Developmental Genetics and the Evolution of Tendon Growth

a dissertation presented
by
Heather L. Dingwall
to
The Department of Human Evolutionary Biology

in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in the subject of

Human Evolutionary Biology

Harvard University
Cambridge, MA
September 2019

Developmental Genetics of Tendon Growth

Abstract

Extant apes are characterized by a high degree of variation in Achilles tendon size, and a tendon's capacity for energy storage is directly related to its relative size. Thus, it has been hypothesized that the relatively long, thin human Achilles tendon is an adaptation to decrease the cost of bipedal locomotion, especially during running. However, there is poor understanding of the mechanisms involved in tendon growth and how these processes lead to differences in adult tendon size among different organisms. Using a mouse model, this dissertation characterizes the transcriptomic and regulatory mechanisms that control the postnatal growth of limb tendons to better understand the process of tendon size evolution. Chapter Two introduces the concept of tissue growth in the context of tendon biology, hominin evolution, and human locomotion. Chapter Three examines the changing proliferative potential of tendon cells during the six weeks immediately following birth. We also characterize expression profiles of known target genes and show that mouse tendon growth occurs in two general phases: proliferative growth and increased matrix production. Chapter Four establishes a comprehensive transcriptomic profile of tendon and presents an open chromatin assay to identify putative regulatory elements involved in the transcriptional regulation of these growth-related genes. These studies reveal the complex gene interactions

that coordinate growth and may drive the observed shift in cell phenotype. Lastly, Chapter Five demonstrates how these putative regulatory modules can be used to address questions about the evolution of Achilles tendon growth in primates in order to contextualize our findings in mice within a human evolutionary framework.

TABLE OF CONTENTS

List of Figures

For my grandparents, Jody and Dave Kendall.

Acknowledgements

This dissertation was truly an interdisciplinary endeavor, and I am grateful to have had the opportunity to work in two labs that exposed me to different perspectives on science. I would like to thank the Capellini Lab members, past and present, for their constant enthusiasm for evolutionary biology, regardless of the specific topic. And many thanks to the Galloway Lab for exposing me to the field of regenerative medicine and for giving me an appreciation for drilling down on the mechanisms underlying interesting phenotypes.

To my advisors, Terry and Jenna: you have both changed my life in your own unique ways. I am amazed that you were both willing to take a chance on me and this project, and I am so grateful that you did. I only hope that the gamble pays off.

I owe a huge thanks to my collaborators, Mor Grinstein, Rishita Shah, Ken Zou, and Luke O'Connor, without whom none of this work would have been possible. Although she has long since left Harvard I will be forever grateful to Jessie Willen for tackling difficult lab projects head-on with me, as well as for her friendship. I must also acknowledge the assistance of the employees of the Harvard Bauer Core and MGH Flow Cytometry Core; their expertise ensured that experiments ran smoothly and successfully. I am especially grateful to Claire Reardon for her help planning and trouble-shooting all of the sequencing assays I have run over the years. Many thanks to the veterinarians and technicians at both the Harvard BRI and the MGH CCM for their help caring for our mice, to the mice themselves for making the ultimate sacrifice, and to Harvard Research Computing for watching over the Odyssey cluster. Thanks also to Radhika Ketani and Meeta Mistry of the HSPH Bioinformatics Core for their reassurance and sanity-checks when I doubted my analyses. The Harvard ATAC-seq Club has also been a great source of inspiration and assistance, for which I

am very grateful. Thank you, also, to the HEB, Peabody, and MGH CRM staff for making our work in the lab so much easier. A very special thanks to Meg Lynch, the head administrator for HEB, without whom I would be lost.

The HGWISE program played an important part in my professional development at Harvard. I am particularly grateful to have been paired with such an amazing HGWISE mentor, Elena Kramer. She has been one of the few constants in my life during my time in graduate school, always ready dive into difficult topics over lunch at HLS. I am also grateful to all of the HEB faculty for their feedback and support over the years, especially David Pilbeam and Rachel Carmody. I am so thankful to have had David in my corner for my entire time in HEB, although I'm not sure what I did to deserve it. And Rachel provided some much-needed advice and encouragement during a particularly difficult time, and has served as a role model for me since the beginning of graduate school. Many thanks to Bernard Wood, Kevin Hatala, Erin-Marie Williams, and Kes Schroer (Domone) for setting me on the path to graduate school and for continuing to cheer me on from a distance.

I have made some incredible friends and colleagues while at HEB. I am particularly grateful for my friendship with Cary Allen-Blevins and Laura Klein, who are incredibly smart, funny, and kind. I have learned so much from them about mother's milk, microbes, and life in general. I am also thankful to have two wonderful young women, Evvie Jagoda and Mariel Young, in the Capellini Lab with me. They have helped to make the lab a more supportive environment and I'm so proud of the support network we've built together. I would also like to thank Linda Reynard, Bridget Alex, Katia Chadaideh, Ian Wallace, Brian Addison, Erik Otarola-Castillo, Eric Castillo, Sam Urlacher, and Liz Brown for their friendship over the years. Neil Roach deserves a special thanks for going out of his way to give me more leadership opportunities as his TF, for regularly giving me solid career advice, and for also giving me a swift kick in

the pants to finish up and defend. And thank you to my amazing therapist, Maggie, who has helped me learn to cope with the many, many failures that come with being a scientist.

Nick Holowka is especially dear to me. Not only is he a careful scientist, but his kind, even demeanor has the ability to calm even the most stressed graduate student. He is a true partner and I am eternally grateful for his support and encouragement, and that of his family.

Lastly, I owe my family and friends the utmost gratitude; their love and support has never wavered, even during my worst moments. My friends Rachel Spitz-Lieberman, Malea Grubb, and Alexis Peterson have been with me since long before any of us dreamed of leaving the ballet world. I am so grateful to have been able to grow and change with them over the many years of our friendship. And finally, a million thanks to Dad and Judy; Mom and Peter; and my grandparents Jody and Dave. They help me remember that I am more than my H-index and have worth outside of the lab.

CHAPTER 1

INTRODUCTION

Comparative primate anatomists have long appreciated the differences in Achilles tendon morphology among extant apes (i.e., hominoids), which includes humans, chimpanzees, bonobos, gorillas, orangutans, gibbons, and siamangs. Upon gross examination it is clear that the human Achilles is much longer and more developed than that of chimpanzees (Figure 1.1), bonobos, gorillas, and orangutans (Swindler and Wood 1973; Diogo 2011; Diogo 2013; Diogo et al. 2013; Aerts et al. 2018). The average reported human Achilles tendon length is approximately 150 mm (110 to 260 mm) (Doral et al. 2010; Rosso et al. 2012) while the Achilles is poorly developed in non-human great apes (Figure 1.1) (Rauwerdink 1991; Diogo 2011; Diogo 2013; Diogo et al. 2013). This variation naturally provokes the question of what selective forces determine Achilles tendon morphology in primates. Addressing this question requires a brief review of tendon structure and function.

1.1 Tendon structure and function

Tendons are dense regular connective tissue structures that connect muscles to bones, enabling muscle force transmission across joints to produce movement. Mature vertebrate tendons are fibrous, matrix-rich structures consisting of extracellular matrix (ECM) proteins embedded within a ground substance that is rich in proteoglycans and water (Kannus 2000). Collagen (predominantly type I) and elastin are the primary tendon ECM proteins, making up ~65-80% and ~1-2% of tendon dry mass respectively (Hess et al. 1989; Kannus 2000), although elastin fibers are not found in all human

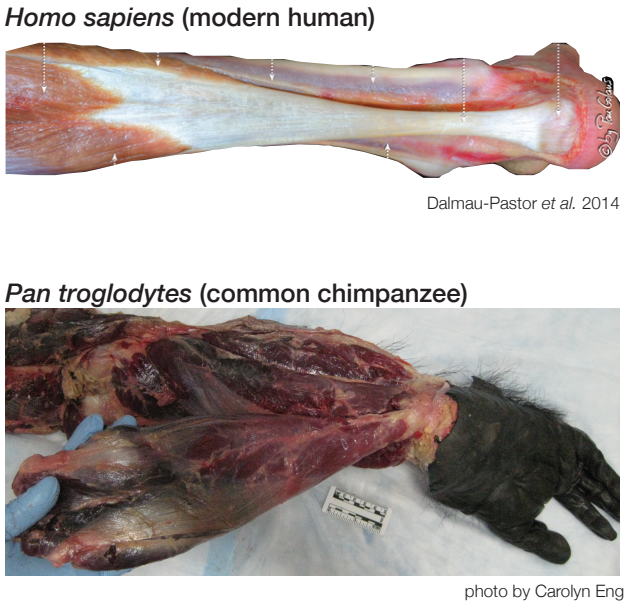


Figure 1.1: Representative images of adult modern human (top) and chimpanzee (bottom) Achilles tendons.

tendons (Józsa, Bálint, and Demel 1978; Carlstedt 1987). These components of the ECM are produced by the tendon cells, which can be found between bundles of collagen fibrils called fibers, which are the basic unit of any tendon (Hess et al. 1989). These fibers are further organized into primary, secondary, and tertiary fiber bundles, creating a tissue out of hierarchically organized collagen (Kannus 2000).

The high proportion of type I collagen in tendon is largely responsible for the viscoelastic properties observed in this tissue (Biewener 2008). This allows limb tendons, such as the Achilles, to store and release elastic strain energy during activities such as running and jumping, thereby decreasing the cost of locomotion and enabling power amplification (Alexander 1984; Alexander 2002; Biewener 2008; Biewener and Patek 2018). The compliance of a tendon, and thus the amount of elastic energy it can store, is directly related to its structure; long, thin tendons can store more energy per unit volume (Biewener and Roberts 2000). Additionally, for a given limb length, having longer tendons necessarily leads to a more proximal grouping of the limb muscles.

This decreases the relative mass of the limb, in addition to decreasing expensive musculature, which can also help decrease the cost of locomotion (Biewener 2008). For these reasons, increased limb tendonization has been proposed as one adaptation for running and jumping in the animal kingdom (Alexander 1984; R. M. Alexander 1991b; Alexander 2002). However, increased tendon length and compliance comes at the cost of positional control of the joint(s) crossed by the tendon (Rack and Ross 1984; Biewener (2008)). Thus the ecology and locomotor demands on a given animal will strongly influence the anatomy of a limb tendon such as the Achilles.

1.2 Evolution of the primate Achilles tendon

The striking differences observed between humans and their closest living relatives (non-human great apes), combined with the findings that limb tendons can store elastic energy during running and jumping in other animals (see above), have led to the hypothesis that a long Achilles tendon is an adaptation for endurance running in humans (Bramble and Lieberman 2004). Meanwhile, the short Achilles observed in extant non-human great apes is hypothesized to be beneficial for arboreal behaviors, giving the animal better positional control of its foot and ankle during climbing (Preuschoft 1992; Jens et al. 2004). Within this evolutionary paradigm, a short, chimpanzee-like Achilles likely represents the ancestral state for all hominins and the derived condition likely arose with the emergence of the genus *Homo* (Bramble and Lieberman 2004).

Although several qualitative reports of relative Achilles length exist, there is remarkably little quantitative data available on primate tendon size variation. A single study by Rauwerdink (1991) contains the most exhaustive data set of primate tendon measurements (normalized to a standardized tibia length) from ten different primate taxa and one outgroup (chinchilla). This work shows that Achilles tendon length

varies widely among extant primates (Figure 1.2) and that this size variation exists on a continuous gradient. While Rauwerdink's sample size and taxonomic breadth are not large enough to statistically investigate the effects of phylogeny and locomotor repertoire on Achilles morphology, Rauwerdink claims that tendon size does not correlate well with locomotor mode among these primates (Rauwerdink 1991). However, locomotor behavior appears to be a better predictor of Achilles tendon morphology than phylogeny alone (see Figure 1.2). Non-human great apes share many arboreal and terrestrial locomotor behaviors, namely terrestrial quadrupedalism, vertical climbing, clambering, and orthograde suspension (Hunt 2016). Rauwerdink only measured the chimpanzee (*Pan troglodytes*) Achilles, but qualitative reports of gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*) musculoskeletal anatomy suggest that their Achilles tendons are of a similar relative size (Diogo 2011; Diogo 2013; Diogo et al. 2013). Interestingly, the woolly monkey (*Lagothrix lagotricha*) Achilles measured by Rauwerdink was also short, similar to that of the chimpanzee (Figure 1.2) (Rauwerdink 1991). These taxa are phylogenetically distant but use similar arboreal locomotor behaviors that may have driven convergence of this trait.

The Achilles tendons measured from cercopithecines (rhesus macaque (*Macaca mulatta*), hamadryas baboon (*Papio hamadryas*), and grivet (*Chlorocebus aethiops*)) generally fall in the intermediate range (Figure 1.2). All three of these species practice terrestrial quadrupedalism, but Rhesus macaques and grivets are also known to be partly arboreal (Fleagle 1988; Schmidt 2011). Hylobatids (gibbons and siamangs), also, exhibit an intermediate relative Achilles length similar to that of baboons, and much longer than that of non-human great apes (Rauwerdink 1991; Aerts et al. 2018); thus they are more similar to taxa outside their clade. Although hylobatids are arboreal and predominantly practice brachiation, they are capable of rapid terrestrial bipedal, "tripedal", and quadrupedal locomotion (Vereecke et al. 2006 JHE), in addition to leaping (Fleagle 1976). Biomechanical analyses have suggested that hylobatid Achilles

possess the potential for elastic energy storage during bipedal locomotion (Vereecke et al. 2006; Vereecke and Aerts 2008). Other studies of leaping in hylobatids suggests that their Achilles could perform power amplification during takeoff (Channon et al. 2009; Channon et al. 2010). Either behavior may have served as a selection pressure for a longer tendon in hylobatids compared to the other apes. However, more recent work asserts that hylobatid bipedalism is not conducive to elastic energy storage by the Achilles tendon and thus the size of the Achilles in gibbons and siamangs is not adaptive (Aerts et al. 2018). Given the similarities between hylobatid and cercopithecine Achilles tendon morphology, Aerts and colleagues argue that perhaps the ancestral hominoid Achilles was more cercopithecine-like, rather than short and chimpanzee-like. Thus, they argue that Achilles morphology was not under selection in hylobatids, but instead was inherited from their common ancestor with cercopithecines (Aerts et al. 2018). However, power amplification during leaping behaviors (Channon et al. 2009; Channon et al. 2010) could have served as a sufficiently strong selection pressue to drive Achilles elongation in hylobatids. This debate about the ancestral state potentially complicates Bramble and Lieberman's hypothesis (Bramble and Lieberman 2004), but does not negate the possibility that a long Achilles tendon is an endurance running adaptation in *Homo*, it merely alters the magnitude of the morphological change that has occurred over the course of primate evolution. It also raises the interesting possibility that the extremely short Achilles in extant non-human great apes is a derived feature that also underwent selection for positional control during climbing behaviors. Regardless of the Achilles morphology of the last common ancestor between humans and chimpanzees, or among hominoids, the limited data from extant primates suggests that Achilles tendon size is evolutionarily labile and has likely undergone intense selection for locomotor and positional behavior many times during primate evolutionary history.

