in cell cycle arrest and myogenic differentiation: involvement of p120 and beta-catenin. *J Biol Chem* **279**: 36795–802.
- Gerber HP, Vu TH, Ryan AM, Kowalski J, Werb Z, Ferrara N. 1999. VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation. *Nat Med* **5**: 623–8.
- Germain PL, Ratti E, Boem F. 2014. Junk or functional DNA? ENCODE and the function controversy. *Biol Philos* **29**: 807–31.
- Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. *Science* **330**: 1775–87.
- Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100.
- Gilbert SF. 2013. Development of the Tetrapod limb. In *Developmental Biology* (ed. A. Aquadron), pp. 489–518, Andrew S. Sinauer, Sunderland, USA.
- Goichberg P, Shtutman M, Ben-Ze'ev A, Geiger B. 2001. Recruitment of beta-catenin to cadherin-mediated intercellular adhesions is involved in myogenic induction. *J Cell Sci* **114**: 1309–19.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644–52.

- Grim M, Wachtler F. 1991. Muscle morphogenesis in the absence of myogenic cells. *Anat Embryol (Berl)* **183**: 67–70.
- Gros J, Serralbo O, Marcelle C. 2009. WNT11 acts as a directional cue to organize the elongation of early muscle fibres. *Nature* **457**: 589–93.
- Gros J, Tabin CJ. 2014. Vertebrate limb bud formation is initiated by localized epithelial-to-mesenchymal transition. *Science* **343**: 1253–6.
- Gross MK, Moran-Rivard L, Velasquez T, Nakatsu MN, Jagla K, Goulding M. 2000. Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development* **127**: 413–24.
- Grün D, Muraro MJ, Boisset J-C, Wiebrands K, Lyubimova A, Dharmadhikari G, van den Born M, van Es J, Jansen E, Clevers H, et al. 2016. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* [Epub ahead of print].
- Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**: 77–88.
- Guerquin MJ, Charvet B, Nourissat G, Havis E, Ronsin O, Bonnin MA, Ruggiu M, Olivera-Martinez I, Robert N, Lu Y, et al. 2013. Transcription factor EGR1 directs tendon differentiation and promotes tendon repair. *J Clin Invest* **123**: 3564–76.
- Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W, Maniatis T, Goodbourn S, Fischer J, Pawson T, Nash P, Tompa M, et al. 2007. Quantifying similarity between motifs. *Genome Biol* **8**: R24.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–512.
- Halbleib JM, Nelson WJ. 2006. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev* **20**: 3199–214.
- Hall BK, Herring SW. 1990. Paralysis and growth of the musculoskeletal system in the embryonic chick. *J Morphol* **206**: 45–56.
- Hamburger V, Hamilton HL. 1951. A series of normal stages in the development of the chick embryo. *J Morphol* **88**: 49–92.
- Hamon MA, Cossart P. 2008. Histone modifications and chromatin remodeling during bacterial infections. *Cell Host Microbe* **4**: 100–09.
- Hansson ML, Behmer S, Ceder R, Mohammadi S, Preta G, Grafström RC, Fadeel B, Wallberg AE. 2012. MAML1 acts cooperatively with EGR1 to activate EGR1-regulated promoters: implications for nephrogenesis and the development of renal cancer. *PLoS One* **7**: e46001.
- Harfe BD, Scherz PJ, Nissim S, Tian H, McMahon AP, Tabin CJ. 2004. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**: 517–28.
- Harms MJ, Ishibashi J, Wang W, Lim H-W, Goyama S, Sato T, Kurokawa M, Won K-J, Seale P. 2014. Prdm16 is required for the maintenance of brown adipocyte identity and function in adult mice. *Cell Metab* **19**: 593–604.
- Hartigan J, Wong M. 1979. A K-means clustering algorithm. *J R Stat Soc* **28**: 100–08.
- Hartmann C, Tabin CJ. 2000. Dual roles of Wnt signaling during chondrogenesis in the chicken limb. *Development* **127**: 3141–59.
- Hasson P, Del Buono J, Logan MPO. 2007. Tbx5 is dispensable for forelimb outgrowth. *Development* **134**: 85–92.
- Hasson P, DeLaurier A, Bennett M, Grigorieva E, Naiche LA, Papaioannou VE, Mohun TJ, Logan MPO. 2010. Tbx4 and Tbx5 acting in connective tissue are required for limb muscle and tendon patterning. *Dev Cell* **18**: 148–56.
- Hatori M, Klätte KJ, Teixeira CC, Shapiro IM. 1995. End labeling studies of fragmented DNA in the avian growth plate: evidence of apoptosis in terminally differentiated chondrocytes. *J Bone Miner Res* **10**: 1960–8.
- Havis E, Bonnin M-A, Olivera-Martinez I, Nazaret N, Ruggiu M, Weibel J, Durand C, Guerquin M-J, Bonod-Bidaud C, Ruggiero F, et al. 2014. Transcriptomic analysis of mouse limb tendon cells during development. *Development* **141**: 3683–96.
- Havis E, Coumailleau P, Bonnet A, Bismuth K, Bonnin MA, Johnson R, Fan C-M, Relaix F, Shi D-L, Duprez D. 2012. Sim2 prevents entry into the myogenic program by repressing MyoD transcription during limb embryonic myogenesis. *Development* **139**: 1910–20.
- Hayashi K, Ozawa E. 1995. Myogenic cell migration from somites is induced by tissue contact with medial region of the presumptive limb mesoderm in chick embryos. *Development* **121**: 661–9.
- He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M, et al. 2010. Nucleosome dynamics define transcriptional enhancers. *Nat Genet* **42**: 343–7.
- He Q, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, Zeitlinger J. 2011. High

- conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat Genet* **43**: 414–20.
- Heidari N, Phanstiel D, He C. 2014. Genome-wide map of regulatory interactions in the human genome. *Genome Res* **24**: 1905–17.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–12.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–8.
- Heymann S, Koudrova M, Arnold H, Köster M, Braun T. 1996. Regulation and function of SF/HGF during migration of limb muscle precursor cells in chicken. *Dev Biol* **180**: 566–78.
- Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MAM, Delany ME. 2004. Sequencing and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.
- Hindi SM, Tajrishi MM, Kumar A. 2013. Signaling mechanisms in mammalian myoblast fusion. *Sci Signal* **6**: re2.
- Hong S-H, Rampalli S, Lee JB, McNicol J, Collins T, Draper JS, Bhatia M. 2011. Cell fate potential of human pluripotent stem cells is encoded by histone modifications. *Cell Stem Cell* **9**: 24–36.
- Horsley V, Jansen KM, Mills ST, Pavlath GK. 2003. IL-4 acts as a myoblast recruitment factor during mammalian muscle growth. *Cell* **113**: 483–94.
- Hosseini A, Hogg DA. 1991. The effects of paralysis on skeletal development in the chick embryo. II. Effects on histogenesis of the tibia. *J Anat* **177**: 169–78.
- Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. *J Educ Psychol* **24**: 417–41.
- Houzelstein D, Auda-Boucher G, Chéraud Y, Rouaud T, Blanc I, Tajbakhsh S, Buckingham ME, Fontaine-Pérus J, Robert B. 1999. The homeobox gene *Msx1* is expressed in a subset of somites, and in muscle progenitor cells migrating into the forelimb. *Development* **126**: 2689–701.
- Hu H, Hilton MJ, Tu X, Yu K, Ornitz DM, Long F. 2005. Sequential roles of Hedgehog and Wnt signaling in osteoblast development. *Development* **132**: 49–60.
- Huang AH, Riordan TJ, Pryce B, Weibel JL, Watson SS, Long F, Lefebvre V, Harfe BD, Stadler HS, Akiyama H, et al. 2015. Musculoskeletal integration at the wrist underlies modular development of limb tendons. *Development* **243**: 1–41.
- Huang AH, Riordan TJ, Wang L, Eyal S, Zelzer E, Brigande J V, Schweitzer R. 2013. Repositioning forelimb superficialis muscles: tendon attachment and muscle activity enable active relocation of functional myofibers. *Dev Cell* **26**: 544–51.
- Hughes SH. 2004. The RCAS system. *Folia Biol* **50**: 107–19.
- Human Microbiome Project Consortium. 2012. A framework for human microbiome research. *Nature* **486**: 215–21.
- Humphrey JD, Dufresne ER, Schwartz MA. 2014. Mechanotransduction and extracellular matrix homeostasis. *Nat Rev Mol Cell Biol* **15**: 802–12.
- Hunter E. 1997. Viral entry and receptors. In *Retroviruses* (eds. J. Coffin, S. Hughes, and H. Varmus), Cold Spring Harbor Laboratory Press.
- Hurle JM, Ros MA, Gañan Y, Macias D, Critchlow M, Hinchliffe JR. 1990. Experimental analysis of the role of ECM in the patterning of the distal tendons of the developing limb bud. *Cell Differ Dev* **30**: 97–108.
- Hutcheson DA, Zhao J, Merrell A, Haldar M, Kardon G. 2009. Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for beta-catenin. *Genes Dev* **23**: 997–1013.
- Ibrahim DM, Hansen P, Rödelsperger C, Stiege AC, Doelken SC, Horn D, Jäger M, Janetzki C, Krawitz P, Leschik G, et al. 2013. Distinct global shifts in genomic binding profiles of limb malformation-associated HOXD13 mutations. *Genome Res* **23**: 2091–102.
- Imanishi T, Nakaoka H. 2009. Hyperlink Management System and ID Converter System: enabling maintenance-free hyperlinks among major biological databases. *Nucleic Acids Res* **37**: W17–22.
- Ito Y, Toriuchi N, Yoshitaka T, Ueno-Kudoh H, Sato T, Yokoyama S, Nishida K, Akimoto T, Takahashi M, Miyaki S, et al. 2010. The Mohawk homeobox gene is a critical regulator of tendon differentiation. *Proc Natl Acad Sci USA* **107**: 10538–42.
- Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, Zaret KS. 2016. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell* **62**: 79–91.

- Jacob M, Christ B, Jacob HJ. 1978. On the migration of myogenic stem cells into the prospective wing region of chick embryos. A scanning and transmission electron microscope study. *Anat Embryol (Berl)* **153**: 179–93.
- James RG, Kamei CN, Wang Q, Jiang R, Schultheiss TM. 2006. Odd-skipped related 1 is required for development of the metanephric kidney and regulates formation and differentiation of kidney precursor cells. *Development* **133**: 2995–3004.
- Jiang J, Chan Y-S, Loh Y-H, Cai J, Tong G-Q, Lim C-A, Robson P, Zhong S, Ng H-H. 2008. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol* **10**: 353–60.
- Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen C-A, Schmitt AD, Espinoza C a, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**: 290–4.
- John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**: 264–8.
- Johnson DS, Mortazavi A, Myers RM. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**: 1497–502.
- Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, et al. 2013. DNA-binding specificities of human transcription factors. *Cell* **152**: 327–39.
- Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**: 384–8.
- Jonkers I, Kwak H, Lis JT. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**: e02407.
- Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**: 167–77.
- Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EE. 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**: 473–86.
- Kahn J, Shwartz Y, Blitz E, Krief S, Sharir A, Breitel DA, Rattenbach R, Relaix F, Maire P, Rountree RB, et al. 2009. Muscle contraction is necessary to maintain joint progenitor cell fate. *Dev Cell* **16**: 734–43.
- Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, et al. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell* **51**: 310–25.
- Kaplan T, Li X-Y, Sabo PJ, Thomas S, Stamatoyannopoulos JA, Biggin MD, Eisen MB. 2011. Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS Genet* **7**: e1001290.
- Kardon G. 1998. Muscle and tendon morphogenesis in the avian hind limb. *Development* **125**: 4019–32.
- Kardon G, Harfe BD, Tabin CJ. 2003. A Tcf4-positive mesodermal population provides a prepattern for vertebrate limb muscle patterning. *Dev Cell* **5**: 937–44.
- Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, et al. 2014. Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* **111**: 6131–8.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996–1006.
- Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**: 2204–7.
- Kharchenko P V, Tolstorukov MY, Park PJ. 2008. Design and analysis of CHIP-seq experiments for DNA-binding proteins. *Nat Biotech* **26**: 1351–9.
- Kieny M, Chevallier A. 1979. Autonomy of tendon development in the embryonic chick wing. *J Embryol Exp Morphol* **49**: 153–65.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.
- Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–7.
- Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436**: 876–80.
- Kim TK, Shiekhhattar R. 2015. Architectural and functional commonalities between enhancers and promoters. *Cell* **162**: 948–59.
- Kim YJ, Björklund S, Li Y, Sayre MH, Kornberg RD. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**: 599–608.
- Kimura W, Machii M, Xue X, Sultana N, Hikosaka K, Sharkar MT, Uezato T, Matsuda M, Koseki H, Miura N.