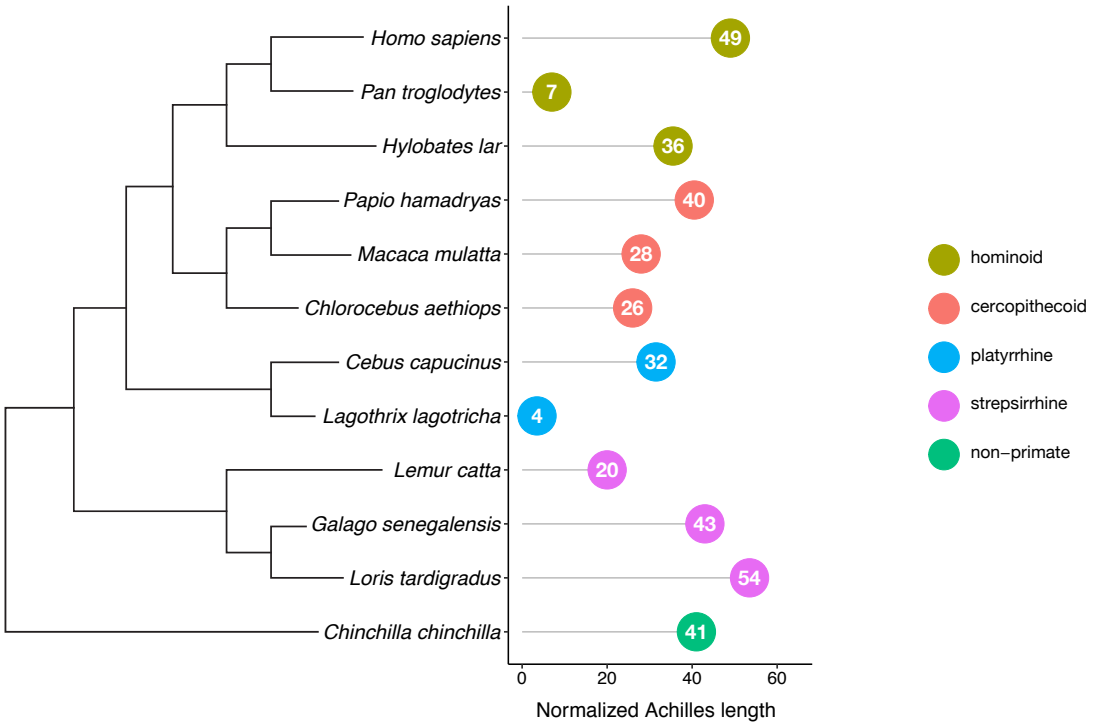


Figure 1.2: Plot of Achilles tendon lengths from eleven primate taxa with one outgroup (Chinchilla). Adapted from Rauwerdink 1991. Human data from Rosso et al. 2012.

1.3 Relevance of postnatal growth

Regardless of which tendon morphology is truly ancestral among apes, the average human Achilles is markedly longer and more developed than anything observed in extant non-human apes and cercopithecines (Figure 1.1 and Figure 1.2). How evolutionary forces created this morphology in modern humans is unknown, although selection almost certainly would have had to act on some mechanism(s) controlling the growth of the tissue. Tendons, like all tissues, can grow via two primary mechanisms: making more cells (proliferation) and making more ECM. During embryonic development, cells proliferate readily in the future tendon as the cells are specified and tissue is patterned. Much of mammalian tendon growth continues postnatally (Comer 1956), but by the time tendons reach their adult size many of the tendon cells are no longer proliferative (Kalson et al. 2015). Even the ECM does not appear to turn over after the tissue reaches maturity, which is around the age of seventeen in humans (Heinemeier et al. 2013). Thus there is a discrete postnatal developmental window during which altering the mechanisms of growth could have a marked effect on adult tendon morphology. Such genetic mechanisms could explain the dramatic differences in tendon length and the evolutionary patterns of variation in tendon morphology among primate species, including humans.

1.4 Brief overview of the dissertation

The biomechanical consequences of such different anatomical phenotypes have been long appreciated, but there is a poor understanding of the molecular mechanisms active during tendon development and growth that lead to these phenotypic differences among primate species. In order to determine how evolution could tinker with cell and/or matrix growth mechanisms in the tendon, we first need a better understanding of 1) how a normal tendon grows postnatally and 2) how this is genetically controlled.

Previous work has shown that morphological diversity evolves primarily through changes in expression of genes with relatively conserved coding sequences via mutations in sequences involved in *cis*-regulation of developmental processes (see Chapter 2). Thus, in this dissertation, I take a developmental regulatory genetics approach to understanding the evolution of the variability in tendon size among primates.

Chapter 1 lays out a framework for using methods from developmental genetics to address questions of musculoskeletal evolution in primates. Chapter 2 reports a novel method for the extraction and purification of intact, high quality RNA from fresh tendon tissue; this is necessary for reliable study of the tendon transcriptome. Chapter 3 examines the changing proliferative potential of tenocytes during the postnatal growth period in a mouse model in order to better understand mechanisms of growth in the tendon. Chapter 4 integrates transcriptomics and genome-wide profiling of chromatin accessibility to identify genetic regulatory modules correlated with postnatal growth in the tendon. Finally, the Discussion and Future Directions explores the implications for the findings of this dissertation, and provides an example of how to apply the results from Chapter 4 to the topic of genetic regulatory evolution in tendon growth.

CHAPTER 2

COMBINING GENETIC AND DEVELOPMENTAL METHODS TO STUDY MUSCULOSKELETAL EVOLUTION IN PRIMATES

This chapter has been published as a book chapter: Capellini TD and Dingwall H. 2017. Combining Genetic and Developmental Methods to Study Musculoskeletal Evolution in Primates. In *Building Bones: Bone Formation and Development in Anthropology*. Cambridge University Press. pp 175-204.

2.1 Introduction

Among mammals, primates exhibit remarkable diversity in skeletal morphology. Much of this diversity is readily apparent in the appendages, body parts that interact with substrates during locomotion and positional behavior. Differences in the lengths, shapes, and proportions of the major long bones of the forelimb (scapula, humerus, radius, ulna) and hindlimb (pelvis, femur, tibia, fibula) reflect the myriad skeletal adaptations primates have evolved to occupy diverse ecological niches. This diversity is not only observable at the level of the entire appendage or individual limb segment, but at specific functional zones, such as growth plates, joints, and muscle-attachment sites. From an evolutionary perspective, this striking morphological diversity reflects the actions of natural selection on variation in pre- and postnatal developmental processes (Carroll, 2008). Historically, this diversity has inspired biologists to search for the developmental and genetic underpinnings of skeletal shape. Yet despite many decades of research, relatively little is known about the molecular mechanisms that control the specific shapes of bones, let alone how modifications to pre- and postnatal

developmental programs influence the morphological variation within and between species. A deeper exploration of these mechanisms is necessary to establish precise connections between genotype and phenotype (Hartl and Ruvolo, 2011), and in doing so to understand the nature of species adaptation and evolution.

A modern synthetic approach, one which integrates experimental findings from developmental biology, genetics, genomics, and bioinformatics, has the potential to provide increased power and resolution in connecting genotype to phenotype and revealing the causative mutations that underlie adaptive morphological evolution. Given the noticeable and marked variation in animal appendages, the identification of the molecular and cellular mechanisms that control limb skeletal development and diversity has been one of the main areas of research within developmental biology for well over 50 years. This is in part a consequence of the findings that experimental disruptions to limbs or natural mutations that impact limb morphology do not necessarily influence embryonic survival, making limbs a tractable system to study developmental principles (Gilbert, 2013). Thus, studies treating the limb as a developmental system have consistently been at the forefront of revealing the basic molecular and cellular mechanisms underlying developmental processes. Not surprisingly, this area of research has also consistently introduced cutting-edge experimental techniques that have aided in the identification of genes, their expression patterns, and their functions within living organisms. Importantly, such achievements have also been matched by advances in genetics, specifically in the development and improvement of methods that serve to map genetic regions to trait variation. For example, the development of genome-wide association studies (GWAS) and refinements in quantitative trait loci (QTL) mapping has achieved a heightened ability to elucidate regions in the genome that explain variation in limb and skeletal morphology. Most recently, functional genomics techniques, which have taken advantage of next-generation sequencing (NGS) technologies, have permitted the genome-wide identification of gene transcripts and

regulatory sequences involved in skeletal development. These newest approaches, when used in the context of bioinformatics and comparative genomics, are beginning to help to refine genomic signals within genetically mapped intervals to causative loci. When all of these approaches are considered collectively, scientists now have a powerful, versatile toolkit to understand limb and skeletal development and to elucidate how nucleotide diversity underlies appendage morphological variation within and between animals.

As alluded to earlier, a major finding that has been experimentally corroborated through achievements in each of the above fields is that as DNA sequences are the units of heredity, modifications to the DNA molecule directly impact molecular processes. These modifications in turn influence the development of a phenotype and phenotypic variation within a species. Under a selective regime, slight perturbations to a developing system, which results in heritable variation, eventually can lead to species-specific adaptations. Indeed, variation in adult skeletal phenotypes often has its roots in changes to early developmental programs in utero and/or processes of growth and maturation that occur during postnatal life (e.g., Young et al., 2006; Chan et al., 2010). Within primates, for example, skeletal morphology and proportions of the major long bones are often established in utero and/or early in postnatal growth (Young et al., 2006), a likely consequence of natural selection operating on the genes that control the patterning of skeletal elements (e.g., determining the number of cell populations that form the femur versus the tibia), and/ or those that control skeletal growth (e.g., regulating proliferation of chondrocytes in the growth plate).

In the context of development, another major finding is that developmentally encoded traits, such as skeletal shape, are often controlled at the level of gene regulation rather than through modifications to the protein-coding portion of the gene (Carroll, 2008). Perhaps because of their visibility in the genome and their relative predictive effects

on protein function, for many years mutations in the protein coding portions of genes were argued to play major roles in adaptive phenotypes. While there are examples of this in the literature, it is now understood that the large majority of genes in the genome have many different roles during development and postnatal growth and that alterations to their function via coding mutations can result in an extensive pleiotropism with deleterious consequences to the organism and its fitness. On a comparative genomics level, one signature of this impact has been the finding that coding portions of genes display high sequence conservation across a large number of vertebrate species, indicating a conserved function for the protein during life (Yue et al., 2014). On the other hand, more recent studies, such as those examining the regulatory control of single genes (e.g., Mortlock et al., 2003) or those focusing of the regulatory architecture of different cell types (ENCODE Consortium, 2012), reveal that the majority of genes have complex regulatory sequences (i.e., on/off switches) that function to drive gene expression in highly specific spatial and temporal domains. By controlling gene expression in a modular fashion, these regulatory switches have the ability to mediate specific anatomical outcomes. This specificity in the control of gene expression helps organisms avoid the extensive pleiotropic effects of coding mutations and provides a mechanism for natural selection to target specific components of functional anatomy. In light of these findings, it is not surprising that recent comparisons of the regulatory architecture of the genomes of different species has revealed considerable divergence in the regulatory control of genes in similar tissues and cell types (Yue et al., 2014).

One additional insight that has emerged into the developmental genetic control of trait variation is that many morphological traits have a polygenic underpinning rather than being controlled by only a single locus. Indeed, it has been known for some time that specific anatomy, even down to a musculotendonous insertion site on bone, may be controlled by tens to hundreds of loci, many of which of are likely regulatory

(Carroll, 2008). However, it is also understood, via findings from comparative genetic mapping experiments, that the extent to which each locus explains heritable variation in a trait is dependent on each species' evolutionary history. For example, in mice, the Growth Differentiation Factor Five (Gdf5) gene, a bone morphogenetic protein that is expressed in growth plates, controls approximately 10–15% of the growth of normal long bones, whereas in humans, GDF5 contributes to less than 1% of growth, even in cases when the gene's function is entirely missing (Capellini et al., unpublished). When considering specific evolutionary histories, it is not surprising that there are cases where only a few loci may have been under intense selection and end up explaining large percentages of variation in the trait. For example, mutations within a tissue-specific regulatory element for the Pitx1 gene explain approximately 65% of variation in pelvic fin presence/absence in some freshwater populations of stickleback fish (Shapiro et al., 2004; Chan et al., 2010). On the other hand, given the nature of selection on highly complex phenotypes, some traits have potentially thousands of underlying loci, none of which control more than a small percentage of variation. For example, human height is likely controlled by thousands of loci; the most potent locus controls only about 1–2% of normal variation in this phenotype (Wood et al., 2014). In lieu of the above, revealing the causative adaptive mutations that control variation in skeletal shape within and/or between species is a daunting task and one that will require insight from multiple scientific angles (see Chapter 1 for further discussion on parsing the genetic basis of complex traits like height).

Given the complicated relationship between genotypes and phenotypes, how then do scientists identify functionally important loci and gauge how much variation they control? How do they sift through the numerous genetic variants within an identified locus to find the variants directly responsible for changes in a species-specific phenotype? Finally, how do they functionally test these sequences to reveal molecular mechanism and their impacts on development? This chapter addresses these questions

and issues in the context of appendage skeletal development and evolution. The goal is to inform the evolutionary developmental anthropologist as to the genetic, molecular, and developmental tools that are available for them to explore aspects of the genotype–phenotype puzzle in the context of primate skeletal variation and evolution.

2.2 Connecting Genotype to Phenotype

To connect genotype to phenotype using a developmental genetics perspective, it is imperative that the DNA base-pair sequences that control limb development, growth, and maturation be identified, and this can be accomplished using multiple approaches. For example, geneticists have historically used mapping approaches to reveal loci that underlie variation in skeletal morphology (see Pardo-Diaz et al., 2015 for review), while more recently, functional geneticists and genomicists have used NGS methods to identify and then hone in on the types of genetic mutations that affect developmental pathways and underlie variation in morphology and disease risk (Gibson and Muse, 2009; Barrett and Hoekstra, 2011). To establish direct functional links between loci and traits and thus identify precisely how genes and their protein products contribute to morphological development, developmental biologists have used techniques such as gene mis-expression in the chick (e.g., Logan and Tabin, 1999) and targeted gene deletion/replacement in the mouse model systems (e.g., Menke, 2013). When used in concert to address evolutionary questions, these approaches can be quite powerful in revealing the sequences that control adaptive trait variation. These topics are addressed in detail below.

2.2.1 Genetic Methods: Finding the Loci that Control Trait Variation

Studies that focus on identifying loci that control trait variation typically fall into one of two main categories: "forward approaches" or "reverse approaches." When a phenotype is known to vary and researchers are seeking to identify the loci that control its variation, these studies are often classified as either "forward genetics" or "forward genomics." Investigative genetic studies involving twins or family pedigrees with known phenotypes fall under the "forward genetics" category, as do genetic association studies such as QTL mapping and GWAS. Most recently, "forward genomics" has emerged and involves using comparisons of known traits across many taxa (i.e., between group variation) in concert with full genome sequences from those taxa to find recurrent genomic regions that control convergent or divergent biological traits (e.g., Hiller et al., 2012). On the other hand, the "reverse genetics" or "reverse genomics" approach employs a de novo scan across a locus or the genome to find sequences that show characteristics of functional and evolutionary change (e.g., evidence of natural selection or drift). Both "reverse approaches" result in the identification of novel sequences, albeit these sequences still need to be linked to biological phenotypes using some of the functional techniques outlined below. This approach is especially powerful in revealing regions of the genome that display evidence of potentially adaptive evolution in different primates and humans (Prabhaker et al., 2006, 2008; McLean et al., 2011; Vitti et al., 2013). To make sense of any biological outcome, both approaches often rely on previous knowledge of the genomic region under investigation and downstream analyses often focus on known genes with established biological effects in chicks, mice, or humans. This "candidate" gene perspective has been particularly useful in twin and family studies to uncover causative mutations involved in limb skeletal variation (Farooq et al., 2013). For example, novel mutations in a functional domain of the GDF5 gene were identified in members of a large Pakistani family exhibiting brachydactyly by sequencing the gene's coding region in affected and unaffected

individuals (Farooq et al., 2013). These mutations disrupted digit joint development, resulting in the absence of the intermediate phalanx and revealed key functions for GDF5 in distal joint development. However, many family studies are inherently limited by the availability of families/cohorts with relevant phenotypes and by the fact that the individual genetic influences on a trait are difficult to identify because they can be hidden by the segregation of other genes and/or noise produced by environmental or experimental variation. Often candidate gene studies are conducted in patients and yield no significant results that are then not reported; this is likely due to the fact that many genes and regulatory regions control the trait in question. Moreover, the candidate gene approach is still limited mainly to coding regions, yet mutations in nearby or faraway non-coding regulatory sequences may be the true culprits responsible for the phenotype of interest.

Recent genomics approaches, such as whole exome sequencing (WES) in which untranslated and coding regions of genes are sequenced genome-wide, allow for a more rapid identification of causative mutations (Tetreault et al., 2015). This is an especially powerful approach when affected and non-affected siblings are both sequenced. For example, using WES, Parry and colleagues (2013) identified mutations in the *Goosecoid* gene that cause short stature, auditory canal atresia, mandibular hypoplasia, and skeletal abnormalities. This technique is even more powerful when it is performed on multiple individuals with and without a given phenotype from different families. Nilsson and colleagues (2014) have identified several different causative mutations in the *Aggrecan* gene, which encodes a proteoglycan in the extracellular matrix of growth plate cartilage, underlying idiopathic short stature syndrome in three different families. While WES and related techniques have been important in mapping relatively simple, monogenic, Mendelian-inherited traits and diseases, they are only now being used to identify the mutations that underlie complex, polygenic traits and diseases with some (limited) success (e.g., autism via Codina-Solà et al., 2015; neural tube defects via

Lemay et al., 2015).

The "forward genetic" approaches of linkage and association mapping (such as GWAS) have more often been used to identify the loci that underlie complex, polygenic traits that display patterns of normally distributed variation. Evolutionary biologists routinely use linkage mapping to identify loci that influence quantitative variation in a particular trait, or QTL (Hartl and Ruvolo, 2011). In QTL mapping, often a broad locus can be uncovered if it is linked genetically via some marker in the genome (e.g., microsatellite repeat, or single nucleotide polymorphism (SNP)) to observable variation in a trait. Such markers, when dispersed across the genome, can be tracked for how they co-segregate with variation in the trait in question, especially after analyses are conducted on offspring cohorts across several generations (e.g., F1 and F2 generations). Typically, QTL mapping is performed in organisms that have marked genetic and phenotypic diversity. In evolutionary biology contexts (i.e., with model and non-model organisms) researchers take advantage of this extensive trait heterogeneity, and improvements in genome-wide genotyping, and experimentally cross individuals in order to track alleles that influence phenotypic variation in the F1 and F2 hybrid generations. Statistical analysis then allows markers to be associated with trait variation and can reveal significant, but often very wide (e.g., megabase) QTL intervals that often harbor many putative functional loci (genes and regulatory sequences). Experimental studies with offspring cohorts possessing many individuals (e.g., thousands) have generally uncovered more loci, including those of slightly smaller effects, as well as loci that are a bit narrower due to the greater number of recombination events that partition meaningful genetic and phenotypic variation into smaller co-segregating blocks. One striking example was by Shapiro and colleagues, who used QTL analyses on stickleback pelvic fin phenotypes to identify a genomic interval that contained numerous genes including Pitx1, a key gene involved in pelvic development (Shapiro et al., 2004; Colosimo et al., 2005). Other examples within mice include

the identification of the cadherin11 locus in the control of femoral microarchitecture (Farber et al., 2011); a locus on chromosome 6 that controls tibial length in LG/J and SM/J inbred lines (Nikolskiy et al., 2015); the PAPP-A2 locus that controls bone shape and size (Christians et al., 2013), and numerous other loci that have been mapped by Cheverud and colleagues and shown to underlie limb and craniofacial skeletal traits (e.g., see Kenney-Hunt et al., 2006). Within non-human primates, fewer examples exist, but include the identification of QTLs governing craniofacial shape in hamadryas baboons (Sherwood et al., 2008); and the identification of a QTL in baboons governing variation in forearm bone mineral density (Havill et al., 2005).

GWAS, as well as other association mapping experiments, have taken advantage of a highly prevalent class of markers called SNPs, and improvements in genotyping technologies, to map population variation in traits to loci in the human genome (reviewed in Hartl and Ruvolo, 2011). GWAS involve genotyping and phenotyping a large number of individuals (tens to hundreds of thousands) from cases and controls that are age- and sex-matched and are often from similar geographic localities. Genotyping involves the use of a SNP-chip that often contains over one million SNP markers spread across the genome. The use of more markers as well as knowledge of the non-random association of alleles at different loci, or linkage disequilibrium (LD), recently acquired from the 1000 Genomes and the HapMap projects (Weir et al., 2005), has improved GWAS studies and helped to narrow down genomic intervals of association. Most recently, rare variant imputation (reviewed in Porcu et al., 2013), along with fine-mapping and refined haplotype analyses in diverse human populations (reviewed in Li and Keating, 2014) has lead to even narrower association windows (tens of thousands of kilobases). Given that most GWAS studies are conducted on complex traits that are extremely variable and highly polygenic, often thousands to hundreds of thousands of individuals will need to be analyzed to identify associated loci; for example, the most recent GWAS of skeletal height variation in human populations

was conducted on over 250,000 individuals and revealed almost 700 loci that explain at least 20% of heritable variation in height (Wood et al., 2014). Importantly, in all GWAS studies to date, the causative base pairs that control variation remain largely unknown due to the fact that many SNPs are often in strong LD within the association interval (Wood et al., 2014).

2.2.2 Functional Genomics Methods: Defining Functional Sequences

Recently, functional genomics approaches have been implemented by large-scale projects such as The Encode Project (ENCODE Project Consortium, 2012), The Roadmap Epigenomics Project (Roadmap Epigenomics Consortium, 2015), and The Fantom Project (Lizio et al., 2015), providing rich data sets upon which to screen putative associated variants from GWAS or other association studies. These projects have taken advantage of NGS to reveal genome-wide transcript production (transcriptome) and usage, as well as the locations of regulatory sequences and their interactions with target genes in a variety of cell and tissue types in humans and mice.

2.2.2.1 Genome-wide Transcript Detection

Detection of expressed transcripts has been carried out on genomic levels by surveying the transcriptome of a specific tissue or cell type (Gibson and Muse, 2009; Dong and Chen, 2013; Roux et al., 2015). Initially, this was accomplished by way of tissue dissections followed by the generation of complementary DNA (cDNA) libraries, in which double-stranded DNA is synthesized from messenger RNA extracted from a tissue of interest. When cDNA is sequenced and mapped to an assembled genome, these libraries produce a list of the expressed transcripts in that tissue (Gibson and Muse, 2009). Early cDNA libraries did yield the locations of transcripts in the genome and interesting differences in transcript abundance and variation between specific

tissues in the body and between different individuals or species (e.g., see Canavez et al., 2001). However, the advent of the DNA microarray (reviewed in Gibson and Muse, 2009), a microchip that contains DNA probes for nearly every known protein coding gene in the genome, made faster, more refined studies of transcript expression and variation possible. In these experiments, cDNA generated from a particular tissue is applied to the chip allowing the sequences to hybridize to gene probes. The hybridized transcripts are then detected using fluorescence or chemiluminescence. This method allows for a relative quantification of which transcripts are either up- or down-represented in the sample, especially when compared to appropriate control genes and tissues. In an effort to characterize the gene expression profile of developing limbs, microarray analyses have been performed on different limb bud zones (Rock et al., 2007; Schreiner et al., 2009), limb types (Shou et al., 2005), limb structures (Pazin et al., 2012), tissue types (Zhang et al., 2008; Chau et al., 2014), and even between different growth zones of developing bones (Wang et al., 2004; Horvat-Gordon et al., 2010; James et al., 2010; Lui et al., 2010).

More recent advancements, such as RNA-sequencing (RNA-seq; reviewed in Ozsolak and Milos, 2011), in which NGS is performed on RNA extracted from whole tissues, cultured cells, and even single cells (see below), provide genome-wide maps of transcriptional output including non-coding RNA transcripts. This technique can be carried out on tissues from different species, although the mapping of reads from NGS can be impacted by the quality of the parent genome. To date, RNA-seq has been performed in a number of different contexts, including on a variety of skeletal tissues and cell types, such as on chondrocyte cell lines and osteoblasts (ENCODE Project Consortium, 2012; Bowen et al., 2014; Mori et al., 2014; Oh et al., 2014; Lizio et al., 2015; Roadmap Epigenomics Consortium, 2015). While these experiments have revealed novel loci that likely reflect cartilage and bone-specific biology, it is important to note that they have been conducted mostly on in vitro derived cell lines

and therefore transcriptomic profiles may be quite dissimilar to profiles acquired from in vivo collected tissues, such as chondrocytes of the growth plate.

Historically, these transcriptomic techniques were more often than not performed on heterogeneous cell populations extracted from a single organ or tissue. However, techniques such as fluorescence-activated cell sorting (FACs) (reviewed in Tung et al., 2007) and laser capture microdissection (reviewed in Datta et al., 2015) now allow for the isolation of specific cell types for use with these functional genomics methods. These approaches enable the identification of all transcripts produced by that cell or tissue type at a given time point. Most recently, microfluidic techniques that permit the isolation of single cells and subsequent "barcoding" of their RNA, permit rapid, single-cell resolution transcriptomics (Macosko et al., 2015).

Using these types of refined data sets, scientists will be able to address a number of issues. For example, they will be able to determine whether GWAS SNPs for height are enriched near genes expressed uniquely within a specific growth plate zone (e.g., in the proliferative zone versus the hypertrophic zone) and this could reveal specific mechanisms that evolution has targeted to drive skeletal variation in human populations. Comparative RNA-seq analyses performed on specific tissues, such as growth plates, or on specific growth plate zones, from several different species with different limb phenotypes will likely shed light on the molecular and evolutionary mechanisms that generate interspecies variation in limb length, segment length, and limb proportions. For example, Cooper and colleagues (2013) have demonstrated that hypertrophic chondrocyte zones from the elongated distal metapodial growth plate of the jerboa, a hopping rodent, are enlarged by over 50% compared to the laboratory mouse. This suggests that genes and regulatory sequences controlling hypertrophic zone formation, differentiation, and maintenance may have been the target of selection during jerboa evolution; if so, it may be interesting to identify

whether similar genetic changes are seen in other hopping or jumping rodents with similar distal metapodial elongations. Hypertrophic cells from species with convergent phenotypes may have very similar transcriptome profiles and be quite dissimilar to more closely related (non-elongated) sister species. On the other hand, hypertrophic chondrocyte zones in each hopping rodent species may exhibit significant transcript expression divergence due to the evolution of distinct hierarchical factors regulating growth plate transcriptional programs. In this case, knowledge of underlying genetic variation within and between species will be paramount for identifying mutations linked to expression differences in these rodents.