2011. Irx11 mutant mice show reduced tendon differentiation and no patterning defects in musculoskeletal system development. *Genesis* **49**: 2–9.
- Kirchhoff C, Obermann H, Behnen M, Davies B. 2006. Role of epididymal receptor HE6 in the regulation of sperm microenvironment. *Mol Cell Endocrinol* **250**: 43–8.
- Kirn-Safran CB, Gomes RR, Brown AJ, Carson DD. 2004. Heparan sulfate proteoglycans: coordinators of multiple signaling pathways during chondrogenesis. *Birth Defects Res C Embryo Today* **72**: 69–88.
- Kitagaki J, Iwamoto M, Liu JG, Tamamura Y, Pacifici M, Enomoto-Iwamoto M. 2003. Activation of beta-catenin-LEF/TCF signal pathway in chondrocytes stimulates ectopic endochondral ossification. *Osteoarthritis Cartilage* **11**: 36–43.
- Kitzmann M, Carnac G, Vandromme M, Primig M, Lamb NJ, Fernandez A. 1998. The muscle regulatory factors MyoD and myf-5 undergo distinct cell cycle-specific expression in muscle cells. *J Cell Biol* **142**: 1447–59.
- Knapska E, Kaczmarek L. 2004. A gene for neuronal plasticity in the mammalian brain: Zif268/Egr-1/NGFI-A/Krox-24/TIS8/ZENK? *Prog Neurobiol* **74**: 183–211.
- Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, et al. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* **18**: 956–63.
- Korinek V, Barker N, Morin PJ, van Wichen D, de Weger R, Kinzler KW, Vogelstein B, Clevers H. 1997. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275**: 1784–7.
- Kowalczyk MS, Hughes JR, Garrick D, Lynch MD, Sharpe JA, Sloane-Stanley JA, McGowan SJ, De Gobbi M, Hosseini M, Vernimmen D, et al. 2012. Intragenic enhancers act as alternative promoters. *Mol Cell* **45**: 447–58.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–80.
- Krüger M, Mennerich D, Fees S, Schäfer R, Mundlos S, Braun T. 2001. Sonic hedgehog is a survival factor for hypaxial muscles during mouse development. *Development* **128**: 743–52.
- Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**: e1000242.
- Kulkarni MM, Arnosti DN. 2003. Information display by transcriptional enhancers. *Development* **130**: 6569–75.
- Kuss P, Villavicencio-Lorini P, Witte F, Klose J, Albrecht AN, Seemann P, Hecht J, Mundlos S. 2009. Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis. *J Clin Invest* **119**: 146–56.
- Kutchuk L, Laitala A, Soueid-Bomgarten S, Shentzer P, Rosendahl AH, Eilott S, Grossman M, Sagi I, Sormunen R, Myllyharju J, et al. 2015. Muscle composition is regulated by a Lox-TGFbeta feedback loop. *Development* **142**: 983–93.
- Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**: 2893–905.
- Kvon EZ, Stampfel G, Yáñez-Cuna JO, Farley E, Levine M, Dickson BJ, Ya JO, Stark A. 2012. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev* **26**: 908–13.
- Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**: 950–3.
- Lam MTY, Li W, Rosenfeld MG, Glass CK. 2014. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* **39**: 170–82.
- Lam PY, Kamei CN, Mangos S, Mudumana S, Liu Y, Drummond IA. 2013. Odd-skipped related 2 is required for fin chondrogenesis in zebrafish. *Dev Dyn* **242**: 1284–92.
- Lampe AK, Bushby KMD. 2005. Collagen VI related muscle disorders. *J Med Genet* **42**: 673–85.
- Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK, Chen S, Iwase S, Alpatov R, Issaeva I, Canaani E, et al. 2007. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* **449**: 689–94.
- Lan Y, Ovitt CE, Cho E-S, Maltby KM, Wang Q, Jiang R. 2004. Odd-skipped related 2 (Osr2) encodes a key intrinsic regulator of secondary palate growth and morphogenesis. *Development* **131**: 3207–16.
- Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**: 1813–31.
- Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, Jung S, Amit I. 2014. Tissue-resident

- macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**: 1312–26.
- Lawrence M, Daujat S, Schneider R. 2016. Lateral thinking: how histone modifications regulate gene expression. *Trends Genet* **32**: 42–56.
- Le Guen L, Marchal S, Faure S, de Santa Barbara P. 2015. Mesenchymal-epithelial interactions during digestive tract development and epithelial stem cell regeneration. *Cell Mol Life Sci* **72**: 3883–96.
- Lee TI, Johnstone SE, Young RA. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1**: 729–48.
- Leikina E, Defour A, Melikov K, Van der Meulen JH, Nagaraju K, Bhuvanendran S, Gebert C, Pfeifer K, Chernomordik L V, Jaiswal JK. 2015. Annexin A1 deficiency does not affect myofiber repair but delays regeneration of injured muscles. *Sci Rep* **5**: 18246.
- Lejard V, Blais F, Guerquin MJ, Bonnet A, Bonnin MA, Havis E, Malbouyres M, Bidaud CB, Maro G, Gilardi-Hebenstreit P, et al. 2011. EGR1 and EGR2 involvement in vertebrate tendon differentiation. *J Biol Chem* **286**: 5855–67.
- Léjard V, Brideau G, Blais F, Salingcarnboriboon R, Wagner G, Roehrl MHA, Noda M, Duprez D, Houillier P, Rossert J. 2007. Scleraxis and NFATc regulate the expression of the pro-alpha(I) collagen gene in tendon fibroblasts. *J Biol Chem* **282**: 17665–75.
- Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**: 1725–35.
- Lewandowski JP, Du F, Zhang S, Powell MB, Falkenstein KN, Ji H, Vokes SA. 2015. Spatiotemporal regulation of GLI target genes in the mammalian limb bud. *Dev Biol* **406**: 92–103.
- Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2009a. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **26**: 493–500.
- Li D, Sakuma R, Vakili NA, Mo R, Puviindran V, Deimling S, Zhang X, Hopyan S, Hui C chung. 2014. Formation of proximal and anterior limb skeleton requires early function of Irx3 and Irx5 and is negatively regulated by shh signaling. *Dev Cell* **29**: 233–40.
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**: 84–98.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–9.
- Li J, Gilmour DS. 2013. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J* **32**: 1829–41.
- Li Q, Brown JB, Huang H, Bickel PJ. 2011a. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* **5**: 1752–79.
- Li X-Y, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. 2011b. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biol* **12**: R34.
- Li Y, Qiu Q, Watson SS, Schweitzer R, Johnson RL. 2010. Uncoupling skeletal and connective tissue patterning: conditional deletion in cartilage progenitors reveals cell-autonomous requirements for Lmx1b in dorsal-ventral limb patterning. *Development* **137**: 1181–8.
- Liang XH, Deng WB, Li M, Zhao ZA, Wang TS, Feng XH, Cao YJ, Duan EK, Yang ZM. 2014. Egr1 protein acts downstream of estrogen-leukemia inhibitory factor (LIF)-STAT3 pathway and plays a role during implantation through targeting Wnt4. *J Biol Chem* **289**: 23534–45.
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–30.
- Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**: 251–5.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–93.
- Liu H, Xu J, Liu CF, Lan Y, Wylie C, Jiang R. 2015a. Whole transcriptome expression profiling of mouse limb tendon development by using RNA-seq. *J Orthop Res* **33**: 840–848.
- Liu H, Zhang C, Zhu S, Lu P, Zhu T, Gong X, Zhang Z, Hu J, Yin Z, Heng BC, et al. 2015b. Mohawk promotes the

- tenogenesis of mesenchymal stem cells through activation of the TGF β signaling pathway. *Stem Cells* **33**: 443–55.
- Liu L, Cheung TH, Charville GW, Hurgo BMC, Leavitt T, Shih J, Brunet A, Rando TA. 2013. Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Rep* **4**: 189–204.
- Liu W, Watson SS, Lan Y, Keene DR, Ovitt CE, Liu H, Schweitzer R, Jiang R. 2010. The atypical homeodomain transcription factor Mohawk controls tendon morphogenesis. *Mol Cell Biol* **30**: 4797–807.
- Logan C, Hornbruch A, Campbell I, Lumsden A. 1997. The role of Engrailed in establishing the dorsoventral axis of the chick limb. *Development* **124**: 2317–24.
- Logan M, Simon HG, Tabin C. 1998. Differential regulation of T-box and homeobox transcription factors suggests roles in controlling chick limb-type identity. *Development* **125**: 2825–35.
- Loganathan PG, Nimmagadda S, Huang R, Scaal M, Christ B. 2005. Comparative analysis of the expression patterns of Wnts during chick limb development. *Histochem Cell Biol* **123**: 195–201.
- Long F, Chung U, Ohba S, McMahon J, Kronenberg HM, McMahon AP. 2004. Ihh signaling is directly required for the osteoblast lineage in the endochondral skeleton. *Development* **131**: 1309–18.
- Loomis CA, Harris E, Michaud J, Wurst W, Hanks M, Joyner AL. 1996. The mouse Engrailed-1 gene and ventral limb patterning. *Nature* **382**: 360–3.
- Lopez-Rios J, Duchesne A, Speziale D, Andrey G, Peterson KA, Germann P, Ünal E, Liu J, Floriot S, Barbey S, et al. 2014. Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* **511**: 46–51.
- Lorda-Diez CI, Montero JA, Martinez-Cue C, Garcia-Porrero JA, Hurle JM. 2009. Transforming growth factors beta coordinate cartilage and tendon differentiation in the developing limb mesenchyme. *J Biol Chem* **284**: 29988–96.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.
- Lu X, Simon MD, Chodaparambil J V, Hansen JC, Shokat KM, Luger K. 2008. The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol* **15**: 1122–4.
- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–60.
- Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**: 1012–25.
- Luscombe NM, Austin SE, Berman HM, Thornton JM. 2000. An overview of the structures of protein-DNA complexes. *Genome Biol* **1**: REVIEWS001.
- Mäki JM. 2009. Lysyl oxidases in mammalian development and certain pathological conditions. *Histol Histopathol* **24**: 651–60.
- Malaval L, Aubin JE, Vico L. 2009. Role of the small integrin-binding ligand N-linked glycoprotein (SIBLING), bone sialoprotein (BSP) in bone development and remodeling. *Osteoporos Int* **20**: 1077–80.
- Martinet C, Monnier P, Louault Y, Benard M, Gabory A, Dandolo L. 2016. H19 controls reactivation of the imprinted gene network during muscle regeneration. *Development* **143**: 962–71.
- Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**: 29–59.
- Mathew SJ, Hansen JM, Merrell AJ, Murphy MM, Lawson J a, Hutcheson D a, Hansen MS, Angus-Hill M, Kardon G. 2011. Connective tissue fibroblasts and Tcf4 regulate myogenesis. *Development* **138**: 371–84.
- McConnell BB, Yang VW. 2010. Mammalian Krüppel-like factors in health and diseases. *Physiol Rev* **90**: 1337–81.
- McKinnell IW, Ishibashi J, Le Grand F, Punch VGJ, Addicks GC, Greenblatt JF, Dilworth FJ, Rudnicki MA. 2008. Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. *Nat Cell Biol* **10**: 77–84.
- McQueen HA, Siriaco G, Bird AP. 1998. Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res* **8**: 621–30.
- Mechem RP. 2012. Overview of extracellular matrix. *Curr Protoc Cell Biol* **Chapter 10**: Unit 10.1.
- Mediero A, Ramkhalawon B, Perez-Aso M, Moore KJ, Cronstein BN. 2015. Netrin-1 is a critical autocrine/paracrine factor for osteoclast differentiation. *J Bone Miner Res* **30**: 837–54.
- Meng X, Brodsky MH, Wolfe S a. 2005. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol* **23**: 988–94.
- Mennerich D, Schäfer K, Braun T. 1998. Pax-3 is necessary but not sufficient for lbx1 expression in myogenic precursor cells of the limb. *Mech Dev* **73**: 147–58.
- Merino R, Rodriguez-Leon J, Macias D, Gañan Y, Economides AN, Hurle JM. 1999. The BMP antagonist Gremlin