Transcriptome experiments, such as those described above, can help bridge the genotype–phenotype gap, especially when they are coupled with direct measurements of genome sequence variation in the same individuals. Accordingly, expression quantitative trait loci or eQTLs (reviewed in Gilad et al., 2008; Majewski and Pastinen, 2011) are loci that show expression variation in relation to underlying genomic variation. SNPs have historically been the primary type of genomic variation examined, but more recently insertion/deletion mutations have been analyzed (Huang et al., 2015). For example, RNA-sequencing and SNP genotyping performed on HapMap and/or 1000 Genomes Project lymphoblastoid cell lines have revealed SNPs that are associated with transcript variation for nearby genes (e.g., Lappalainen et al., 2013). These eQTLs will help to narrow down causative mutations among the many associated variants uncovered via GWAS and are useful for honing in on variants within QTL intervals.

2.2.2.2 Genome-wide Regulatory Element Detection

One insight gained from The Encode Project (ENCODE Project Consortium, 2012), The Roadmap Epigenomics Project (Roadmap Epigenomics Consortium, 2015), and The Fantom Project (Lizio et al., 2015) is that most genomic variation (including

eQTLs) resides in close proximity to, or directly overlaps with, a sequence with known biochemical function, the majority of which is regulatory in nature. This finding suggests that regulatory portions of the genome are important drivers of developmental variation and evolutionary change. Recent advancements in the genome-wide identification of regulatory sequences are improving our understanding of this domain, specifically with respect to the spatiotemporal control of gene regulation in a cell-type specific manner.

Techniques that reveal regulatory sequences (e.g., enhancers, repressors, promoters) across the genome have been based on several fundamental observations. First, it has been known for some time, via the development of chromatin immunoprecipitation (ChIP) (reviewed in Orlando, 2000), that regulatory sequences are physically bound by transcription factors (TF), which facilitate the expression of target genes (reviewed in Krebs et al., 2014). Some TFs have been shown to act in a general manner; that is, they bind to the same regulatory sequence in many different cell types (e.g., P300) (e.g., Visel et al., 2009). Other TFs have been shown to be functional only in a particular cell type, bound to many targets, and/or responsible for hierarchically controlling its transcriptional profile (e.g., Aziz et al., 2010). A typical hypothesis that emerges from these findings is that if a TF is shown to interact biochemically with a regulatory sequence, then a DNA modification (e.g., SNP) at the specific location of binding may reduce or enhance TF binding, producing a functional impact on gene transcription and phenotype. Importantly, recently ChIP has been combined with NGS (i.e., ChIP-seq) to identify all locations in the genome bound by a specific factor (reviewed in Furey, 2012). Projects such as The ENCODE Project (ENCODE Project Consortium, 2012) have performed ChIP-seq on over 125 different human cell types for a number of transcription factors, some of which universally mark enhancers and others that mark enhancers for specific cell types.

ChIP-seq studies have only recently been performed on tissues related to limb and skeletal development. Menke and colleagues performed ChIP-seq on early mouse limb buds for the TF Pitx1 (Infante et al., 2013). As Pitx1 is a hindlimb-specific regulator (Lanctôt et al., 1999), their work has revealed many targets genome-wide that likely regulate hindlimb development in mammals; such targets can be screened for mutations that potentially underlie variation in hindlimb morphology across mammals and primates. ChIP-seq has also been performed on skeletogenic tissues for Sox9 (Oh et al., 2014), a hierarchical regulator of mesenchymal condensation and early chondrocyte development in long bones (Bi et al., 1999; Ohba et al., 2015). This data set of regulatory sequences can be used to identify whether any GWAS variants for height reside within and thus potentially disrupt chondrocyte enhancer function. Likewise, comparative genomic sequence analyses on such an enhancer data set can be used to reveal suites of enhancers that have experienced evolutionary changes between species that have different long bone skeletal phenotypes.

Second, it has also been discovered that chromatin that is wound around histones in the form of nucleosomes is actively unwound before transcription (reviewed in Krebs et al., 2014). Importantly, where chromatin is unwound or "open," it can then be experimentally digested using nucleases, enzymes that cut DNA (Wu et al., 1979a,b; Gross and Garrard, 1988). Numerous assays have been developed that use NGS to sequence "open" digested nucleosome sequences (reviewed in Meyer and Liu, 2014) with the goal of identifying the locations of these potential regulatory sequences across the genome. These techniques include: DNase-seq (Crawford et al., 2004; Sabo et al., 2004), FAIRE-seq (Giresi et al., 2007), and ATAC-seq (Buenrostro et al., 2013). DNase-seq involves the digestion of "open" chromatin by the enzyme DNase I followed by sequencing (Crawford et al., 2004; Sabo et al., 2004). DNase-seq protocols performed on hundreds of cell and tissue types have revealed millions of active regulatory regions across the genome (reviewed in Madrigal and Krajewski, 2012) and about one-third of

these DNase sites are specific to individual cell types, reflecting the cell-type-specific control of gene regulation (ENCODE Project Consortium, 2012). To date, DNase-seq has been performed on mouse fore- and hindlimb buds at gestational day (E) 11.5 revealing a number of loci that regulate early limb bone patterning; on mouse limb buds at E14.5 identifying loci involved in limb chondrogenesis (ENCODE Project Consortium, 2012), and on in vitro-derived osteoblasts to study bone development (Inoue and Imai, 2014; Tai et al., 2014).

Of these approaches, ATAC-seq, or assay for transposase-accessible chromatin and sequencing (Buenrostro et al., 2013), is quite promising. This approach uses a specific transposase that preferentially digests open chromatin regions and simultaneously integrates built-in adaptor tags to the ends of the digested sequence that coincide with the regions adjacent to nucleosomes. Using primers that recognize these tagged sites, polymerase chain reaction (PCR) is then performed to amplify the library, which then undergoes NGS and mapping of reads to the parent genome to reveal regulatory sequences. This approach is quite promising because it is relatively simple, that is, it can be accomplished in less than a day, and requires very little starting cellular material. This latter property of ATAC-seq allows scientists to identify regulatory regions on rare tissue samples (for example, human embryonic limb buds) or even highly specific anatomies or cell types. While ATAC-seq has not yet been used on skeletal tissues, one elegant study by Shubin and colleagues (Gehrke et al., 2015) has compared genome-wide ATAC-seq signatures derived from mammalian limb-bud and fish fin-bud tissues to understand the evolution of gene regulation during appendage development over deep evolutionary time.

Third, chromatin wound around nucleosomes experiences chemical modifications that lead to its relaxation and thus potential for active transcription (reviewed in Krebs et al., 2014). At the nucleosome, core histone proteins possess exposed amino acid residues

or tails that can be chemically modified or "marked" by a number of processes including acetylation and methylation (reviewed in Rivera and Ren 2013). Specific histone tail residues, when acetylated, can lead to the loosening of a chromatin–histone complex, which exposes DNA for future occupancy by transcription factors. On the other hand, methylation at specific histone tail residues can lead to repression and a highly wound DNA–histone complex. Thus, based on the type of mark, it has been possible to identify using ChIP-seq active and repressed regulatory sequences on a genome-wide and cell-type specific level. Some of the most studied histone marks include: H3K27ac, a mark of active enhancers (Heintzman et al., 2009; Creyghton et al., 2010; Rada-Iglesias et al., 2011); H3K4me3, a mark of active promoters (Bernstein et al., 2005; Pokholok et al., 2005); and H3K27me3, a mark of a repressed region (Bernstein et al., 2006). The ENCODE Project (ENCODE Project Consortium, 2012) has studied a series of these marks in human cell lines and tissues related to limb development, while the Roadmap Epigenomics Project (Roadmap Epigenomics Consortium, 2015) has focused on a variety of human fetal and adult tissues. With respect to skeletal development, each consortium has performed a number of experiments for different histone marks on chondrocytes and osteoblasts derived from adult long bone marrow cavities. These data sets have not yet been extensively examined.

All three of the approaches discussed above yield regulatory sequences which can be screened and/or filtered for mutations that either have been associated with trait variation, or are different between two species or individuals under study. In addition, these techniques provide an understanding of the localized control of gene expression via the identification of regulatory sequences within a candidate locus. Thus, they can be used to substantially narrow down the number of putative regulatory mutations that need to be considered between affected and unaffected individuals in candidate gene studies or in association intervals. Because these data sets are new, there have not been many examples of their use in the above applications. However, one recent

study has revealed that specific enhancers for SOX9 are physically removed from its coding region due to chromosomal inversions in patients with limb defects such as campomelic dysplasia (Gordon et al., 2009b; Fukami et al., 2012). It is through the identification of the enhancers and their physical displacement that the causative mutations underlying this phenotype were discovered.

2.2.2.3 Intrachromosomal Interactions between Regulatory Elements and Genes

In order to connect regulatory elements to their specific target genes, and thus be able to understand how mutations within them impact gene expression and phenotypes, assays that gauge biophysical connections, known as chromosomal conformation capture assays, have been developed (reviewed in Rivera and Ren, 2013). These techniques can be performed in a localized manner to reveal intralocus interactions (i.e., chromosomal conformation capture (3C) (Dekker et al., 2002) or chromosomal conformation capture carbon copy (5C) (Dostie et al., 2006)), or on a broader genome-wide level (i.e., circular chromosome conformation capture (4C) (Zhao et al., 2006), chromatin interaction analysis using paired-end tag sequencing (ChIA-PET) (Fullwood et al., 2010), or Hi-C (Lieberman-Aiden et al., 2009)). One common finding of all of these capture techniques is that regulatory sequences can target more than one gene during development, making understanding the regulatory impacts of sequence variants quite complicated. To date, few studies have been conducted using capture assays for limb and skeletal development, although some notable examples exist. Amano and colleagues (2009) used 3C to identify intralocus interactions in the Sonic hedgehog (SHH) locus, revealing important expression kinematics of the SHH protein in the limb bud. Mutations in SHH have been shown to alter digit morphology and likely explain some variation in digit number in different animals (see below). 3C was also used on the SHOX locus and revealed that disruptions in enhancer interactions may underlie some cases of idiopathic short stature (Benito-Sanz et al.,

2012). Finally, Lupiáñez and colleagues (2015) used 4C assays on patients with brachydactyly and polydactyly and revealed that genomic disruptions (via deletions, inversions, or duplications) to intrachromosomal interactions between enhancers and promoters for the WNT6;IHH;EPHA4;PAX3 locus are likely the causative mutations underlying these skeletal phenotypes. In time, these methods will be used in a comparative framework to reveal how intrachromosomal interactions vary between species displaying considerable variation in skeletal morphology.

2.2.3 Developmental Biology Methods: Testing Putative Functional Variants and Understanding their Developmental Context and Phenotypic Impacts

While genetic mapping experiments are ideal for identifying loci associated with phenotypic traits, and functional genomic studies can help narrow down association and QTL intervals to a smaller number of putative functional mutations, both lack the ability at this time to directly test mutations for functional impacts on phenotypes. To acquire this level of understanding for a given variant (or region of the genome that markedly differs between individuals or species), developmental biology and molecular biology methods must be utilized, and thus far they have been important for understanding: (1) the identification of the spatiotemporal patterns of expression for transcripts and proteins in specific tissues and cells; (2) the nature of signaling interactions between and within tissues and cells; (3) the tracking of cellular contributions to developing and mature tissues; (4) the identification of specific regulatory sequences for genes used during development and growth; and (5) the functions of gene and regulatory elements in the embryo at specific times and places. Below, each of these contributions is discussed in the context of skeletal development and in the ways they have been helpful in linking genotypes to phenotypes.

2.2.3.1 Detecting Gene Expression and Protein Localization

For a coding or regulatory variant to impact morphological variation, the mutation will likely alter the expression and/or function of the gene in a specific cell type of interest. Therefore, methods have been developed that allow for the precise identification of the spatiotemporal expression pattern of the gene and its protein product. The earliest expression detection methods involved the qualitative detection of just one individual gene or protein product either on a histological section or in a whole embryo using in situ hybridization (reviewed in Hauptmann, 2015) or immunohistochemistry (reviewed in Buchwalow and Böcker, 2014). In these cases, the endogenously expressed transcript or protein of interest is first targeted using either a labeled-mRNA probe or antibody, and once the probe is hybridized to the transcript, or the antibody is bound to the protein, they are then detected via a colorimetric or fluorescence reaction. Assays such as these, when simultaneously performed for cell-type specific marker genes or proteins, and/or in conjunction with histological staining techniques (e.g., hematoxylin and eosin staining), allow the researcher to define specific expression zones at the level of cells and tissues (see Chapter 9 for an example). These methods provide important information about where and when a gene is expressed and, if used in the context of gene loss-of-function or mis-expression experiments (see below), they can serve to reveal how the expression of downstream genes and specific molecular pathways are disrupted (e.g., see Capellini et al., 2006, 2010). They can also be used to determine whether species-specific mutations in regulatory elements lead to transcript down- or upregulation at a localized subdomain of an entire gene's expression pattern (e.g., see Shapiro et al., 2004).

2.2.3.2 Identifying How Signaling Interactions Sculpt Phenotypes

Because phenotypic variation arises during development, it is important to understand how genetic mutations alter, ever so finely, signaling interactions within and between tissues. A key component of this approach is characterizing these interactions in model systems, so that the functional ramifications of mutations can be contextualized, especially as they lead to the production of variation. Indeed, some of the earliest studies in developmental biology used model systems such as the chick and the axolotl to examine tissue interactions during development and how they control the formation of distinct tissues (e.g., see Gilbert, 2013). These experiments involved either placing foil barriers between tissues of interest or removing tissues to alter signaling and observe the phenotypic consequences on the development of appendage skeletal elements (e.g., see Summerbell, 1979; Stephens and McNulty, 1981). For example, placement of a foil barrier between the lateral plate mesoderm and somitic dermomyotome at the level of the chick forelimb bud lead to the downregulation of bone morphogenetic protein expression and severe scapula blade phenotypes. Likewise, removal of the somites at the forelimb and hindlimb levels has revealed that signaling interactions between the somites and lateral plate mesoderm are important for the formation of the scapula but not necessarily the pelvis (Huang et al., 2006). These experiments also reveal that one potential source of variation in scapula blade morphology is from the actions of genes in adjacent non-scapula tissues.

Another early experimental approach involved grafting tissues from one part of the body onto another to observe the phenotypic effects of disrupted signaling. For example, early studies in chick limb development revealed that when a specific subpopulation of posterior limb bud cells was grafted onto the anterior portion of a similarly staged limb bud, an ectopic mirror image digit duplication occurred (MacCabe et al., 1973). Furthermore, the grafting of the homologous mouse tissues onto the anterior chick limb

bud resulted in similar patterns (Tickle et al., 1976), revealing a conservation of digit patterning mechanisms. These early experiments identified an important signaling center called the zone of polarizing activity (ZPA) that is now known to express Shh, which encodes for a protein that signals across the limb bud to drive posterior digit formation (Riddle et al., 1993). Researchers have also used beads soaked in specific proteins or chemical antagonists to explore how signaling occurs within and between tissues. For example, the application of SHH protein-soaked beads to the anterior portion of the chick limb bud yielded mirror image duplications, thus reproducing the result of ectopic application of ZPA tissues (Yang et al., 1997). In addition, distal limb truncations and digit loss occurred in response to the application of SHH antagonists to the ZPA region, revealing the important role of this pathway in digit patterning and outgrowth (Scherz et al., 2007). Beads soaked in bone morphogenetic proteins, such as GDF5 or its antagonist, have helped to reveal their role in the formation of synovial joints (Francis-West et al., 1996; Merino et al., 1999), whereas the application of chemical inhibitors to developing growth plates has demonstrated the important effects of signaling interactions in endochondral ossification (e.g., Nagai and Aoki, 2002; Wu and De Luca, 2006).

2.2.3.3 Fate-mapping

Many protocols have been designed to track or fate-map cellular contributions to distinct tissues during development, thereby providing a context for understanding how functional mutations in specific cells can influence morphological development and variation. For example, studies in chick embryos using DiI labeling have fated cell populations in the early limb bud to different signaling centers as well as to proximal and distal skeletal elements (e.g., Vargesson et al., 1997; Dudley et al., 2002). They have also shed light on the cellular progenitors of both synovial joints (Koyama et al., 2008) and muscles of the limb (e.g., Pacifici et al., 2006). Interspecies (quail–chick) cell

labeling experiments, which take advantage of the use of species-specific antibodies for detection, have also revealed the migratory pathways of muscle cells in the limb (e.g., Valasek et al., 2005), as well as the tissues that give rise to both girdles (reviewed in Huang et al., 2006).

The laboratory mouse has been an exceptionally important model system for tracing cell lineages during development (reviewed in Kraus et al., 2014). Using gene-targeting procedures, researchers have created mouse strains harboring specific regulatory sequences placed upstream of the Cre gene to drive its expression. The protein product of this gene is the Cre-recombinase enzyme, which can excise artificial sequence tags called loxP sites. For the purpose of fate-mapping, loxP sites have been engineered to flank "stop" sequences upstream of regulatory elements capable of constitutively activating a reporter gene, such as one that produces lacZ or green fluorescent protein (GFP) (e.g., Soriano, 1999). Mouse lines harboring this reporter sequence can be crossed to a line with a tissue-specific regulatory sequence driving the expression of Cre-recombinase. The enzyme will then excise the stop sequence via recombination at the loxP sites, thus activating lacZ or GFP gene expression in only that cell type. Using detection methods, scientists can then track lacZ- or GFP-labeled cells as they proliferate and migrate to the tissues they help form. For example, this procedure has been used to fate cells that have expressed Shh to posterior digits (Harfe et al., 2004), expressed Gdf5 to developing and mature joints (Koyama et al., 2008), and expressed Sox9 to chondrogenic populations in the growth plates (Akiyama et al., 2005).

2.2.3.4 Regulatory Element Identification

The discovery that phenotypic variation within and between species is due to regulatory mutations (King and Wilson, 1975), which has been recently documented by genome-wide studies in humans (Grossman et al., 2013) and stickleback fish (Jones et al., 2012), has lead scientists to interrogate conserved non-coding sequences for regulatory

function using in vitro or in vivo assays (reviewed in Davidson 2001). In these assays, candidate non-coding sequences are cloned upstream of reporter genes (lacZ or GFP) to see if they can activate reporter expression when transfected in cells or injected into living embryos. While cell assays have been useful in determining that a sequence has regulatory potential, in vivo approaches, such as enhancer transgenesis, reveal a sequence's precise spatial–temporal control of gene regulation in the three-dimensional embryo (Mortlock et al., 2003). For example, based on its high sequence conservation in mammals, a specific regulatory enhancer for Shh was discovered that drives this gene specifically in the ZPA (Lettice et al., 2003) in mouse embryos. Interestingly, single point mutations in this enhancer in mice, cats, and humans each result in an extra digit in the forepaw and hindpaw (Gurnett et al., 2007; Furniss et al., 2008; Lettice et al., 2008; Sun et al., 2008), demonstrating the role of regulatory mutations underlying phenotypic diversity. Recently, specific long bone and joint regulatory sequences for the Gdf5 (Capellini, unpublished) and Gdf6 (Mortlock et al., 2003) genes, along with several other musculoskeletal and limb genes such as Bmp5 (Guenther et al., 2008), Fgf8 (Marinić et al., 2013), and Myf5 (Summerbell et al., 2000) have been discovered also based on strong vertebrate conservation profiles. Importantly, these sequences control variation, as for example, mutations in a long-bone growth plate specific GDF5 regulatory element underlies limb length variation in humans (Capellini et al., unpublished).

2.2.3.5 Identification of Gene and Regulatory Element Function

A major step in connecting genotype to phenotype is determining the biological function of the RNA transcript, regulatory region, and/or specific DNA base-pair mutation under study. Numerous protocols have been developed that allow for the targeted interference or alteration of a DNA or RNA molecule in vivo (reviewed in Behringer et al., 2014). Some of these assays, initially performed in chicks and mice,

focused on using small hairpin RNA molecules (shRNAs) and the RNA interference pathway to knock down transcript level in order to reveal a gene's developmental role (reviewed in Campeau and Gobeil, 2011). Other techniques involved gene mis-expression by injecting replication-competent retroviral vectors (RCAS) possessing complementary DNAs for target genes (reviewed in Gordon et al., 2009a). For example, targeted RCAS mis-expression for the genes TBX4, TBX5, and PITX1 helped identify their roles in the determination of chick limb identity and early outgrowth (Logan and Tabin, 1999; Rodriguez-Esteban et al., 1999; Takeuchi et al., 1999), whereas similar assays for *Ihh*, *PTHrP*, *Wnt*, *Bmp*, and *Notch* revealed their actions during growth plate chondrocyte biology (Vortkamp et al., 1996; Zou et al., 1997; Hartmann and Tabin, 2000; Church et al., 2002; Provot et al., 2006).

Gene targeting directly in mice has long been the gold standard for identifying how a specific DNA mutation influences biological function (reviewed in Menke, 2013). The earliest techniques allowed for the removal of a target sequence (known as a "knockout"), or a replacement of an endogenous mouse sequence with a foreign sequence (known as a "knockin"). These early techniques revolved around the process of homologous recombination, which allowed scientists to use the cell's own repair machinery in concert with foreign sequence constructs to replace a targeted sequence with a new sequence or none at all. Some of the earliest "knockout" and "knockin" alleles were performed on loci involved in limb skeletal development. For example, knockout of the *Shh* gene or its long-range regulatory element lead to a severe digit reduction and truncation, which roughly phenocopied early tissue removal and bead experiments in chicks (Chiang et al., 1996; Sagai et al., 2005). Many genes involved in growth plate regulation and joint formation have been experimentally excised using this approach (see Decker et al., 2014; Kozhemyakina et al., 2015). When data generated using gene targeting are coupled with tissue expression and other functional studies, not only has the understanding of the molecular circuitry of limb development, growth

plate function, and joint formation been greatly expanded, but now scientists have the ability to contextualize how sequence mutations associated with phenotypes (and apparent within/between species) impact phenotypic variation.

Recently, techniques in genome editing using zinc finger nucleases (Urnov et al., 2010), transcription activator-like effector nucleases (TALENs) (Christian et al., 2010), and clustered regularly interspaced short palindromic repeats (CRISPR) with the CRISPR-associated system (CRISPR-Cas9) (Jinek et al., 2012) have afforded scientists more rapid, cost-effective, and artifact-free ways of making targeted mutations at a locus (reviewed in Gupta and Musunuru, 2014). Of these three techniques, CRISPR-Cas9 editing has proven to be the most popular due to its higher efficiency, greater ease, and lower cost. This technique works when short guide RNA (sgRNA) molecules, which are complementary to the targeted DNA region of interest, are artificially expressed together with a Cas9 nuclease in a cell (reviewed in Zhang et al., 2014). The guides lead Cas9 to the specific DNA site of interest, so that it can cut and induce a double-strand break. CRISPR-Cas9 has been used to create frameshift mutations resulting in premature stop codons and loss-of-function of key developmental genes (e.g., Fossat et al., 2015), as well as to excise enhancers (e.g., Zhou et al., 2014). CRISPR-Cas9 has also been used to "knockin" human sequence (see Zhang et al., 2014), which has revolutionized the study of the functional basis of normal and rare variation underlying animal phenotypes (e.g., Gennequin et al., 2013). Techniques have improved so rapidly that CRISPR-Cas9 can now be used to generate mutations at multiple loci simultaneously (Wang et al., 2013) and it has been adapted to study large-scale structural changes in chromosomes (Kraft et al., 2015) and to recreate human structural rearrangements in the mouse model (Lupiáñez et al., 2015). More recently, CRISPR-Cas9 has been used to manipulate the genomes of other organisms, including humans (Liang et al., 2015), non-human primates (Niu et al., 2014; Chen et al., 2015; Wan et al., 2015), and rats (Shao et al., 2014). CRISPR-Cas9 is also

being used extensively *in vitro*, such as in human, primate, and mouse cell lines, to characterize the effects that coding and regulatory mutations have on cellular phenotypes (see Zhang et al., 2014).

2.3 Anthropological Perspectives

The last 10 years have witnessed remarkable progress in the development of a number of important genetic, genomic, and developmental biology tools that help to identify and test the specific base-pairs that control phenotypic variation. Given the specialized nature of these techniques, there has not been a single study to date that has combined all approaches to link a genotype to a skeletal phenotype. However, one landmark project involved the discovery of a recurrent adaptive genotype in different freshwater stickleback populations (Shapiro et al., 2004; Chan et al., 2010). In this study, the causative allele, an approximate 500-bp enhancer deletion near the *Pitx1* gene, was identified due to the use of extensive genetic mapping, population genetic approaches, and functional tests in vivo. These approaches helped identify the *Pitx1* locus as a driver of variation in pelvic morphology in sticklebacks, discover freshwater stickleback populations with *Pitx1* haplotypes harboring recurrent enhancer deletions, reveal that these haplotypes were under selection, and focus stickleback transgenic and knockin studies to pinpoint the functional adaptive base-pair deletion controlling pelvic fin loss. This discovery was borne out of the development of several genetic, molecular biology, and developmental biology tools by David Kingsley's laboratory over the last 15 years that has made the stickleback a model system to study the evolutionary mechanisms controlling phenotypic variation.

For practical, ethical, and monetary reasons, Evolutionary Developmental Anthropologists are unlikely going to acquire many of these tools and samples for their primates of interest. For example, given the endangered status of many primates, there will

likely always be a dearth of embryonic, infant, and juvenile tissues for functional genomic and developmental biology studies. Even if these samples were made available to a few experts (as in Chapters 4 and 6), they may not be high enough in number to meet the requirements for biological replication. In fact, the most cutting-edge functional genomics studies to date were performed on heterogeneous tissues from only a few individuals of each primate species, each of which differed dramatically in sex, age, health status, and post-mortem processing (e.g., see Khaitovich et al., 2005; Perry et al., 2012). Another example includes the difficulties in acquiring enough wild (and/or captive) individuals in large numbers and of diverse genetic makeup to perform association mapping experiments.

There may be some improvements on some of these fronts in the next several years, although they will only improve a portion of the tool types discussed above. For example, regarding sample size, one possible remedy is the establishment of a unified international system that fosters the opportunistic acquisition of tissues from animals that have died due to injury, illness, or during pregnancy. In the proper organizational setting, low cell number RNA-seq (Ozsolak and Milos 2011) and ATAC-seq (Buenrostro et al., 2013) protocols can produce important functional genomic data sets from a host of different tissues and cell types from a single animal. Another improvement will be the continued generation, expansion, and use of lymphoblastoid cell lines (LCLs) and induced pluripotent stem cells (IPS) from multiple individuals of a single species (e.g., see Khan et al., 2013; Gallego Romero et al., 2015). These lines will allow for functional genomic studies, although their use will be of limited value to directly understanding three-dimensional skeletal morphology. To strengthen genetic mapping experiments in primates, large-scale efforts currently underway through a number of projects, such as the International Vervet Research Project (IVRP) (Jasinska et al., 2012); the Southwest National Primate Research Center (SNPRC) (https://www.txbiomed.org/primate-research-center); the UC Davis California National Primate Research

Center (UCDCNPRC) (http://www.cnprc.ucdavis.edu/); the Cayo Santiago Rhesus Macaque Study (CSRMS) (e.g., Widdig et al., 2016); and the German Primate Center (GPC) (http://www.dpz.eu/en/home.html). These projects all hold great promise for increasing colony size, expanding knowledge of genetic diversity, improving animal pedigrees, and enhancing the potential for mapping of genotype to phenotype in captive and wild settings.