- regulates outgrowth, chondrogenesis and programmed cell death in the developing limb. *Development* **126**: 5515–22.
- Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. 2010. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* **38**: D204–10.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553–60.
- Millay DP, O'Rourke JR, Sutherland LB, Bezprozvannaya S, Shelton JM, Bassel-Duby R, Olson EN. 2013. Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* **499**: 301–5.
- Minina E, Kreschel C, Naski MC, Ornitz DM, Vortkamp A. 2002. Interaction of FGF, Ihh/Pthlh, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation. *Dev Cell* **3**: 439–49.
- modENCODE Consortium, Roy S, Ernst J, Kharchenko P V, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, et al. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**: 1787–97.
- Moorman C, Sun L V, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu X-J, White KP, Bussemaker HJ, et al. 2006. Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **103**: 12027–32.
- Morgan B a, Fekete DM. 1996. Manipulating gene expression with replication-competent retroviruses. *Methods Cell Biol* **51**: 185–218.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621–8.
- Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, Sartorelli V. 2013. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* **51**: 606–17.
- Mugford JW, Sipilä P, McMahon JA, McMahon AP. 2008. *Osrl* expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an *Osrl*-dependent nephron progenitor compartment within the mammalian kidney. *Dev Biol* **324**: 88–98.
- Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**: 197–208.
- Murakami S, Ohki-Hamazaki H, Watanabe K, Ikenaka K, Ono K. 2010. Netrin 1 provides a chemoattractive cue for the ventral migration of GnRH neurons in the chick forebrain. *J Comp Neurol* **518**: 2019–34.
- Murchison ND, Price BA, Conner DA, Keene DR, Olson EN, Tabin CJ, Schweitzer R. 2007. Regulation of tendon differentiation by scleraxis distinguishes force-transmitting tendons from muscle-anchoring tendons. *Development* **134**: 2697–708.
- Murphy MM, Lawson J a, Mathew SJ, Hutcheson D a, Kardon G. 2011. Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. *Development* **138**: 3625–37.
- Nacu E, Glausch M, Le HQ, Damanik FFR, Schuez M, Knapp D, Khattak S, Richter T, Tanaka EM. 2013. Connective tissue cells, but not muscle cells, are involved in establishing the proximo-distal outcome of limb regeneration in the axolotl. *Development* **140**: 513–8.
- Naiche LA, Papaioannou VE. 2007. *Tbx4* is not required for hindlimb identity or post-bud hindlimb outgrowth. *Development* **134**: 93–103.
- Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugge B. 2002. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108**: 17–29.
- Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny L a, Dekker J. 2013. Organization of the mitotic chromosome. *Science* **342**: 948–53.
- Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grützner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**: 635–40.
- Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, et al. 2011. A cis-regulatory map of the *Drosophila* genome. *Nature* **471**: 527–31.
- Noma K, Allis CD, Grewal SI. 2001. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**: 1150–5.
- Oberlender SA, Tuan RS. 1994. Expression and functional involvement of N-cadherin in embryonic limb chondrogenesis. *Development* **120**: 177–87.
- Odelberg SJ, Kollhoff A, Keating MT. 2000. Dedifferentiation of mammalian myotubes induced by *msx1*. *Cell* **103**: 1099–109.

- Ohno S. 1972. So much “junk” DNA in our genome. *Brookhaven Symp Biol* **23**: 366–70.
- Ojeda JL, Barbosa E, Bosque PG. 1970. Selective skeletal staining in whole chicken embryos; a rapid Alcian blue technique. *Stain Technol* **45**: 137–8.
- Olson EN. 1992. Interplay between proliferation and differentiation within the myogenic lineage. *Dev Biol* **154**: 261–72.
- Otabe K, Nakahara H, Hasegawa A, Matsukawa T, Ayabe F, Onizuka N, Inui M, Takada S, Ito Y, Sekiya I, et al. 2015. Transcription factor Mohawk controls tenogenic differentiation of bone marrow mesenchymal stem cells in vitro and in vivo. *J Orthop Res* **33**: 1–8.
- Otani K, Dong Y, Li X, Lu J, Zhang N, Xu L, Go MYY, Ng EKW, Arakawa T, Chan FKL, et al. 2014. Odd-skipped related 1 is a novel tumour suppressor gene and a potential prognostic biomarker in gastric cancer. *J Pathol* **234**: 302–15.
- Pagel J-I, Deindl E. 2011. Early growth response 1--a transcription factor in the crossfire of signal transduction cascades. *Indian J Biochem Biophys* **48**: 226–35.
- Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA. 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1**: 299–312.
- Panne D, Maniatis T, Harrison SC. 2007. An atomic model of the interferon-beta enhanceosome. *Cell* **129**: 1111–23.
- Parang B, Rosenblatt D, Williams AD, Washington MK, Revetta F, Short SP, Reddy VK, Hunt A, Shroyer NF, Engel ME, et al. 2015. The transcriptional corepressor MTGR1 regulates intestinal secretory lineage allocation. *FASEB J* **29**: 786–95.
- Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. *PLoS Genet* **9**: e1003748.
- Park CS, Shen Y, Lewis A, Lacorazza HD. 2016. Role of the reprogramming factor KLF4 in blood formation. *J Leukoc Biol* **99**: 673–85.
- Parr B a, Shea MJ, Vassileva G, McMahon a P. 1993. Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**: 247–61.
- Parslow TG, Jones SD, Bond B, Yamamoto KR. 1987. The immunoglobulin octanucleotide: independent activity and selective interaction with enhancers. *Science* **235**: 1498–501.
- Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau J-C, Ferrier P, Spicuglia S. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J* **30**: 4198–210.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**: 785–6.
- Pizette S, Niswander L. 2001. Early steps in limb patterning and chondrogenesis. *Novartis Found Symp* **232**: 23–36; discussion 36–46.
- Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**: 517–27.
- Porrua O, Libri D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**: 190–202.
- Pourquié O, Fan CM, Coltey M, Hirsinger E, Watanabe Y, Bréant C, Francis-West P, Brickell P, Tessier-Lavigne M, Le Douarin NM. 1996. Lateral and axial signals involved in avian somite patterning: a role for BMP4. *Cell* **84**: 461–71.
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**: 1851–4.
- Pryce BA, Watson SS, Murchison ND, Staverosky JA, Dünker N, Schweitzer R. 2009. Recruitment and maintenance of tendon progenitors by TGFbeta signaling are essential for tendon formation. *Development* **136**: 1351–61.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res* **40**: D290–301.
- Querfurth E, Schuster M, Kulesa H, Crispino JD, Döderlein G, Orkin SH, Graf T, Nerlov C. 2000. Antagonism between C/EBPbeta and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev* **14**: 2515–25.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–2.

- Rabinowitz AH, Vokes SA. 2012. Integration of the transcriptional networks regulating limb morphogenesis. *Dev Biol* **368**: 165–180.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–83.
- Rallis C, Bruneau BG, Del Buono J, Seidman CE, Seidman JG, Nissim S, Tabin CJ, Logan MPO. 2003. Tbx5 is required for forelimb bud formation and continued outgrowth. *Development* **130**: 2741–51.
- Rankin SA, Gallas AL, Neto A, Gómez-Skarmeta JL, Zorn AM. 2012. Suppression of Bmp4 signaling by the zinc-finger repressors Osr1 and Osr2 is required for Wnt/ β -catenin-mediated lung specification in *Xenopus*. *Development* **139**: 3010–20.
- Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–80.
- Relaix F, Rocancourt D, Mansouri A, Buckingham M. 2005. A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature* **435**: 948–53.
- Riddle RD, Johnson RL, Laufer E, Tabin C. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**: 1401–16.
- Riffo-Campos ÁL, Castillo J, Tur G, González-Figueroa P, Georgieva EI, Rodríguez JL, López-Rodas G, Rodrigo MI, Franco L. 2015. Nucleosome-specific, time-dependent changes in histone modifications during activation of the early growth response 1 (Egr1) gene. *J Biol Chem* **290**: 197–208.
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–30.
- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**: 651–7.
- Robson LG, Hughes SM. 1996. The distal limb environment regulates MyoD accumulation and muscle differentiation in mouse-chick chimaeric limbs. *Development* **122**: 3899–910.
- Robson LG, Kara T, Crawley a, Tickle C. 1994. Tissue and cellular patterning of the musculature in chick wings. *Development* **120**: 1265–76.
- Rong PM, Teillet MA, Ziller C, Le Douarin NM. 1992. The neural tube/notochord complex is necessary for vertebral but not limb and body wall striated muscle differentiation. *Development* **115**: 657–72.
- Rosenfeld JA, Wang Z, Schones DE, Zhao K, DeSalle R, Zhang MQ. 2009. Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomics* **10**: 143.
- Rot C, Stern T, Blecher R, Friesem B, Zelzer E. 2014. A mechanical Jack-like mechanism drives spontaneous fracture healing in neonatal mice. *Dev Cell* **31**: 159–70.
- Rozario T, DeSimone DW. 2010. The extracellular matrix in development and morphogenesis: A dynamic view. *Dev Biol* **341**: 126–40.
- Rudnicki JA, Brown AM. 1997. Inhibition of chondrogenesis by Wnt gene expression in vivo and in vitro. *Dev Biol* **185**: 104–18.
- Saccone V, Consalvi S, Giordani L, Mozzetta C, Barozzi I, Sandoná M, Ryan T, Rojas-Muñoz A, Madaro L, Fasanaro P, et al. 2014. HDAC-regulated myomiRs control BAF60 variant exchange and direct the functional phenotype of fibro-adipogenic progenitors in dystrophic muscles. *Genes Dev* **28**: 841–57.
- Sagai T, Hosoya M, Mizushima Y, Tamura M, Shiroishi T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**: 797–803.
- Sainsbury S, Bernecky C, Cramer P. 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* **16**: 129–43.
- Sambrook J, Russel DW. 2001. *Molecular cloning: a laboratory manual*. Third edit. Cold Spring Harbor Laboratory Press.
- Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, Schreiber SL, Mellor J, Kouzarides T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**: 407–11.
- Saunders JW. 1948. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J Exp Zool* **108**: 363–403.
- Schäfer K, Braun T. 1999. Early specification of limb muscle precursor cells by the homeobox gene Lbx1h. *Nat Genet* **23**: 213–6.
- Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, et al. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor

- binding. *Science* **328**: 1036–40.
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**: 671–5.
- Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat Cell Biol* **6**: 73–7.
- Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, et al. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev* **18**: 1263–71.
- Schwanhäusser B, Busse D, Li N. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337–42.
- Schweitzer R, Chyung JH, Murtaugh LC, Brent a E, Rosen V, Olson EN, Lassar A, Tabin CJ. 2001. Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments. *Development* **128**: 3855–66.
- Schweitzer R, Zelzer E, Volk T. 2010. Connecting muscles to tendons: tendons and musculoskeletal development in flies and vertebrates. *Development* **137**: 2807–17.
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent transcription from active promoters. *Science* **322**: 1849–51.
- Sen Banerjee S, Feinberg MW, Watanabe M, Gray S, Haspel RL, Denking DJ, Kawahara R, Hauner H, Jain MK. 2003. The Krüppel-like factor KLF2 inhibits peroxisome proliferator-activated receptor- γ expression and adipogenesis. *J Biol Chem* **278**: 2581–4.
- Serafini T, Colamarino SA, Leonardo ED, Wang H, Beddington R, Skarnes WC, Tessier-Lavigne M. 1996. Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**: 1001–14.
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**: 37–46.
- Sharir A, Stern T, Rot C, Shahar R, Zelzer E. 2011. Muscle force regulates bone shaping for optimal load-bearing capacity during embryogenesis. *Development* **138**: 3247–59.
- Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov V V, et al. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**: 116–20.
- Shimizu H, Yokoyama S, Asahara H. 2007. Growth and differentiation of the developing limb bud from the perspective of chondrogenesis. *Dev Growth Differ* **49**: 449–54.
- Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844–7.
- Siggers T, Duyzend MH, Reddy J, Khan S, Bulyk ML. 2011. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol Syst Biol* **7**: 555.
- Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**: 978–81.
- Simionescu-Bankston A, Kumar A. 2016. Noncoding RNAs in the regulation of skeletal muscle biology in health and disease. *J Mol Med (Berl)* [Epub ahead of print].
- Simons EV, van Horn JR. 1971. A new procedure for whole-mount alcian blue staining of the cartilaginous skeleton of chicken embryos, adapted to the clearing procedure in potassium hydroxide. *Acta Morphol Neerl Scand* **8**: 281–92.
- Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, et al. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**: 1270–82.
- Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D, Morrice DR, Law AS, Masabanda J, Sazanov A, et al. 2000. Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet* **31**: 96–103.
- Soler C, Daczewska M, Da Ponte JP, Dastugue B, Jagla K. 2004. Coordinated development of muscles and tendons of the *Drosophila* leg. *Development* **131**: 6041–51.
- Soler C, Laddada L, Jagla K. 2016. Coordinated development of muscles and tendon-like structures: early interactions in the *drosophila* leg. *Front Physiol* **7**: 22.
- Solursh M, Ahrens PB, Reiter RS. 1978. A tissue culture analysis of the steps in limb chondrogenesis. *In Vitro* **14**: 51–61.
- Solursh M, Drake C, Meier S. 1987. The migration of myogenic cells from the somites at the wing level in avian embryos. *Dev Biol* **121**: 389–96.
- Soufi A, Donahue G, Zaret KS. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**: 994–1004.