As these endeavors will improve some of the aforementioned tools, Evolutionary Developmental Anthropologists will still be forced to use model systems to understand the developmental and genetic basis of variation and this poses some important issues. For example, recent comparisons of The ENCODE Project (ENCODE Project Consortium, 2012) the Mouse ENCODE Project (Yue et al., 2014), and smaller-scale (e.g., Cotney et al., 2013) data sets have revealed considerable divergence in the regulatory genomes between humans and mice, suggesting that the use of the mouse as a model system may lead to erroneous claims on how a locus controls variation. Yet, for a portion of the regulatory genome, that is, orthologous regulatory sequences that remain syntenic to their putative target genes, there appears to be significant functional conservation (Yue et al., 2014). Additionally, Yue and colleagues (2014) and other researchers (Cheng et al., 2014; Stergachis et al., 2014) have discovered that human and mouse transcription factor binding at target loci, and the associated genetic networks they control, are substantially more conserved than previously believed, indicating that important molecular cascades remain intact from mouse to human. Both findings support the continued use of the mouse to gauge the functional impacts of mutations that alter conserved sequences have on phenotypic variation; although they indicate that the use of the mouse as a model system for functional genomics studies should be carefully undertaken and in a comparative framework. For this reason, recent studies cataloging transcript and regulatory element usage in the mouse have also performed assays on precious human and macaque embryonic tissues (Cotney

et al., 2013; Reilly et al., 2015). While this level of experimentation is preferable, it is not very feasible or likely because the acquisition of samples is unpredictable and difficult. In the absence of the appropriate primate tissue data sets, detailed comparisons of novel mouse data sets with those compiled across many tissues from The ENCODE Project (ENCODE Project Consortium, 2012) and The Roadmap Epigenomics Project (Roadmap Epigenomics Consortium, 2015) should be helpful in revealing the context of important mutations in primates.

Regardless of these suggested difficulties, the laboratory mouse has already had tremendous utility, especially in the context of testing the functional consequences of mutations that differ between human populations and/or primate species. For example, Kamberov and colleagues (2013) generated a coding mutation in the human *Ectodysplasin Receptor* (*EDAR*) gene in mice to recreate a human mutation that is high in frequency in Chinese populations and displays strong evidence of past selection. This mutation resulted in several human-like phenotypes in mice, including morphological changes in hair thickness, mammary biology, and eccrine gland density. Mice will also continue to provide a unique in vivo three-dimensional mammalian context for interpreting and/or characterizing morphological differences that result from specific regulatory mutations. For example, by comparing the transcriptome of human and mouse progenitor cell populations in the cortex, Florio and colleagues (2015) identified over 50 genes that were preferentially expressed in humans. One specific human gene, *ARHGAP11B*, when knocked into the orthologous mouse locus, increased basal progenitor generation and self-renewal and increased cortical plate area and gyrification, likely underlying some of the major developmental changes in human brain evolution. Indeed, as more and more gene transcripts and regulatory enhancers are discovered that are found to be expressed in and/or control highly specific anatomy, the mouse will be the only system available to assess how mutations in regulatory elements influence anatomy.

Another major issue that Evolutionary Developmental Anthropologists are up against concerns the acquisition of adequate phenotypic data within and between species. While thousands of papers have been published on morphological variation in primates offering tremendous insight into primate biology, most if not all have not acquired genotypic information and many have measured the same trait differently, which makes combining phenotypic data sets for future association mapping studies difficult, misleading, and highly problematic. In addition, measured phenotypes do not necessarily reflect important developmental information, and their use in genetic mapping experiments or comparative genomic studies may lead to many false positives. Fortunately, there have been some recent attempts to remedy these issues. Centers like the IVRP, SNPRC, UCDCNPRC, CSRM, and GPC, along with large-scale phenotype data acquisition and cataloging by Morphobank (www. morphobank.org), are starting to collect the relevant information and standardize measurements.

Finally, for studies to be informative on an evolutionary level, improvements must also be made in addressing how functional mutations reflect and have affected the fitness landscape (Barrett and Hoekstra, 2011; Vitti et al., 2013). While there is agreement that this last aspect is critically important, being able to identify the adaptive value of a specific phenotype encoded by a genetic mutation is extremely difficult. There are a few experimental techniques that help reveal how natural selection directly impacts variation at any functional, putatively adaptive variant and some experiments have laid the foundations for how these studies may be conducted (reviewed in Pardo-Diaz et al. 2015). Other than these insights, many scientists have reverted to a "reverse genomics" approach that relies on genome-wide scans to identify loci that display characteristic signatures of natural selection. One comprehensive measure developed by Pardis Sabeti and colleagues, called the composite of multiple signals (CMS) score (Grossman et al., 2013), involves the integration of genome-wide datasets from five different selection methods into one score. CMS has been performed on human

genomes spanning different continents and populations and has produced lists of candidate regions in the modern human genome, each of which can be functionally interrogated using all of the methods described above (e.g., see Kamberov et al., 2013). However, these tests do not reveal loci that have experienced selection between species (e.g., between humans and chimpanzees from a last common ancestor), and further improvements in methods that can detect such signatures are highly needed.

As powerful as it is, the CMS test statistic (or any approach that aims to reveal selected regions of primate genomes) relies on sequence and SNP data sets generated on many individuals (e.g., the 1000 Genomes Project). In human genomics, data sets like these are the tip of the iceberg, and they will be greatly expanded upon to include tens of thousands to millions of genomes in the next 5–10 years. Unfortunately, such expansive data sets have not been generated for primates, although there are a few projects that have augmented the number of sequenced genomes. For example, the Great Ape Genome Diversity Project has sequenced genomes from approximately 100 hominoids, and this has improved our understanding of hominoid phylogenetic history, allele sharing, genome-wide methylation patterns, chromosome evolution, and species divergence at the sequence level (Hernando-Herraez et al., 2013; Prado-Martinez et al., 2013; Sudmant et al., 2013; Nam et al., 2015). However, for each sequenced individual, phenotype data are missing. Additionally, the sample size for each hominoid species is still not large enough to perform CMS testing or tests to identify fixed, selected regions of the genome. Given that for African apes we have genetic resources (e.g., linkage maps (e.g., Auton et al., 2012)), reference genomes (Chimpanzee Sequencing and Analysis Consortium, 2005; Prüfer et al., 2012; Scally et al., 2012; Xue et al., 2015), samplings of genomic or exomic diversity (e.g., Prado-Martinez et al., 2013; Bataillon et al., 2015), biological reagents (e.g., iPS and LCLs cells (Khan et al., 2013; Gallego Romero et al., 2015)), molecular tools such as microarrays (e.g., Khaitovich et al., 2004), transcriptomic data sets (e.g., Khaitovich et al., 2005; Perry et al., 2012)

and phenotypic data sets (e.g., 100+ years of field observation and sample collection at multiple sites and museums), we should be focusing on large-scale genomic projects that capture nucleotide variation, haplotype diversity, and phenotypic data in thousands of remaining apes from around the world. These data sets may permit novel "reverse genomics" scans for selection, the possibility of revealing broadly mapped loci, and they will help partially transform these wild primates into "model-like" systems so that we can explore adaptation and the genotype–phenotype relationship in animals very pertinent to understanding the evolution of the human condition.

CHAPTER 3

A ROBUST METHOD FOR RNA EXTRACTION AND PURIFICATION FROM A SINGLE ADULT MOUSE TENDON

This chapter has been published as an article in PeerJ: Grinstein M*, Dingwall HL*, Shah RR, Capellini TD, Galloway JL. 2018. A robust method for RNA extraction and purification from a single adult mouse tendon. PeerJ, 6, e4664.

3.1 Introduction

Tendon injuries are common problems for active individuals and the aging population (Kaux et al. 2011). Treatment options include physical therapy and surgical intervention, but pain and limited mobility often persist, making complete restoration of tendon function challenging (Nourissat et al. 2015). Our current understanding of the molecular and cellular pathways regulating tendons during homeostasis, healing, and aging are limited. Several studies using large animal models such as sheep, rabbits, and rats have provided important information about tendon injury, biomechanics, surgical techniques, and bioengineering strategies for tendon repair (Voleti 2012). Other studies have used mouse genetics to gain an understanding of the molecular and cellular response of tendons to acute injuries, changing load environments, and in gene loss-of-function models (Dunkman et al. 2014; Dyment et al. 2014; Howell et al. 2017; Mendias et al. 2008; Wang et al. 2017). The mouse system offers unique advantages for implementing mechanistic studies of tendon biology as they permit genetic lineage tracing and conditional knockout strategies, and they can be housed simply and in large numbers to improve sample sizes for functional studies. Even with inbred mouse

strains, inter-animal variation can affect the conclusions drawn from gene expression analyses (Sultan et al. 2007; Watkins-Chow & Pavan 2008). Therefore, the use of several biological replicates of tendon tissues obtained from individual mice for RNA analysis is essential for furthering our mechanistic understanding of tendon biology.

Mature tendons are comprised of type I collagen, which are arranged in a highly ordered hierarchical manner along the long axis of the tissue (Kannus 2000). Tendon cells lie between these organized fibrils and are surrounded by a hydrophilic, glycoprotein-rich ground substance (Bi et al. 2007; Kannus 2000; Yoon & Halper 2005). This dense, fibrous, water-rich matrix that surrounds the tendon cells poses a significant challenge for the acquisition of high-quality RNA. In addition, tendons have low cell density compared with other tissues such as muscle or liver, resulting in minimal RNA yield per gram of tissue (Kannus 2000; Reno et al. 1997).

Previous studies have described protocols for RNA extraction from human or larger mammalian animal models such as rabbit (Ireland & Ott 2000; Reno et al. 1997), but analyzing RNA from small animal models such as mouse can be more difficult. This issue has led to several different strategies for achieving RNA yield and quality sufficient for gene expression analysis by RT-qPCR or RNA-seq. RNA amplification methods have permitted gene expression analysis of single injured and uninjured tendons (Dunkman et al. 2014), but this can be prohibitively expensive for analyzing a large number of samples or target genes. In addition, studies in other tissues have shown that such global pre-amplification can lead to biased results and increased false negative rates, especially for low- and medium-copy transcripts (Dunkman et al. 2014). Targeted pre-amplification methods have been developed to minimize PCR bias by using multiplexed primer pools at low concentrations combined with few PCR cycles (Jang et al. 2011). The resulting amplified cDNA can be used either in SYBR Green-based or probe-based qPCR assays. While such amplification methods have

been shown to be highly sensitive and yield less-variable RT-qPCR results compared to global cDNA amplification (Kroneis et al. 2017), the complex nature of the PCR amplification reactions imposes some limitations. First, targeted pre-amplification reactions require precision to yield usable cDNA. Because input sample concentrations, mRNA copy number, PCR cycle number, the specific combination of targets, and primer pool concentration can all affect the success of targeted pre-amplification, each assay must be individually optimized (Korenková et al. 2015; Kroneis et al. 2017). Improperly formulated reactions can lead to poor specificity and sensitivity of downstream qPCR, especially when using SYBR Green chemistry (Andersson et al. 2015). Such optimizations are time consuming and are not generalizable to different samples and different gene sets. Additionally, the nature of amplifying a specific set of targets inherently limits the possible downstream use of the pre-amplified samples.

Mendias and colleagues and Nielson and colleagues have performed gene expression analysis on a single mouse Achilles or plantaris tendon in different loss-of-function mouse models or in altered loading conditions (Mendias et al. 2008; Mendias et al. 2012; Nielsen et al. 2014). However, this approach is not widespread in the literature and these studies do not report on the RNA integrity, although they do report sample purity (260/280 ratio). However, there are examples of many studies that pool a large number of tendons (e.g., 12-20 individual tendons) (Bell et al. 2013; Trella et al. 2017). Not only does this increase the mouse cohort size and experimental costs, but it can also inflate the inter-individual variation, which may explain some of the large variability in transcript abundance that was found in subsets of their gene expression analysis (Trella et al. 2017). Lastly, other studies have focused on tendon-derived cell populations such as tendon stem/progenitor cells (Bi et al. 2007). This approach results in robust RNA yields, but it queries a cell population that has been expanded in culture and could have altered transcriptomic and epigenomic states compared with that of native tendon tissue.

The various technical limitations associated with obtaining high-quality, high-yield RNA using existing protocols enlarges the cohorts of mice needed for statistical analysis, and can hinder the use of RT-qPCR or functional genomic assays such as RNA-seq on single adult mouse tendons. Here, we present a robust, low-cost, and straightforward RNA isolation protocol that enables the isolation of high-integrity RNA from a single mouse Achilles tendon. We show that pooling tendon samples inflates estimates of biological variance for gene expression data in RT-qPCR analysis. We apply this method to analysis of injured and contralateral uninjured tendons to demonstrate the detection of significant and reproducible gene expression changes. In addition, this method can be used to purify high quality RNA from other musculoskeletal tissues, making it easily adaptable to multiple connective and skeletal tissue types, or from difficult to obtain tissues from humans or other organisms.

3.2 Methods

3.2.1 Mouse Studies

Achilles tendons were collected from wildtype C57BL/6J mice between 3-5 months of age (Jackson Laboratories 00664, n = 30 total). To compare gene expression levels between injured and uninjured Achilles tendons in the same mouse, excisional Achilles tendon injuries were performed using a 0.3 mm biopsy punch as described (Beason et al. 2012). The incision was closed with 6-0 Ethilon nylon sutures and the tendons were harvested 30 days after injury for analysis. Mice were housed, maintained, and euthanized according to American Veterinary Medical Association guidelines. All experiments were performed according to our Massachusetts General Hospital Institutional Animal Care and Use Committee (IACUC: 2013N000062) approved protocol.

3.2.2 RNA Extraction and Purification

Dissected Achilles tendons were placed immediately into 1.5 ml tubes containing 500 μl of TRIzol reagent (Invitrogen Cat# 15596026) and high impact zirconium 1.5 mm beads (30-40 beads per tube, D1032-15 Benchmark). Samples were homogenized in two 180-second rounds of bead beating at 50 Hz (BeadBug microtube homogenizer). Samples were then moved directly to dry ice or -80°C for longer storage up to 6 months.