- Später D, Hill TP, O'sullivan RJ, Gruber M, Conner DA, Hartmann C. 2006. Wnt9a signaling is required for joint integrity and regulation of Ihh during chondrogenesis. *Development* **133**: 3039–49.
- Spitz F, Furlong EE. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613–26.
- Srinivasan K, Strickland P, Valdes A, Shin GC, Hinck L. 2003. Netrin-1/Neogenin interaction stabilizes multipotent progenitor cap cells during mammary gland morphogenesis. *Dev Cell* **4**: 371–82.
- Stampfel G, Kazmar T, Frank O, Wienerroither S, Reiter F, Stark A. 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**: 147–51.
- Stefflova K, Thybert D, Wilson MD, Streeter I, Aleksic J, Karagianni P, Brazma A, Adams DJ, Talianidis I, Marioni JC, et al. 2013. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* **154**: 530–40.
- Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim J-E, Chen J, Lazar MA, et al. 2008. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol* **28**: 2825–39.
- Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41–5.
- Stricker S, Brieske N, Haupt J, Mundlos S. 2006. Comparative expression pattern of Odd-skipped related genes *Osr1* and *Osr2* in chick embryonic development. *Gene Expr Patterns* **6**: 826–34.
- Stricker S, Fundele R, Vortkamp A, Mundlos S. 2002. Role of Runx genes in chondrocyte differentiation. *Dev Biol* **245**: 95–108.
- Stricker S, Mathia S, Haupt J, Seemann P, Meier J, Mundlos S. 2012. Odd-skipped related genes regulate differentiation of embryonic limb mesenchyme and bone marrow mesenchymal stromal cells. *Stem Cells Dev* **21**: 623–33.
- Stumm J. 2016. The role of *Osr1* in postnatal muscle development and muscle regeneration. PhD Thesis, Freie Universität Berlin.
- Sugimoto Y, Takimoto A, Akiyama H, Kist R, Scherer G, Nakamura T, Hiraki Y, Shukunami C. 2013a. *Scx*+/*Sox9*+ progenitors contribute to the establishment of the junction between cartilage and tendon/ligament. *Development* **140**: 2280–8.
- Sugimoto Y, Takimoto A, Hiraki Y, Shukunami C. 2013b. Generation and characterization of *Scx*Cre transgenic mice. *Genesis* **51**: 275–83.
- Sunadome K, Yamamoto T, Ebisuya M, Kondoh K, Sehara-Fujisawa A, Nishida E. 2011. ERK5 regulates muscle cell fusion through *Klf* transcription factors. *Dev Cell* **20**: 192–205.
- Suva LJ, Ernst M, Rodan GA. 1991. Retinoic acid increases *zif268* early gene expression in rat preosteoblastic cells. *Mol Cell Biol* **11**: 2503–10.
- Swartz ME, Eberhart J, Pasquale EB, Krull CE. 2001. EphA4/ephrin-A5 interactions in muscle precursor cell migration in the avian forelimb. *Development* **128**: 4669–80.
- Swinehart IT, Schlientz AJ, Quintanilla CA, Mortlock DP, Wellik DM. 2013. *Hox11* genes are required for regional patterning and integration of muscle, tendon and bone. *Development* **140**: 4574–82.
- Tabin C, Wolpert L. 2007. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev* **21**: 1433–42.
- Tabin CJ, McMahon AP. 2008. Developmental biology: grasping limb patterning. *Science* **321**: 350–2.
- 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**: 4155–62.
- Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**: 663–76.
- Takata H, Terada K, Oka H, Sunada Y, Moriguchi T, Nohno T. 2007. Involvement of Wnt4 signaling during myogenic proliferation and differentiation of skeletal muscle. *Dev Dyn* **236**: 2800–7.
- ten Berge D, Brugmann SA, Helms JA, Nusse R. 2008. Wnt and FGF signals interact to coordinate growth with cell fate specification during limb development. *Development* **135**: 3247–57.
- Tepluyuk NM, Galindo M, Tepluyuk VI, Pratap J, Young DW, Lapointe D, Javed A, Stein JL, Lian JB, Stein GS, et al. 2008. Runx2 regulates G protein-coupled signaling pathways to control growth of osteoblast progenitors. *J Biol Chem* **283**: 27585–97.
- Tessarz P, Kouzarides T. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* **15**: 703–8.
- Thanos D, Maniatis T. 1995. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83**: 1091–100.

- Thurman R, Rynes E, Humbert R, Vierstra J, Maurano M, Haugen E, Sheffield N, Stergachis A, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.
- Tidball JG, Lin C. 1989. Structural changes at the myogenic cell surface during the formation of myotendinous junctions. *Cell Tissue Res* **257**: 77–84.
- Tozer S, Bonnin M-A, Relaix F, Di Savino S, García-Villalba P, Coumailleau P, Duprez D. 2007. Involvement of vessels and PDGFB in muscle splitting during chick limb development. *Development* **134**: 2579–91.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–5.
- Tsang AP, Visvader JE, Turner CA, Fujiwara Y, Yu C, Weiss MJ, Crossley M, Orkin SH. 1997. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90**: 109–19.
- Tschopp P, Duboule D. 2014. The genetics of murine Hox loci: TAMERE, STRING, and PANTHERE to engineer chromosome variants. *Methods Mol Biol* **1196**: 89–102.
- Tu S, Narendra V, Yamaji M, Vidal SE, Rojas LA, Wang X, Kim SY, Garcia BA, Tuschl T, Stadtfeld M, et al. 2016. Co-repressor CBFA2T2 regulates pluripotency and germline development. *Nature* **534**: 387–90.
- Tufan AC, Tuan RS. 2001. Wnt regulation of limb mesenchymal chondrogenesis is accompanied by altered N-cadherin-related functions. *FASEB J* **15**: 1436–8.
- Uysal-Onganer P, Kypta RM. 2012. Wnt11 in 2011 - the regulation and function of a non-canonical Wnt. *Acta Physiol* **204**: 52–64.
- Vallecillo García P, Orgeur M, Stumm J, Poehle-Kronawitter S, Keppert V, Börno ST, Shinichiro H, Relaix F, Koch M, Timmermann B, et al. Odd skipped-related 1 (Osr1) identifies the fibro-adipogenic progenitor (FAP) lineage and regulates a pro-myogenic transcriptional program during limb development. *Nat Commun* [In revision].
- Vasyutina E, Stebler J, Brand-Saberi B, Schulz S, Raz E, Birchmeier C. 2005. CXCR4 and Gab1 cooperate to control the development of migrating muscle progenitor cells. *Genes Dev* **19**: 2187–98.
- Veitia RA, Bottani S, Birchler JA. 2013. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet* **29**: 385–93.
- Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, Ljungman M. 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res* **24**: 896–905.
- Vettese-Dadey M, Grant PA, Hebbes TR, Crane- Robinson C, Allis CD, Workman JL. 1996. Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J* **15**: 2508–18.
- Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* **160**: 554–66.
- Villar D, Flicek P, Odom DT. 2014. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat Rev Genet* **15**: 221–33.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama J a, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–8.
- Vortkamp A, Lee K, Lanske B, Segre G V, Kronenberg HM, Tabin CJ. 1996. Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* **273**: 613–22.
- Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**: 281–5.
- Wall ME, Hlavacek WS, Savageau MA. 2004. Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **5**: 34–42.
- Wan Y, Lewis AK, Colasanto M, van Langeveld M, Kardon G, Hansen C. 2012. A practical workflow for making anatomical atlases for biological research. *IEEE Comput Graph Appl* **32**: 70–80.
- Wang H, Noulet F, Edom-Vovard F, Le Grand F, Duprez D. 2010. Bmp signaling at the tips of skeletal muscles regulates the number of fetal muscle progenitors and satellite cells during development. *Dev Cell* **18**: 643–54.
- Wang J, Kumar RM, Biggs VJ, Lee H, Chen Y, Kagey MH, Young RA, Abate-Shen C. 2011a. The Msx1 homeoprotein recruits polycomb to the nuclear periphery during development. *Dev Cell* **21**: 575–88.
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011b. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**: 120–4.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh T-Y, Peng W, Zhang MQ, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**:

- Widelitz RB, Jiang TX, Murray B a, Chuong CM. 1993. Adhesion molecules in skeletogenesis: II. Neural cell adhesion molecules mediate precartilaginous mesenchymal condensations and enhance chondrogenesis. *J Cell Physiol* **156**: 399–411.
- Wienert B, Funnell APW, Norton LJ, Pearson RCM, Wilkinson-White LE, Lester K, Vadolas J, Porteus MH, Matthews JM, Quinlan KGR, et al. 2015. Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. *Nat Commun* **6**: 7085.
- Witte F, Dokas J, Neuendorf F, Mundlos S, Stricker S. 2009. Comprehensive expression analysis of all Wnt genes and their major secreted antagonists during mouse limb development and cartilage differentiation. *Gene Expr Patterns* **9**: 215–23.
- Wong ES, Thybert D, Schmitt BM, Stefflova K, Odom DT, Flicek P. 2015. Decoupling of evolutionary changes in transcription factor binding and gene expression in mammals. *Genome Res* **25**: 167–78.
- Wu J, Srinivasan S V., Neumann JC, Lingrel JB. 2005. The KLF2 transcription factor does not affect the formation of preadipocytes but inhibits their differentiation into adipocytes. *Biochemistry* **44**: 11098–105.
- Wunderle VM, Critcher R, Hastie N, Goodfellow PN, Schedl A. 1998. Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci U S A* **95**: 10649–54.
- Wunderlich Z, Mirny LA. 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet* **25**: 434–40.
- Xu J, Liu H, Park J-S, Lan Y, Jiang R. 2014. *Osr1* acts downstream of and interacts synergistically with *Six2* to maintain nephron progenitor cells during kidney organogenesis. *Development* **141**: 1442–52.
- Yamashita R, Suzuki Y, Sugano S, Nakai K. 2005. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* **350**: 129–36.
- Yebra M, Montgomery AMP, Diaferia GR, Kaido T, Silletti S, Perez B, Just ML, Hildbrand S, Hurford R, Florkiewicz E, et al. 2003. Recognition of the neural chemoattractant netrin-1 by integrins $\alpha6\beta4$ and $\alpha3\beta1$ regulates epithelial cell adhesion and migration. *Dev Cell* **5**: 695–707.
- Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder M, et al. 2012. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol* **13**: R48.
- Yoon BS, Ovchinnikov DA, Yoshii I, Mishina Y, Behringer RR, Lyons KM. 2005. *Bmpr1a* and *Bmpr1b* have overlapping functions and are essential for chondrogenesis in vivo. *Proc Natl Acad Sci U S A* **102**: 5062–7.
- Young RS, Hayashizaki Y, Andersson R, Sandelin A, Kawaji H, Itoh M, Lassmann T, Carninci P, Bickmore WA, Forrest AR, et al. 2015. The frequent evolutionary birth and death of functional promoters in mouse and human. *Genome Res* **25**: 1546–57.
- Zakany J, Duboule D. 2007. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* **17**: 359–66.
- Zakany J, Kmita M, Duboule D. 2004. A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* **304**: 1669–72.
- Zalc A, Hayashi S, Auradé F, Bröhl D, Chang T, Mademtzoglou D, Mourikis P, Yao Z, Cao Y, Birchmeier C, et al. 2014. Antagonistic regulation of *p57kip2* by *Hes/Hey* downstream of Notch signaling and muscle regulatory factors regulates skeletal muscle growth arrest. *Development* **141**: 2780–90.
- Zhang D, Schwarz EM, Rosier RN, Zuscik MJ, Puzas JE, O’Keefe RJ. 2003. ALK2 functions as a BMP type I receptor and induces Indian hedgehog in chondrocytes during skeletal development. *J Bone Miner Res* **18**: 1593–604.
- Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**: 1311–20.
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.
- Zhang Z, Lan Y, Chai Y, Jiang R. 2009. Antagonistic actions of *Msx1* and *Osr2* pattern mammalian teeth into a single row. *Science* **323**: 1232–4.
- Zhao Q, Eberspaecher H, Lefebvre V, De Crombrughe B. 1997. Parallel expression of *Sox9* and *Col2a1* in cells undergoing chondrogenesis. *Dev Dyn* **209**: 377–86.
- Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, Orlov YL, Sung W-K, Shahab A, Kuznetsov VA, et al. 2007. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**: 286–98.
- Zimmerman LB, De Jesús-Escobar JM, Harland RM. 1996. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**: 599–606.
- Zou H, Wieser R, Massagué J, Niswander L. 1997. Distinct roles of type I bone morphogenetic protein receptors in

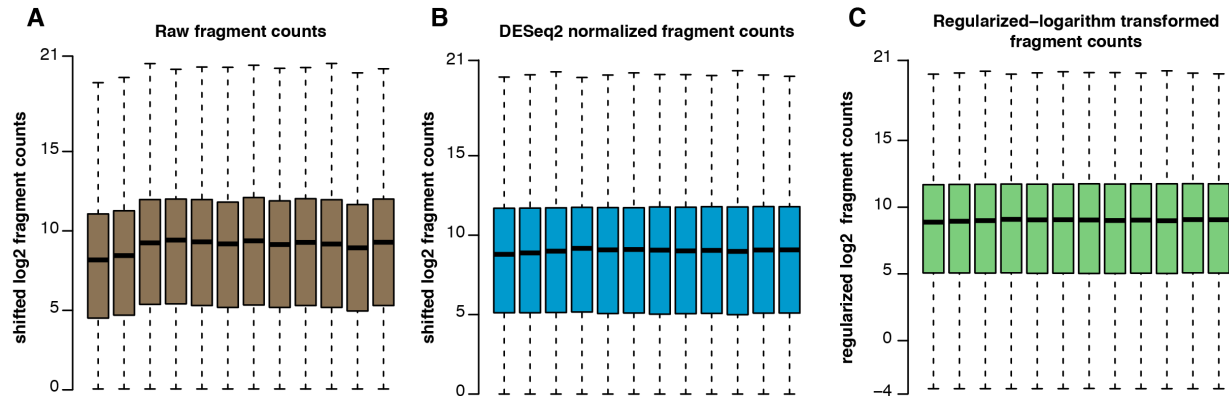
the formation and differentiation of cartilage. *Genes Dev* **11**: 2191–203.

Zuniga A. 2015. Next generation limb development and evolution: old questions, new perspectives. *Development* **142**: 3810–20.

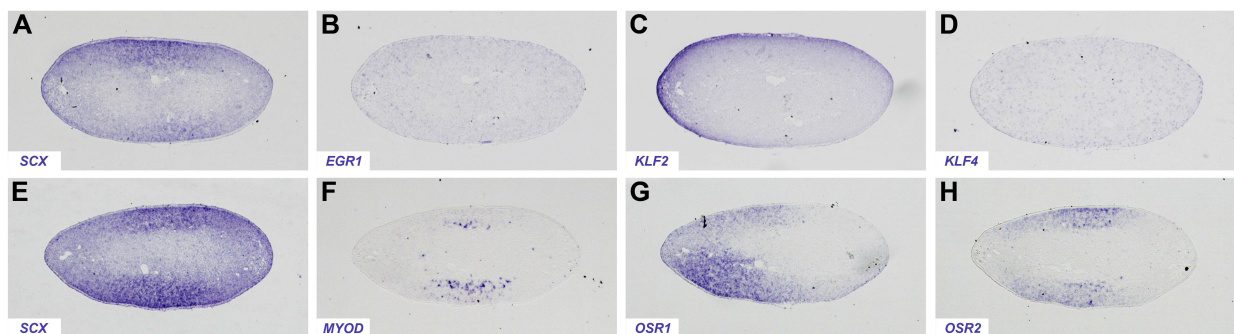
Zuniga A, Haramis AP, McMahon AP, Zeller R. 1999. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**: 598–602.

SUPPLEMENTAL INFORMATION

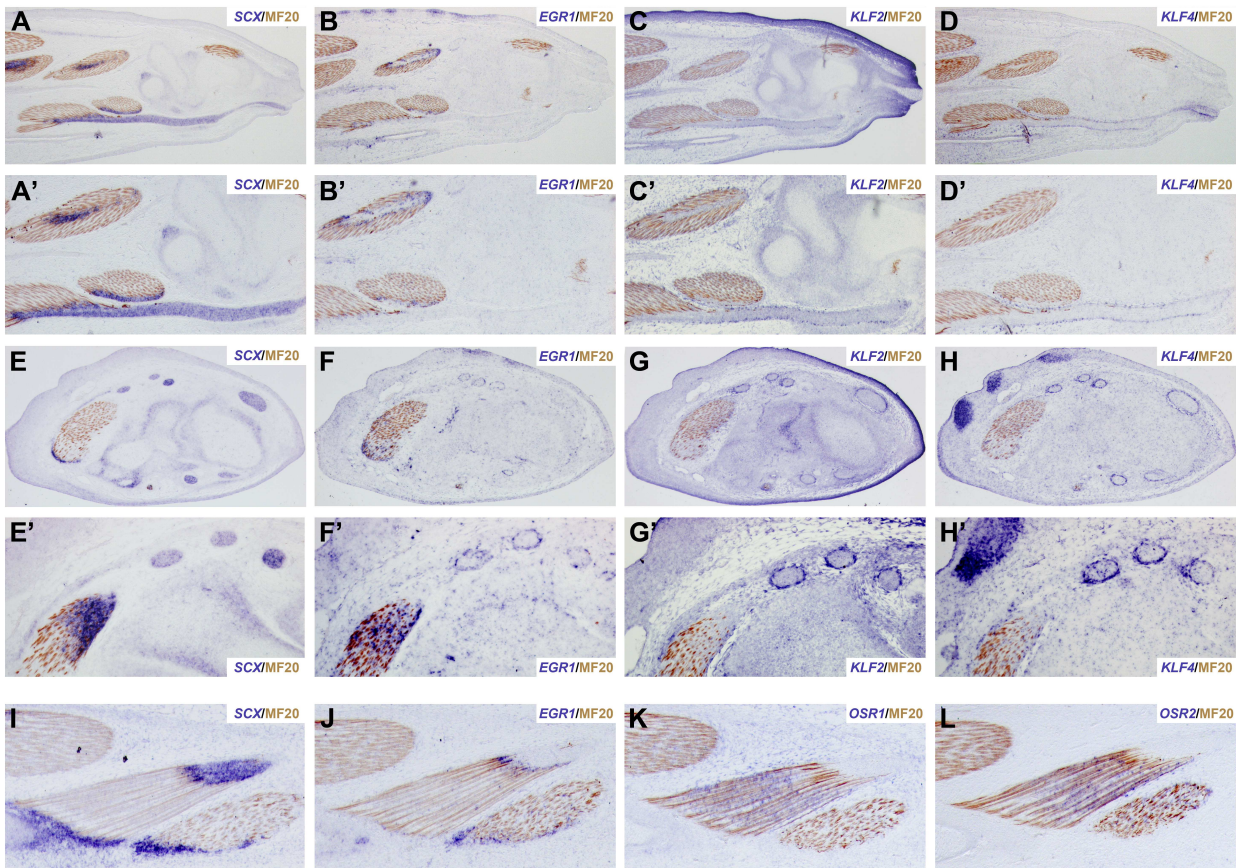
Supplemental figures



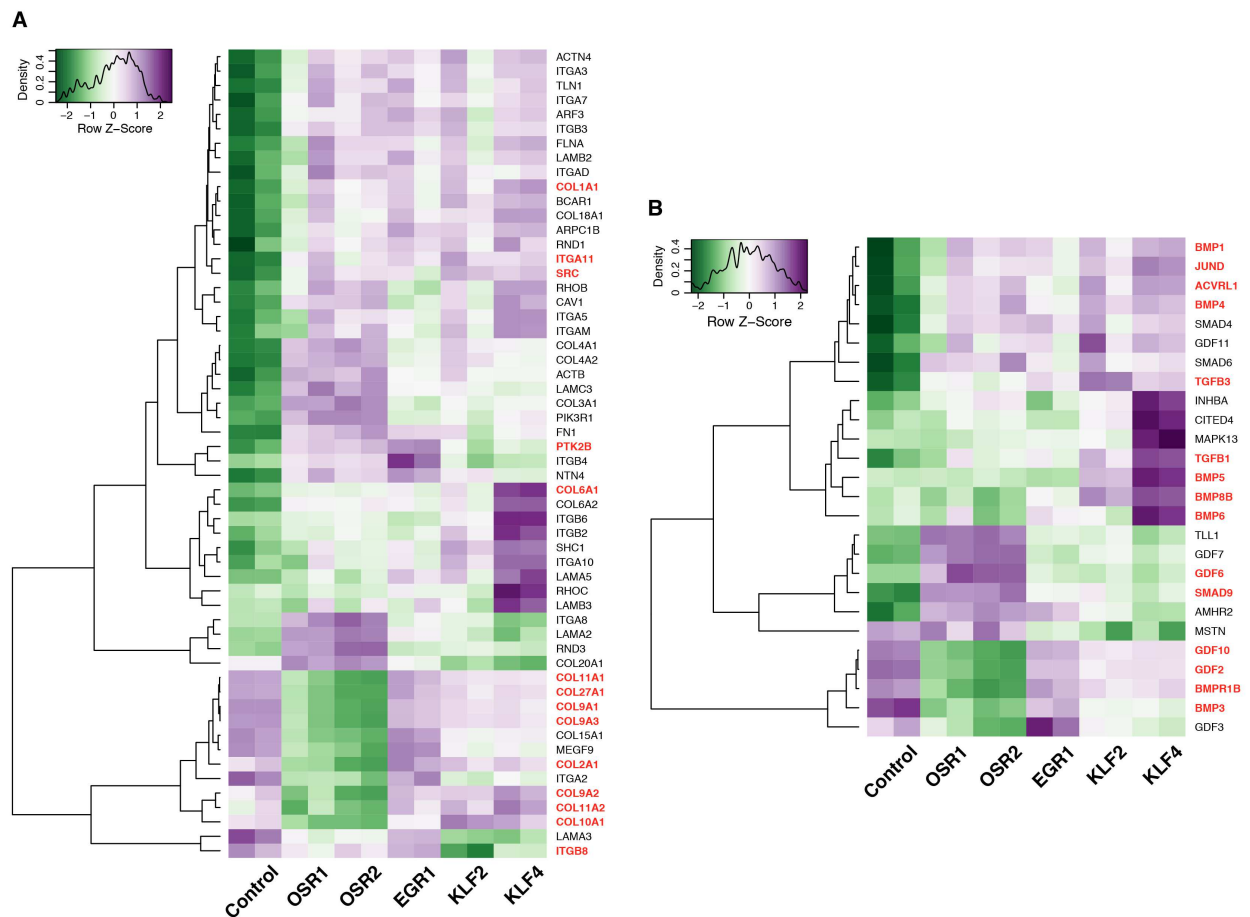
Supplemental figure S1. RNA-seq fragment count normalization. Distribution of raw (A), DESeq2 normalized (B) and regularized-logarithm transformed (C) fragment counts across all samples. The normalization factor computed by DESeq2 enables to correct library size and RNA composition bias, while the regularized-logarithm transformation shrinks together genes with low fragment counts.



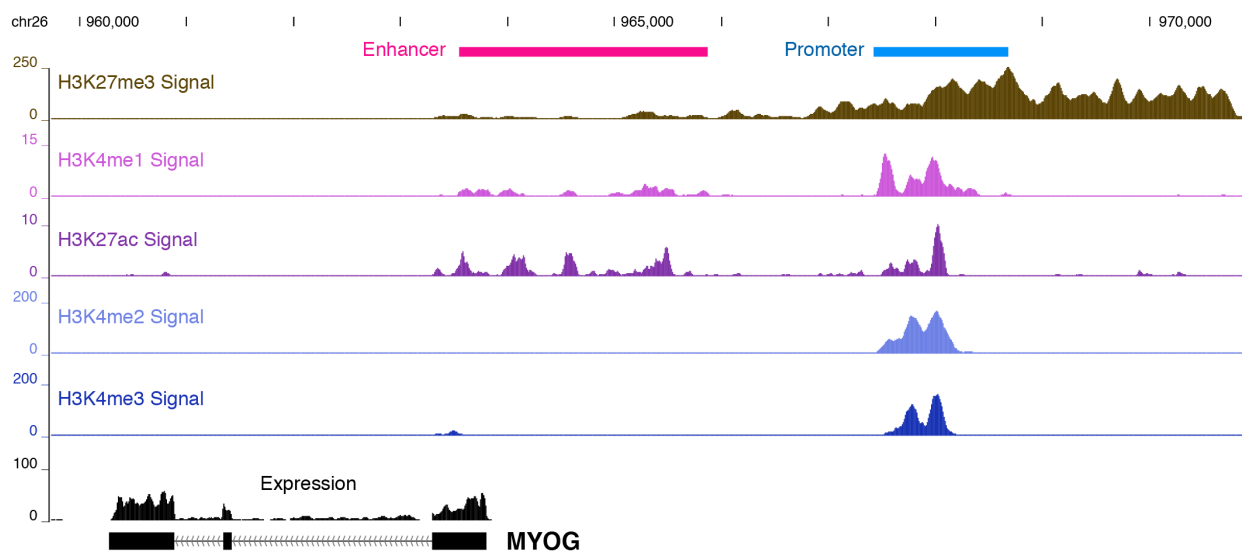
Supplemental figure S2. Endogenous expression of the transcription factors in forelimbs of E4.5 chick embryos. ISH for *SCX* (A,E), *EGR1* (B), *KLF2* (C), *KLF4* (D), *MYOD* (F), *OSR1* (G) and *OSR2* (H) genes in chick forelimbs of E4.5 (HH24-25) chick embryos. (A-D) Adjacent and transverse sections were hybridized with *SCX*, *EGR1*, *KLF2* and *KLF4* probes (blue). Magnification 5X. (E-H) Adjacent and transverse sections were hybridized with *SCX*, *MYOD*, *OSR1* and *OSR2* probes (blue). Magnification 5X. Top dorsal; left, posterior. *SCX* is a tendon-specific marker; *MYOD* is a muscle-specific marker.