Alternative tissue disruption procedures that were tested included homogenization of both fresh and frozen tendons in 500 μl TRIzol with a Polytron handheld homogenizer (PT 1200E, Kinematica AG) until tissue was visibly disrupted (60 to 90 seconds). Cryo-grinding of samples was tested using a freezer mill (SPEX 6875). Achilles tendons were snap frozen in liquid nitrogen and transferred to a super-cooled SPEX grinding cylinder (SPEX 6751C4) and pulverized in a bath of liquid nitrogen for 3 minutes. Ground samples were collected by rinsing the cylinder with 500 μl TRIzol and transferred to a 1.5 ml tube. For enzymatic digestion, tendons were placed in 2 ml Eppendorf tubes with 1 ml digestion solution containing 0.2% collagenase II (Worthington, LS004176) in DMEM (Gibco 11965) containing 0.1% Penicillin/Streptomycin (Corning 30002cl) and 1% Hepes (Gibco 15630-080). Tubes were kept on ice during the dissection period and were incubated together in a 37°C shaking water bath for 90 minutes. In order to digest remaining matrix, we added 200 µL of 0.2% collagenase I (Gibco 17100-017) and 300 µl of 0.4% Dispase (Gibco 17105-041) to the partially digested samples and incubated at 37°C for an additional 30 minutes. Following the digestion, the samples were centrifuged at 500 RCF (g) for 5 minutes, the supernatant was aspirated, and 500 μL TRIzol was added. All homogenized samples were stored at -80°C until RNA isolation.

To extract RNA, the samples were thawed on ice followed by a 5-minute incubation at

room temperature. Samples were quickly spun in the sample tubes and the homogenate was moved to a new Eppendorf tube, leaving behind the beads and residual tissue. Next, a chloroform extraction was performed, using double the recommended ratio of chloroform to TRIzol, which has been shown to increase RNA yield in small samples (Macedo & Ferreira 2014). One hundred microliters of chloroform was added to the homogenate and vortexed well for approximately 1 minute. The TRIzol /chloroform mixture was then moved to a 1.5 ml MaXtract high density tube (Qiagen Cat No. 129046), incubated at room temperature for 2-3 minutes, and spun at $\geq$ 12,000 x g at 4°C for 15 minutes. MaXtract tubes contain a sterile gel that forms a barrier between the RNA-containing aqueous phase and the TRIzol/chloroform upon centrifugation at 4°C, thus minimizing carryover of organic solvents leading to an overall reduction in sample contamination. After centrifugation, the aqueous phase was transferred to a clean 1.5 ml Eppendorf tube and an equal volume of 100% ethanol was added to the aqueous phase and mixed well. At this stage, the RNA/ethanol mix was typically stored at -80°C. We have found that brief incubation of this mixture at -80°C improved the total RNA yield, but it is not required.

RNA purification was next performed using the ZR Tissue & Insect RNA MicroPrep kit (Zymo Research R2030) or the Direct-Zol systems (Zymo Research R2050, R2060). Based on typical tendon yields, the ZymoSpin IC spin columns are optimal for use with RNA extracted from single tendons as these columns can purify up to 5 μg of RNA in as little as 6 μl eluate. However, this protocol also has been successfully used with ZymoSpin IIC columns, which require a larger elution volume. After adding the RNA/ethanol mix to the spin column, the standard Zymo purification protocol was used with the following modifications. First, a 15-minute on-column DNase I treatment was added to minimize genomic DNA contamination. An extra wash step was included to improve sample purity. Prior to elution, columns were spun for an additional 2 minutes at maximum speed to remove residual ethanol. RNA was eluted in 15 μl

RNase/DNase free water that was pre-warmed to 55-60°C to maximize the RNA recovery from the spin column. RNA concentration was measured via fluorometric quantitation (Qubit HS RNA assay, Invitrogen, CAT# Q32852) and sample quality was determined by spectrophotometric analysis (NanoDrop 2000c, ThermoFisher Scientific) as well as capillary electrophoresis (2100 Bioanalyzer, Agilent). The final RNA product was stored at -80°C for RT-qPCR analysis.

3.2.3 RT-qPCR, Data Analysis, and Statistics

One hundred nanograms total RNA was reverse transcribed with $oligo(dT)_{20}$ primers using the SuperScript IV First Strand Synthesis System (Thermo Fisher 18091050) and a no-reverse transcriptase control was included for every sample. A total of 2 ng cDNA template was amplified for 40 cycles in each SYBR green qPCR assay (Applied Biosystems 4367659) using a final primer concentration of 200 nM. All assays were performed in technical triplicate using either a LightCyclerII 480 (Roche; pooled samples) or a StepOnePlus Real Time PCR system (Applied Biosystems; injury samples). Three independent biological samples were run per condition for both sets of RT-qPCR. *Gapdh* was used as the reference gene for all samples.

All analyses were conducted in R 3.4.3 (@R Core Team (2019)). For the pooling experiment, summary statistics were calculated for *Scleraxis* (*Scx*) and *Gapdh* technical and biological replicate cycle threshold (CT) values independently. Variance estimates for *Scx* ΔC_T relative expression were calculated using standard error propagation techniques. Relative expression values for *Collagen Ia2* (*Col1a2*) and *SRY-Box9* (*Sox9*) were calculated for the injury analysis using the $\Delta\Delta C_T$ method (Livak & Schmittgen 2001) and injury samples were normalized to their corresponding uninjured contralateral controls. Statistical differences between injured and uninjured samples from three biological replicates (n = 3 mice) were analyzed via Welch's t-test (Welch

1947) on the ΔC_T values.

3.3 Results

Several tissue disruption methods were tested in order to achieve optimal RNA quality and quantity from a single mouse tendon. Among those tested were enzymatic digestion, cryogenic grinding (manual and mill), shearing with a handheld homogenizer (i.e., rotor-stator), and bead beating. Capillary electrophoresis was performed on purified RNA using a Bioanalyzer RNA Nano chip (Agilent). RNA integrity number (RIN), a quantification of degradation, was calculated by the accompanying Agilent software based on the electropherogram for a given sample; a RIN of 10 indicates completely intact RNA whereas a RIN of 1 indicates severely degraded RNA. Enzymatic digestion produced intact RNA (RIN > 7), but low RNA yield ($\leq$ 1ng/μl). Cryogenic grinding and handheld homogenizer dissociation methods resulted in low yield ($\leq$ 5ng/μl) and poor RNA purity and integrity (RIN ≤ 3). Bead beater homogenization was found to produce the best results in terms of RNA quality (i.e., RIN ≥ 6.5) and quantity ($\geq$ 50 ng/µl), and minimized carryover between samples. Additionally, bead beating was easily combined with standard TRIzol extraction and commercially available purification methods.

To further evaluate our bead beating homogenization method, we performed additional experiments examining the level of degradation that occurs prior to homogenization as well as during homogenization. To address the former, single Achilles tendons from similarly aged mice were left in sterile 1x PBS on ice following dissection for up to 9 minutes before homogenization to simulate waiting times involved in batch dissection. The shortest time between dissection and homogenization (0-30 seconds) yielded more intact RNA (RIN = 6.5) while longer wait times resulted in more degraded RNA (9 minutes processing time RIN = 5.4; Figure 3.1). This demonstrates that

measurable degradation can occur prior to sample homogenization, and occurs with increases in time after dissection on the order of only minutes (Figure 3.1). Therefore, processing the tendon(s) immediately following dissection is essential for preserving RNA integrity. We next tested how the duration of bead beating affects RNA quality by varying homogenization times of single and four pooled Achilles tendons. Samples were homogenized for 30 seconds, 60 seconds, 180 seconds, or 360 seconds (in two consecutive rounds of 180 seconds; Figure 3.2A,B). RNA from samples homogenized for less than 60 seconds suffered more degradation than those that underwent longer homogenization times (Figure 3.2B), indicating incomplete homogenization of the tissue during the shorter bead-beating periods. Homogenization times of 720 seconds did not improve RIN numbers (RIN = 6). This could be due to prolonging the amount of time until RNA extraction or elevation of the temperature with longer homogenization periods, leading to degradation. The temperature of TRIzol is a likely factor as other studies with similar homogenization techniques used methods to lower its temperature (Leite et al. 2012).

To test whether pooling tendons from multiple individuals into one sample prior to homogenization influences RNA integrity, we measured RNA quality from single Achilles tendons as well as pools of differing sizes (2, 4, 6, and 8 tendons, n = 3 biological replicates per pooling level; Figure 3.3A,B). Electropherograms and RIN measurements show that RNA from all pooling levels suffer levels of degradation similar to single Achilles samples (Figure 3.3A,B). Purity measurements were also similar among single and pooled samples. Therefore, pooling tendons from multiple individuals is not protective against RNA degradation; the only measure that improved with increased pool size was RNA yield (Figure 3.3C). To determine if pooling multiple samples affects gene expression measurements, we evaluated gene expression in single and differentially pooled tendon samples described above (n = 3 per pooling level) via RT-qPCR. Although we find no gain in RNA quality from pooling, treating pools

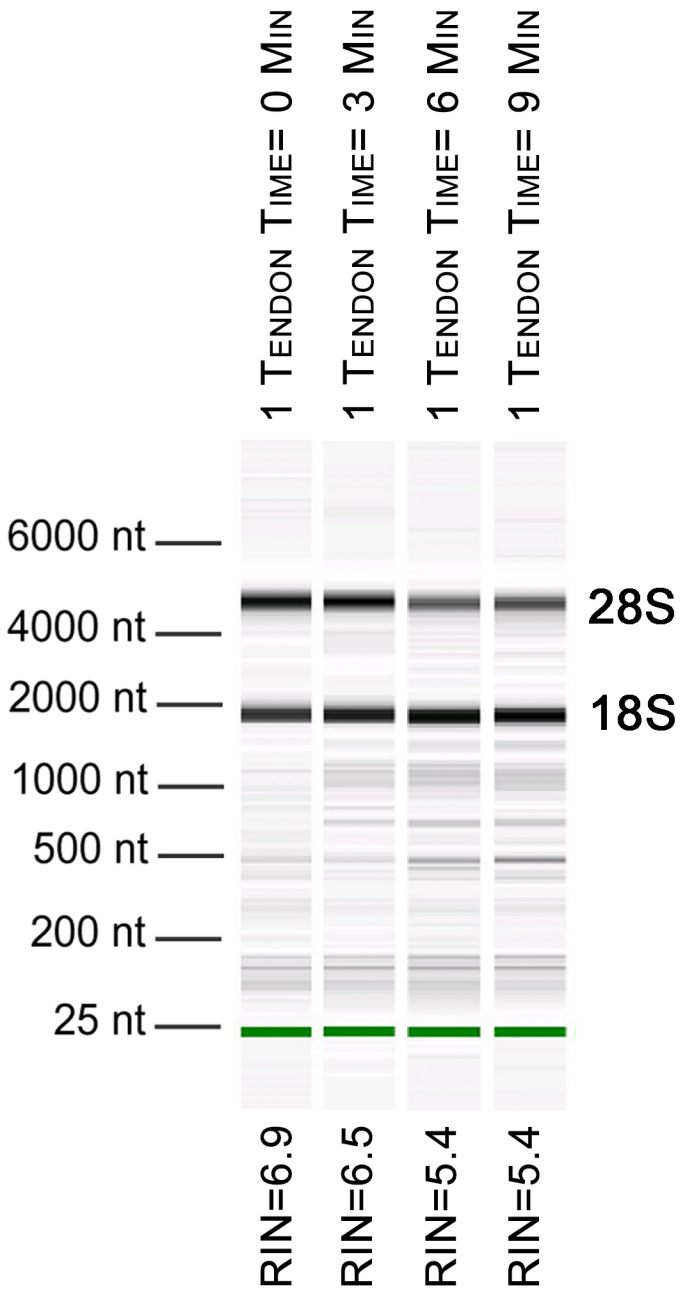


Figure 3.1: Length of time between dissection and processing affect RNA integrity.

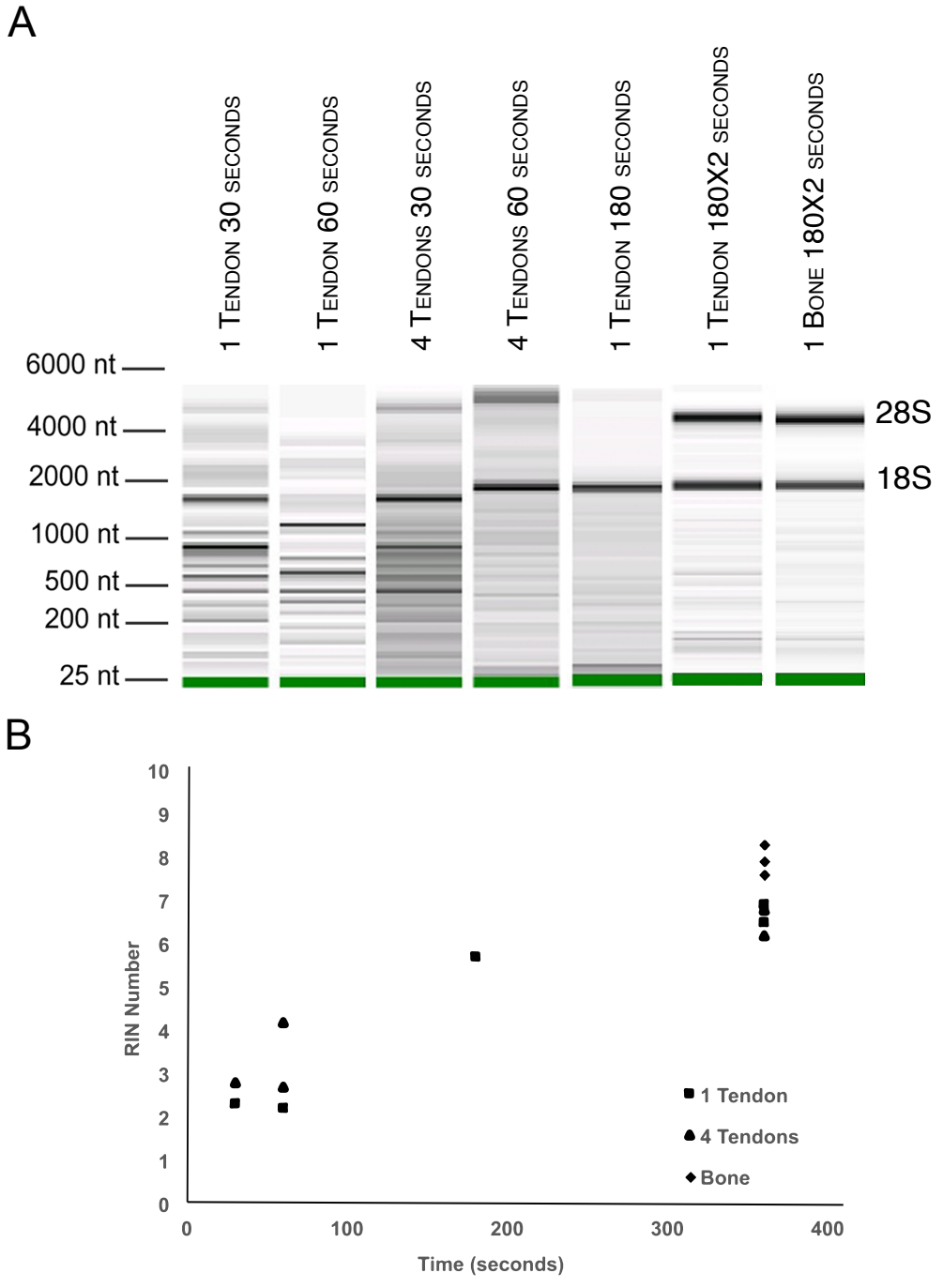


Figure 3.2: Optimization of homogenization regime.

of tendons from multiple individuals as single biological replicates results in larger standard deviations in C_T measurements in assays for *Scx* and *Gapdh* (Figure 3.4). This leads to larger sample variance for larger pools, driven by differences in ΔC_T between biological replicates within a group, which impedes the detection of small gene expression changes. Such increases in variance for pooled versus single samples have also been reported for RNA-seq datasets (Rajkumar et al. 2015).