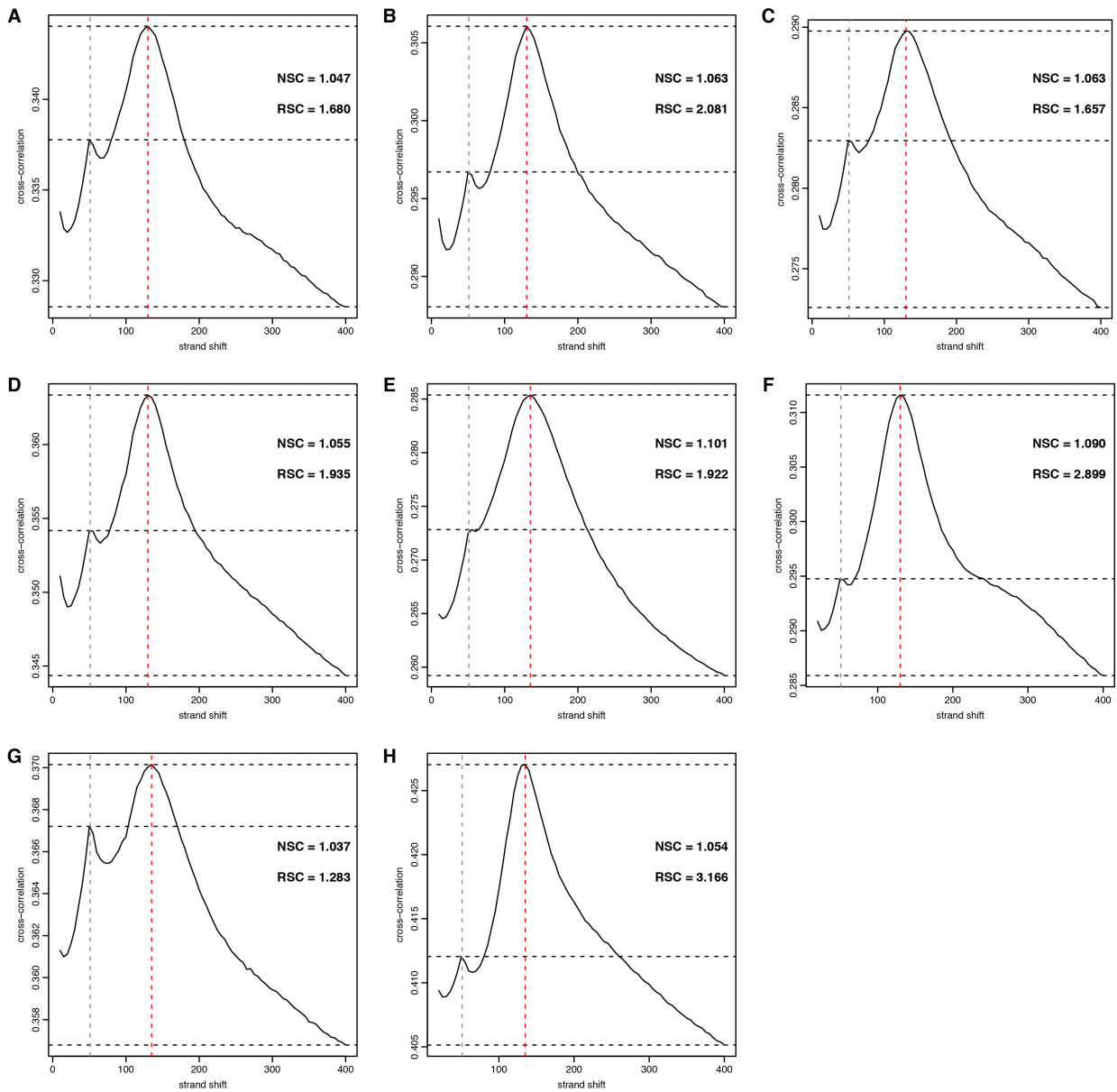
Supplemental figure S3. Endogenous expression of the transcription factors in forelimbs of E9.5 chick embryos. ISH for *SCX* (A,A',E,E',I), *EGR1* (B,B',F,F',J), *KLF2* (C,C',G,G'), *KLF4* (D,D',H,H'), *OSR1* (K) and *OSR2* (L) genes in forelimbs of E9.5 (HH35-36) chick embryos. (A-D) Adjacent and longitudinal sections were hybridized with *SCX*, *EGR1*, *KLF2* and *KLF4* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 2.5X at the zeugopod level. Top, anterior; left, proximal. (A'-D') Magnification 5X of posterior regions of sections (A-D). (E-H) Adjacent and transverse sections were hybridized with *SCX*, *EGR1*, *KLF2* and *KLF4* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the wrist level. Top, posterior; left, dorsal. (E'-H') Magnification 10X of posterior regions of sections (E-H). (I-L) Adjacent and longitudinal sections were hybridized with *SCX*, *EGR1*, *OSR1* and *OSR2* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X. *SCX* is a tendon-specific marker; MF20 targets skeletal muscle myosin.



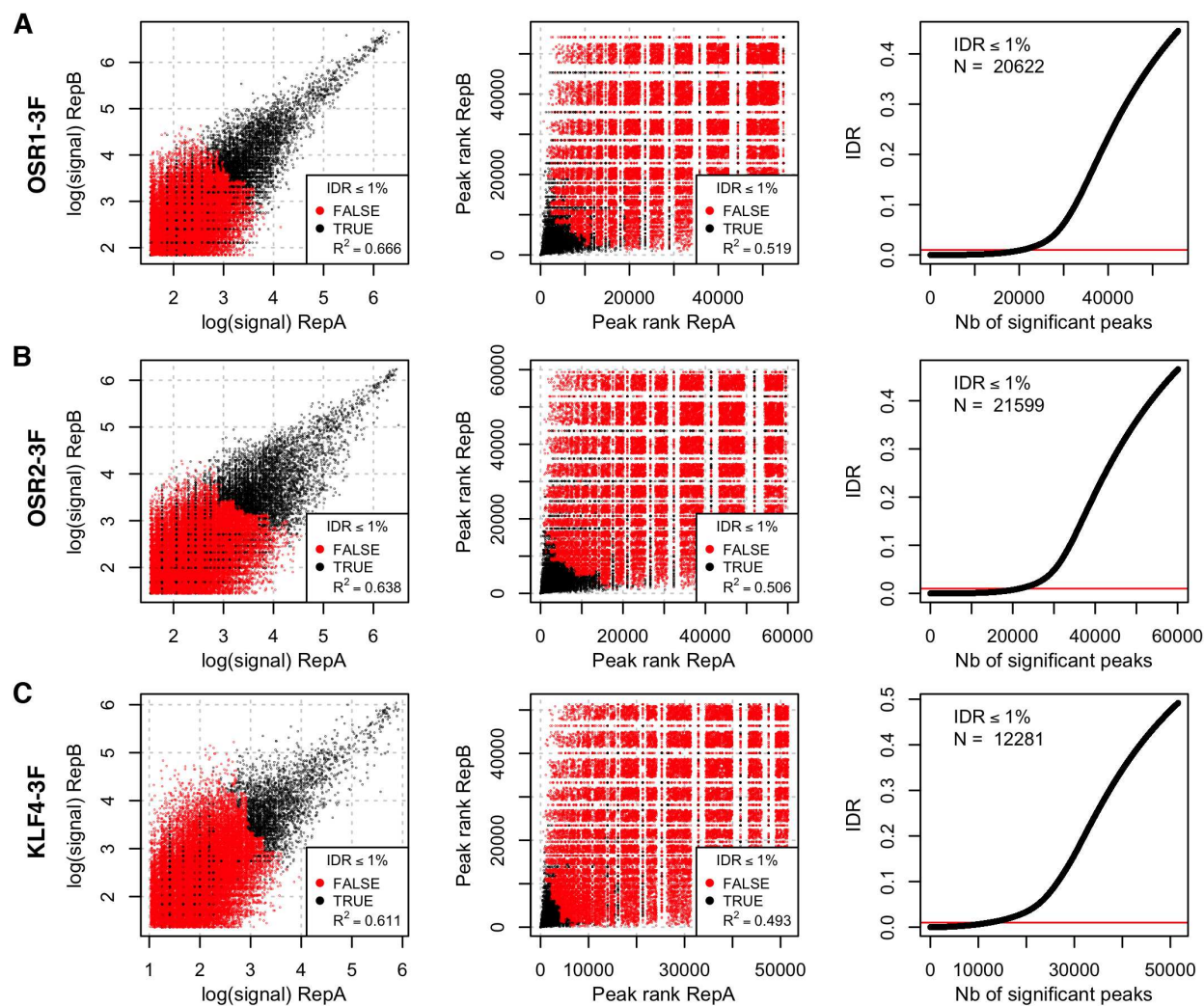
Supplemental figure S4. Differentially expressed genes belonging to the Integrin and TGF-beta signalling pathways. (A) Heatmap of the 56 DE genes associated with the Panther Integrin signalling pathway. **(B)** Heatmap of the 26 DE genes associated with the Panther TGF- β signalling pathway. Hierarchical clustering was performed by using the one minus Pearson correlation. Genes related to cartilage and bone development/differentiation are depicted in red.



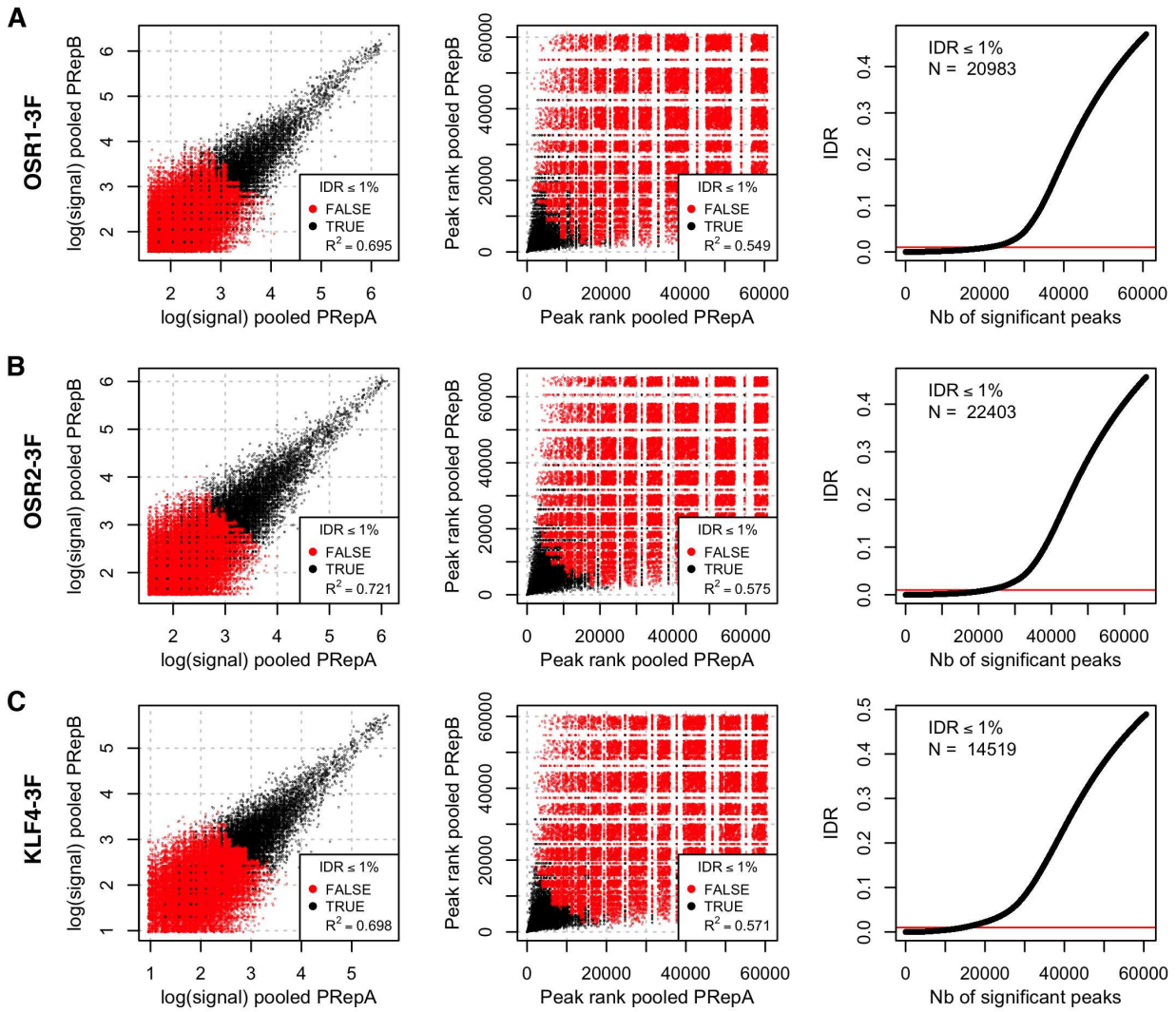
Supplemental figure S5. Chromatin landscape in the vicinity of *MYOG* in chMM cultures. Five covalent histone tail modifications were investigated by ChIP-seq in chMM cultures infected with retroviruses carrying no recombinant protein: H3K4me1, H3K4me2, H3K4me3, H3K27ac and H3K27me3. These post-translational modifications are frequently found in promoter and/or enhancer regulatory domains. H3K4me3 and H3K27ac are associated with transcriptional activity, whereas H3K27me3 is a mark of facultative heterochromatin.



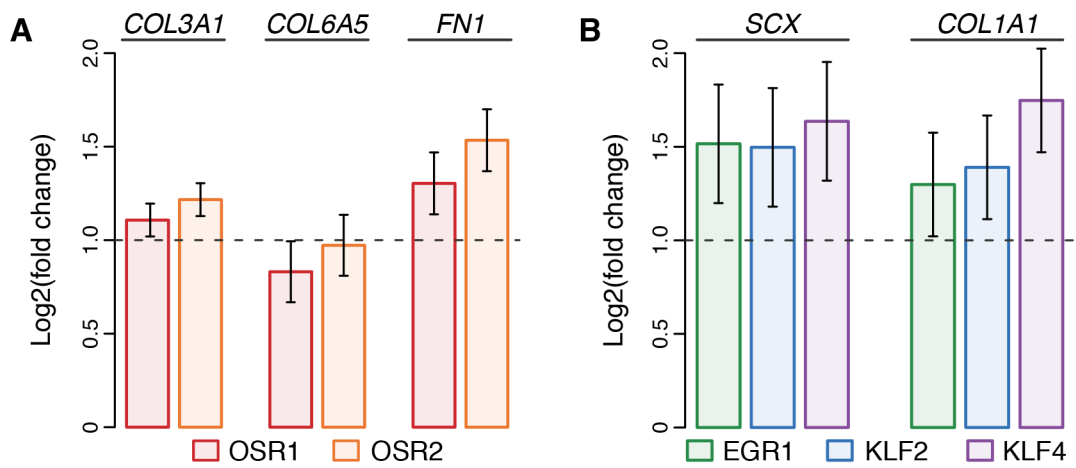
Supplemental figure S6. Cross-correlation analysis of ChIP-seq enrichment. Strand cross-correlation plot of OSR1-3F RepA (A), RepB (B); OSR2-3F RepA (C), RepB (D); EGR1-3F RepA (E); KLF2-3F RepB (F); KLF4-3F RepA (G), RepB (H), to evaluate the signal-to-noise ratio of the ChIP-seq enrichment profiles. NSC, normalized strand correlation (≥ 1.05); RSC, relative strand correlation (≥ 0.8). The first peak corresponds to the phantom peak identified at the sequencing read length; the second peak corresponds to the CHIP peak identified at the half of the fragmentation length.



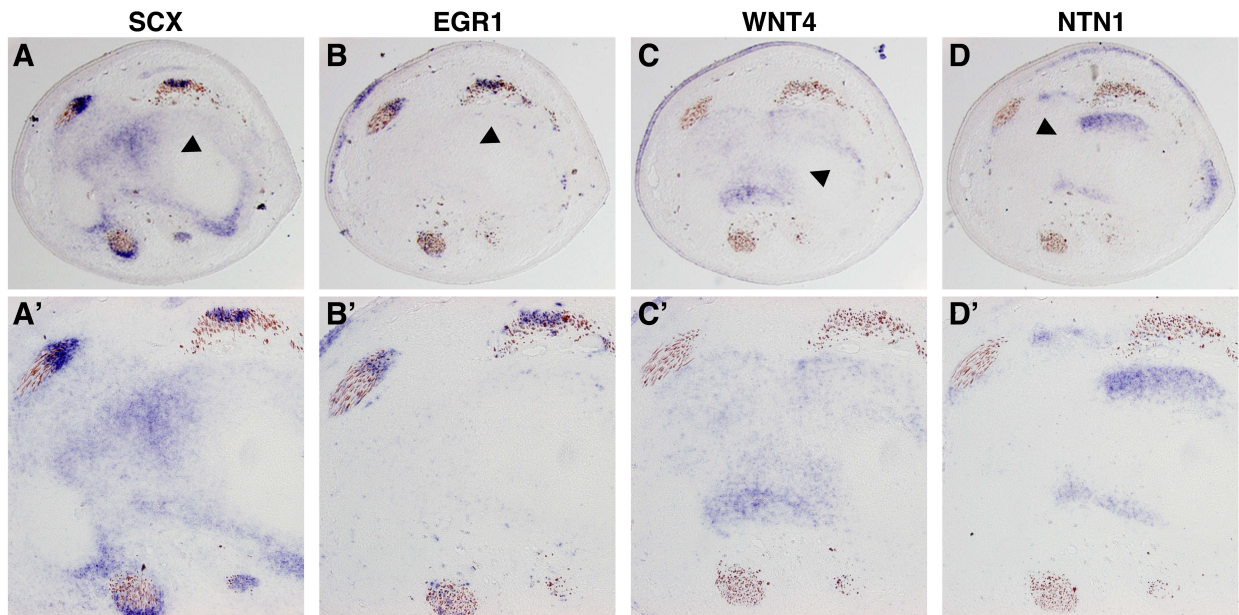
Supplemental figure S7. Replicate consistency IDR analysis. IDR analysis on the peaks identified for both biological replicates of OSR1-3F (A), OSR2-3F (B) and KLF4-3F (C) ChIP-seq. Peaks (N) passing an IDR threshold of 0.01 were considered as consistent between both replicates.



Supplemental figure S8. Pooled-replicate consistency IDR analysis. IDR analysis on the peaks identified for the pooled biological replicates of OSR1-3F (A), OSR2-3F (B) and KLF4-3F (C) ChIP-seq. Both biological replicates of each TF ChIP-seq were merged and randomly partitioned into two pseudo-replicates. Peaks (N) passing an IDR threshold of 0.01 were considered as consistent between both pseudo-replicates.



Supplemental figure S9. Expression levels of connective tissue-associated genes in chMM cultures. (A) Relative expression of the MCT-related genes *COL3A1*, *COL6A5* and *FN1* in OSR1- and OSR2-overexpressing chMM cultures as compared to control cultures ($p_{adj} < 10^{-5}$). (B) Relative expression of the tendon-related genes *SCX* and *COL1A1* in EGR1-, KLF2- and KLF4-overexpressing chMM cultures as compared to control cultures ($p_{adj} < 10^{-5}$).



Supplemental figure S10. Endogenous expression of *WNT4* and *NTN1* genes in forelimbs of E8 chick embryos. ISH for *SCX* (A,A'), *EGR1* (B,B'), *WNT4* (C,C') and *NTN1* (D,D') genes in forelimbs of E8 (HH34) chick embryos. (A-D) Adjacent and transverse sections were hybridized with *SCX*, *EGR1*, *WNT4* and *NTN1* probes (blue) followed by immunohistochemistry with the MF20 antibody (brown). Magnification 5X at the wrist level. (A'-D') Magnification 10X of regions marked with black arrowheads of sections (A-D). MF20 targets skeletal muscle myosin. Top, dorsal; left, posterior.

Supplemental tables

Supplemental table S1. Primer list.

Target	Accession	Sense	Sequence	Purpose
<i>OSR1</i>	NM_001168709.1	forward	GCATCCATGGGAAGTAAGACTCTGCCGG	RCAS-BP(A) constructs
		reverse	GCATGAATTCGCACTTAAGTTTGGCAGGTTTCAG	
<i>OSR2</i>	NM_001170344.1	forward	GCATCCATGGGCAGCAAGGCGCTGCCGG	RCAS-BP(A) constructs
		reverse	GCATGAATTCGGCTGTGCCCGGTAGATCACAG	
<i>EGR1</i>	NM_204136.2	forward	GCATCGTCTCCCATGGCTGCGGCCAAGGCAGAG	RCAS-BP(A) constructs
		reverse	GCATCGTCTCGAATTCGCAAATCTCAATTGTCCTTGGAGAAAAG	
<i>KLF2</i>	XM_418264.4	forward	GCATCGTCTCCCATGGCGCTGAGCGATACCATC	RCAS-BP(A) constructs
		reverse	GCATCGTCTCGAATTCATGTGCCGCTTCATGTGCAG	
<i>KLF4</i>	XM_004949369.1	forward	GCATCGTCTCCCATGCGGCAGCCCCCGG	RCAS-BP(A) constructs
		reverse	GCATCGTCTCGAATTCAAAGTGCCTCTTCATGTGTAAGGCG	
<i>OSR1</i>	NM_001168709.1	forward	TTCTGACCACCTTTCCAACC	ISH probe (CDS)
		reverse	TCCTTTACTTGCGTGTGCAG	
<i>KLF4</i>	XM_004949369.1	forward	CATCCTCTCCAACCTCCCTCA	ISH probe (CDS)
		reverse	CGGCATCAGCTCTTGGTACT	
<i>ADGRG2</i>	XM_015272749.1	forward	TGGGGATTTGCTTTCTTCAC	ISH probe
		reverse	AGCTTCCTCCCCATTTTGTT	
<i>ANXA1</i>	NM_206906.1	forward	TATGAAGGGGCTTGGAACTG	ISH probe
		reverse	ACATAAAGCGACCAGGATGG	
<i>CBFA2T2</i>	NM_001011689.1	forward	TGGAGATGGTGGAGAAGACC	ISH probe
		reverse	CCCAAACCTCCCCCTTAGAG	
<i>NTN1</i>	L34549.1	forward	GACGAGAACGAGGACGACTC	ISH probe
		reverse	GACGAGAACGAGGACGACTC	
<i>WNT4</i>	NM_204783.1	forward	GAGCTGGACAAGTGTGGATG	ISH probe
		reverse	GCAAGAGAGAAGCAGGGTCA	
<i>WNT11</i>	NM_204784.1	forward	TGATATCAGGCCGGTTAAGG	ISH probe
		reverse	GAAGGTCCCATTGGAAGTCA	

Supplemental table S2. RNA-seq mapping and assignment metrics.

Cells	TF	Replicate	Genome	Number of sequenced read pairs	Number of mapped read pairs	Percentage of mapped read pairs	Number of genes	Number of read pairs in gene features	Percentage of assigned read pairs	DESeq2 size factor	Number of genes with TPM >= 1	Number of DE genes (DESeq2; FDR 0.01; absolute FC ≥ 2; padj < 0.01)
chMM d5	Empty	RepA	galGal4	61,293,829	57,226,361	93.4%	21,350	51,233,721	83.6%	0.6251886	16,407	-
chMM d5	Empty	RepB	galGal4	70,336,622	65,997,861	93.8%	21,350	59,297,870	84.3%	0.7125234	16,075	-
chMM d5	OSR1-3F	RepA	galGal4	108,450,226	106,731,154	98.4%	21,350	95,874,397	88.4%	1.1457763	15,667	1,997
chMM d5	OSR1-3F	RepB	galGal4	112,332,929	109,359,879	97.4%	21,350	95,309,147	84.8%	1.1472096	15,773	-
chMM d5	OSR2-3F	RepA	galGal4	105,566,039	103,430,587	98.0%	21,350	93,496,499	88.6%	1.1373227	15,752	2,289
chMM d5	OSR2-3F	RepB	galGal4	96,930,250	94,485,092	97.5%	21,350	85,072,771	87.8%	1.0213997	15,565	-
chMM d5	EGR1-3F	RepA	galGal4	119,974,859	117,063,128	97.6%	21,350	106,079,374	88.4%	1.2019170	15,489	1,369
chMM d5	EGR1-3F	RepB	galGal4	101,138,411	99,474,007	98.4%	21,350	90,175,670	89.2%	1.0538920	15,591	-
chMM d5	KLF2-3F	RepA	galGal4	110,560,901	107,231,699	97.0%	21,350	97,733,305	88.4%	1.1331060	15,434	2,150
chMM d5	KLF2-3F	RepB	galGal4	102,353,299	99,545,240	97.3%	21,350	91,316,919	89.2%	1.0990220	15,411	-
chMM d5	KLF4-3F	RepA	galGal4	85,196,270	81,486,895	95.6%	21,350	74,139,243	87.0%	0.8774570	15,539	2,907
chMM d5	KLF4-3F	RepB	galGal4	109,766,442	105,073,271	95.7%	21,350	94,891,254	86.4%	1.1156930	15,581	-

TPM, Transcripts per million
DE, Differentially expressed
FDR, False-discovery rate
FC, Fold change
padj, Benjamini-Hochberg adjusted p-value

Supplemental table S3. Histone ChIP-seq metrics.

Cells	TF	Replicate	Antibody	Genome	Number of sequenced reads	Number of filtered reads	Number of uniquely mapped reads	Number of non-redundant reads	MAC2 fragment length (bp)	Number of broad peaks containing at least one narrow peak	Consistent peaks between replicates (≥ 50% of overlap)	FRiP
chMM d5	Empty	RepA	H3K4me1	galGal4	32,728,349	32,346,896	27,046,822	24,999,383	144	70,275	60,263	0.591
chMM d5	Empty	RepB	H3K4me1	galGal4	77,812,312	76,941,575	65,749,235	58,690,985	135	44,714	34,995	0.526
chMM d5	Empty	RepA	H3K4me2	galGal4	42,765,665	41,831,465	31,567,636	24,271,924	124	21,238	18,274	0.755
chMM d5	Empty	RepB	H3K4me2	galGal4	26,874,878	26,424,579	19,593,355	15,425,235	132	53,960	42,481	0.797
chMM d5	Empty	RepA	H3K4me3	galGal4	29,152,101	28,419,579	20,485,834	11,703,101	130	95,918	71,864	0.877
chMM d5	Empty	RepB	H3K4me3	galGal4	52,421,293	51,395,575	34,920,103	16,148,365	123	144	114	0.861
chMM d5	Empty	RepA	H3K27ac	galGal4	24,041,688	23,753,596	19,273,271	17,945,631	134	53,960	42,481	0.473
chMM d5	Empty	RepB	H3K27ac	galGal4	41,319,840	40,875,224	31,786,381	26,662,123	130	95,918	71,864	0.499
chMM d5	Empty	RepA	H3K27me3	galGal4	30,007,876	29,634,749	22,549,312	19,045,315	144	144	114	0.616
chMM d5	Empty	RepB	H3K27me3	galGal4	26,254,540	25,863,980	19,495,679	16,656,352	145	144	114	0.642

FRiP, Fraction of reads in peaks

Supplemental table S4. Transcription factor ChIP-seq quality metrics.

Cells	TF	Replicate	Antibody	Genome	Number of sequenced reads	Number of filtered reads	Number of uniquely mapped reads	Number of non-redundant reads	PBC (≥ 0.8)	NSC (≥ 1.05)	RSC (≥ 0.8)	SPP fragment length (bp)	Self-consistent peaks (IDR<0.01)	Self-consistent IDR (<2)	Replicate-consistent peaks (IDR<0.01)	Pooled self-consistent peaks (IDR<0.01)	Replicate-consistent IDR (<2)	FRiP (≥ 0.01)
chMM d5	OSR1-3F	RepA	anti-FLAG	galGal4	56,540,298	56,011,868	26,455,037	19,193,921	0.726	1.047	1.680	130	19,472					0.014
chMM d5	OSR1-3F	RepB	anti-FLAG	galGal4	32,725,733	32,506,774	18,122,153	13,210,252	0.729	1.063	2.081	130	21,839	1.122	20,622	20,983	1.018	0.022
chMM d5	OSR2-3F	RepA	anti-FLAG	galGal4	43,416,394	43,028,638	19,420,420	14,401,100	0.742	1.063	1.657	130	20,311					0.019
chMM d5	OSR2-3F	RepB	anti-FLAG	galGal4	41,411,543	41,137,694	23,477,039	17,417,387	0.742	1.055	1.935	130	22,169	1.091	21,599	22,403	1.037	0.024
chMM d5	EGR1-3F	RepA	anti-FLAG	galGal4	43,318,998	42,892,862	19,258,524	13,628,197	0.708	1.101	1.922	135	20,435					0.019
chMM d5	EGR1-3F	RepB	anti-FLAG	galGal4	26,740,191	26,556,705	15,155,849	11,268,462	0.744	1.121	1.968	135	18,588	1.099	11,737	16,627	1.417	0.033
chMM d5	KLF2-3F	RepA	anti-FLAG	galGal4	24,741,066	24,500,869	17,146,390	14,032,420	0.818	1.023	0.962	130	7,128					0.017
chMM d5	KLF2-3F	RepB	anti-FLAG	galGal4	35,046,332	34,782,452	17,222,767	13,485,356	0.783	1.090	2.899	130	23,391	3.282	6,542	21,352	3.264	0.027
chMM d5	KLF4-3F	RepA	anti-FLAG	galGal4	60,269,846	59,991,651	29,222,190	21,664,401	0.741	1.037	1.283	135	14,274					0.012
chMM d5	KLF4-3F	RepB	anti-FLAG	galGal4	35,627,835	35,369,563	26,294,445	17,348,739	0.660	1.054	3.166	135	24,595	1.723	12,281	14,519	1.182	0.024

PBC, PCR bottleneck coefficient
NSC, Normalized strand correlation
RSC, Relative strand correlation
IDR, Irreproducible discovery rate
FRiP, Fraction of reads in peaks

Publications

A manuscript gathering all the results presented in this thesis is currently in preparation. In addition, two publications and two manuscripts, which I sign as co-author, emerged from internal and external collaborative contributions.

- Berkholz J, Orgeur M, Stricker S and Munz B, 2015. skNAC and Smyd1 in transcriptional control. *Exp Cell Res* 336(2): 182–191.
- Knierim E, Hirata H, Wolf NI, Morales-Gonzalez S, Schottmann G, Tanaka Y, Rudnik-Schöneborn S, Orgeur M, Zerres K, Vogt S, van Riesen A, Gill E, Seifert F, Zwirner A, Kirschner J, Goebel HH, Hübner C, Stricker S, Meierhofer D, Stenzel W and Schuelke M, 2016. Mutations in subunits of the Activating Signal Cointegrator 1 complex are associated with prenatal spinal muscular atrophy and congenital bone fractures. *Am J Hum Genet* 98(3): 473-489.
- Vallecillo García P, Orgeur M, Stumm J, Poehle-Kronawitter S, Keppert V, Börno ST, Shinichiro H, Relaix F, Koch M, Timmermann B, Marazzi G, Sassoon DA, Duprez D and Stricker S. Odd skipped-related 1 (Osr1) identifies the fibro-adipogenic progenitor (FAP) lineage and regulates a pro-myogenic transcriptional program during limb development. *Nat Commun*: in revision.
- Milet C, O'Brien Allbright K, Orgeur M, Couplier F, Duprez D and Havis E. Egr1 loss-of-function leads to constitutive browning of mouse inguinal subcutaneous adipose tissue. To be submitted.

Scientific communications

- Poster presentation during the PhD retreat of the Max-Planck-Institute for Molecular Genetics, Brandenburg an der Havel, Germany, November 2012.
- Workshop preparation during the MyoGrad Summer School for Myology, Berlin, Germany, June 2013. Exploring the ENCODE data via the UCSC browser.
- Poster presentation during the PhD retreat of the Max-Planck-Institute for Molecular Genetics, Brandenburg an der Havel, Germany, September 2013.
- Oral presentation during the MyoGrad PhD retreat, Berlin, Germany, December 2014. Transcriptional network underlying irregular connective tissue and tendon formation.
- Poster presentation during the joint meeting of the German and French societies of developmental biologists, Nuremberg, Germany, March 2015.
- Poster presentation during the EMBO François Jacob Symposium: Genetic control of development and evolution, Paris, France, September 2015.

LIST OF FIGURES AND TABLES

List of figures and supplemental figures

Figure 1. Development of the vertebrate limb.	1
Figure 2. Schematic representation of the chondrogenesis process during limb development.	3
Figure 3. Schematic representation of the myogenesis process during chick forelimb development.	5
Figure 4. Schematic representation of the tenogenesis process during mouse limb development.	7
Figure 5. Schematic representation of muscle connective tissue formation during limb development.	10
Figure 6. Representation of the limb musculoskeletal system in the mouse and the fruit fly.	13
Figure 7. Quantitative model of gene expression.	15
Figure 8. Schematic representation of initiation of gene transcription.	17
Figure 9. Post-translational modifications of histone protein tails.	20
Figure 10. Schematic representation of promoter and enhancer architecture.	22
Figure 11. Schematic representation of transcription factor binding at DNA regulatory elements.	26
Figure 12. Models of transcription factor-binding at <i>cis</i> -regulatory regions.	27
Figure 13. Schematic representation of RNA polymerase II recruitment.	29
Figure 14. Endogenous expression of the transcription factors in hindlimbs of E4.5 chick embryos.	51
Figure 15. Endogenous expression of the transcription factors in hindlimbs of E9.5 chick embryos.	52
Figure 16. Overexpression of the transcription factors in chick embryo limbs.	53
Figure 17. Cell differentiation within the chMM culture conditions.	55
Figure 18. Transcript abundances within GO terms related to cell differentiation and development.	56
Figure 19. Workflow to design the gene annotation model.	59
Figure 20. Characteristics of the established gene annotation model.	60
Figure 21. Organization of the viral genome.	61
Figure 22. Sample-to-sample distance across chMM cultures.	62
Figure 23. 10,712 differentially expressed genes in all chMM cultures.	63
Figure 24. 4,298 unique differentially expressed genes across all chMM cultures.	64
Figure 25. <i>K</i> -means gene clustering of the differentially expressed genes.	65
Figure 26. Signalling pathway enrichment analysis of the differentially expressed genes.	67
Figure 27. GO enrichment analyses on the differentially expressed genes.	68
Figure 28. Histone modification signal profiles.	70
Figure 29. Regulatory domains.	71
Figure 30. Promoter states associated with genes.	73
Figure 31. Histone mark coverage profiles at the TSS of differentially expressed genes.	73
Figure 32. Promoter state surrounding the TSS of differentially expressed genes.	74
Figure 33. Detection of the 3F-tagged transcription factors within the chMM cultures.	76
Figure 34. Principal components analysis of the transcription factor-signal profiles.	76
Figure 35. Cross-correlation analysis of ChIP-seq enrichment.	77
Figure 36. Replicate consistency IDR analysis.	78
Figure 37. Pooled-replicate consistency IDR analysis.	79

Figure 38. Transcription factor-binding events within DNA regulatory domains.	80
Figure 39. Candidate target genes of the transcription factors.	82
Figure 40. Expression of selected target genes in chick forelimbs at E5.5.	84
Figure 41. <i>NTN1</i> is directly regulated by all transcription factors within the chMM cultures.	85
Figure 42. Endogenous expression of OSR1 and OSR2 target genes in forelimbs of E8 chick embryos.	86
Figure 43. Endogenous expression of <i>ANXA1</i> , a target gene of KLF4, in forelimbs of E8 chick embryos.	86
Figure 44. Overview of cell differentiation processes in chMM cultures.	96
Supplemental figure S1. RNA-seq fragment count normalization.	123
Supplemental figure S2. Endogenous expression of the transcription factors in E4.5 forelimbs.	123
Supplemental figure S3. Endogenous expression of the transcription factors in E9.5 forelimbs.	124
Supplemental figure S4. DE genes belonging to the Integrin and TGF-beta signalling pathways.	125
Supplemental figure S5. Chromatin landscape in the vicinity of <i>MYOG</i> in chMM cultures.	125
Supplemental figure S6. Cross-correlation analysis of ChIP-seq enrichment.	126
Supplemental figure S7. Replicate consistency IDR analysis.	127
Supplemental figure S8. Pooled-replicate consistency IDR analysis.	128
Supplemental figure S9. Expression levels of connective tissue-associated genes in chMM cultures.	129
Supplemental figure S10. Endogenous expression of <i>WNT4</i> and <i>NTN1</i> genes in E8 forelimbs.	129

List of tables and supplemental tables

Table 1. Promoter and enhancer regulatory domains.	72
Supplemental table S1. Primer list.	130
Supplemental table S2. RNA-seq mapping and assignment metrics.	131
Supplemental table S3. Histone ChIP-seq metrics.	131
Supplemental table S4. Transcription factor ChIP-seq quality metrics.	132