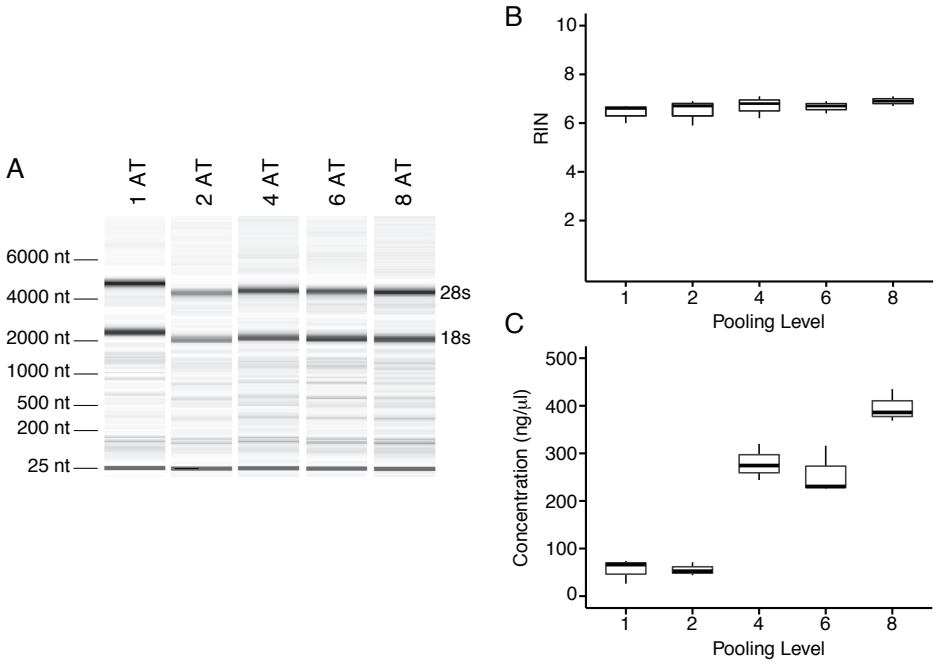


Figure 3.3: Tendon pooling affects RNA yield but not quality.

To validate the performance of the RNA obtained using this protocol, we performed RT-qPCR for *Sox9* and *Col1a2* expression on single Achilles tendons at 30 days following an acute excision Achilles tendon injury. All samples were obtained from single injured and contralateral uninjured Achilles tendons from the same mouse. Using this protocol, we found significantly increased expression of *Sox9* and *Col1a2* in injured Achilles tendons compared with their uninjured contralateral counterparts ($p < 0.05$ for *Sox9* and $p < 0.01$ for *Col1a2*; Figure 3.5). These results are consistent

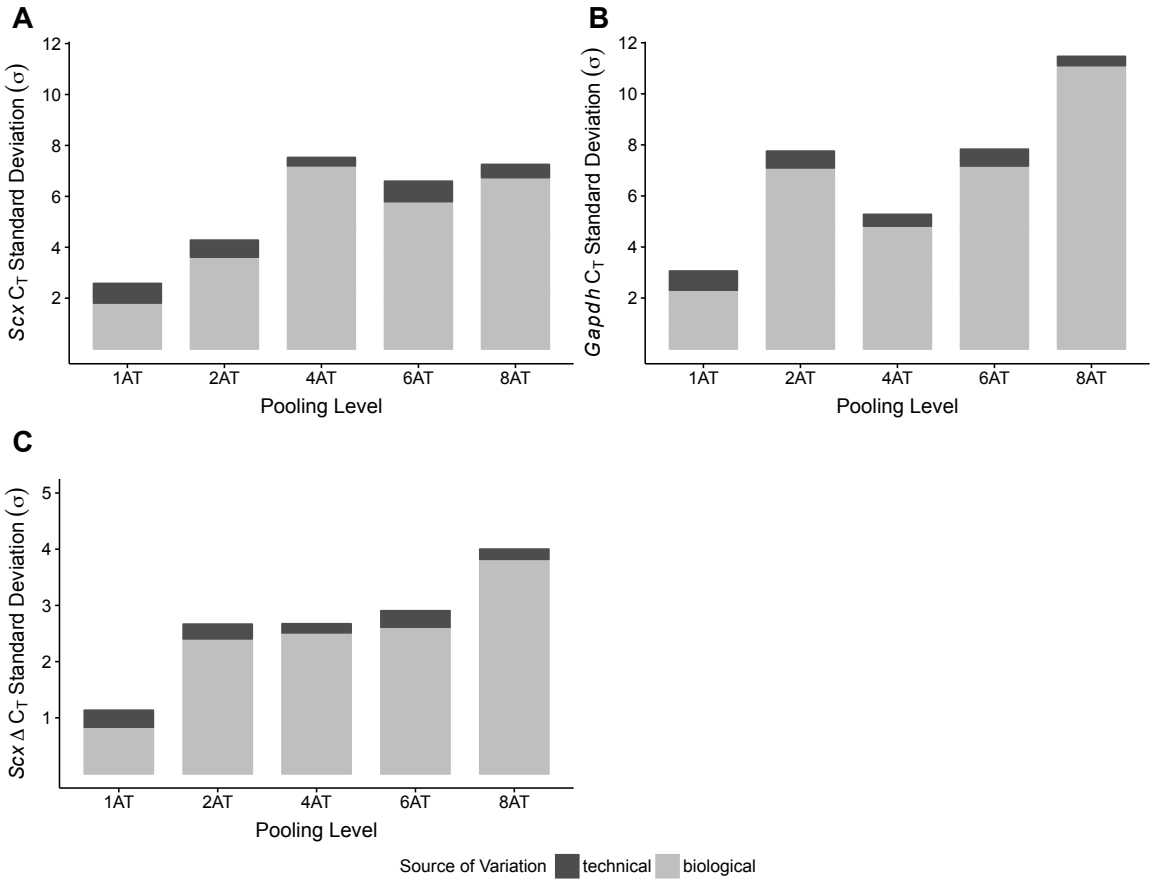


Figure 3.4: Sample pooling affects estimates of sample variance in RT-qPCR.

with previous studies showing increased expression of *Sox9* and *Col1a2* following tendon injury (Guerquin et al. 2013; Zhang & Wang 2013), and also show that our method is robust enough to detect gene expression changes in single tendon samples.

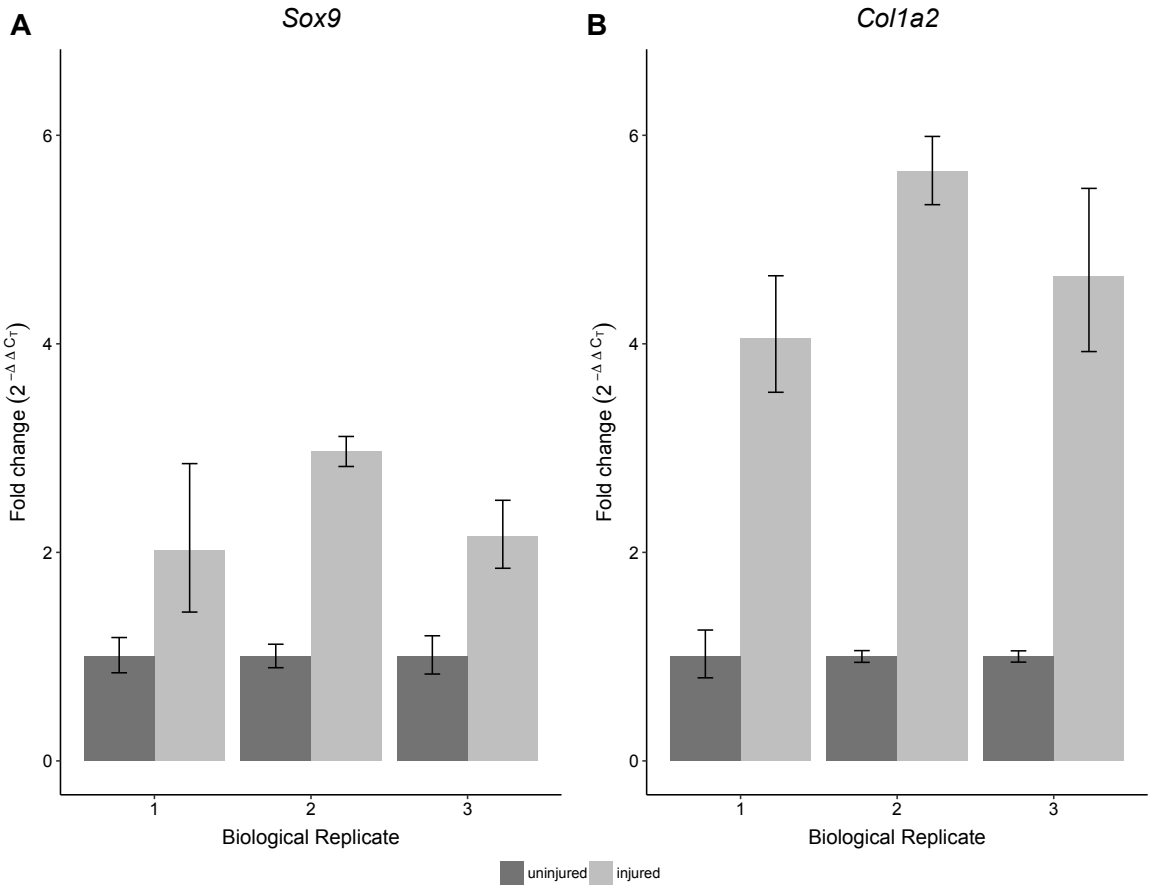


Figure 3.5: Sensitivity and reproducibility of RT-qPCR on single tendon RNA.

3.4 Discussion and Conclusions

Obtaining high quality RNA from tendons can be challenging, and this can limit the direction and scope of studies focused on analyzing adult mouse tendon tissues. Whereas a few studies have used single tendons without amplification, many other studies have used amplification or pooling of greater than 12 samples to detect gene expression changes. Both approaches can be expensive due to the high costs associated with amplification kits for multi-gene analysis or the number of mice used for one

biological replicate. Dissociation, followed by culture and expansion of tendon-derived cells can yield greater RNA concentrations of high quality, but such approaches cannot be used to study gene expression changes after injury. The approach we described above provides a straightforward method to consistently obtain high yields of RNA from one Achilles tendon of sufficient quality to perform RT-qPCR analysis without amplification. Our reported RIN scores are also acceptable for standard RNA-seq differential expression analysis. In addition, studies have shown minimal variance among polyA-selected samples of high (RIN $>$ 7) to moderate (RIN = 6-7) RNA integrity as well as efficacy in correcting for variation in RNA integrity in the differential expression analysis (Jaffe et al. 2017; Romero et al. 2014). Although high RIN values should be the goal, there are some options for studies in which there are limitations in sample quality with moderate RIN scores. However, it must be noted that each RNA-seq library preparation system has specific input RNA requirements, and researchers should ensure that their RNA samples meet all manufacturer qualifications prior to use in a sequencing study.

Our analysis also uncovered key steps that are integral towards generating intact, high yield RNA from the single tendon samples. In particular, we find that the time from dissection to homogenization and storage can significantly impact the quality of the RNA, causing measurable degradation. In this regard, even small delays on the order of minutes could affect overall RNA quality, which could greatly affect differential gene expression analysis. In addition, the duration of homogenization is important for maximizing RNA yield and quality. Homogenization times that are too short or long can result in dramatically different RIN and concentrations regardless of the level of sample pooling.

Similar to previous RNA-seq studies, our RT-qPCR analysis of single and pooled tendon samples revealed that pooling increases the variance of gene expression measurements

(Rajkumar et al. 2015). It has been argued that pooling samples from multiple individuals into single biological replicates results in biological averaging and is therefore an appropriate, and even useful, practice in gene expression studies via microarray (Kendziorski et al. 2005). However, genes that are lowly expressed or exhibit subtle differences between conditions would require a larger sample size of pools to achieve adequate statistical power, which would further inflate mouse and reagent cost for RT-qPCR, microarray, or RNA-seq analyses (Shih et al. 2004). This study also highlights the problem of performing RT-qPCR comparisons on a single pool per group (run in technical triplicate), under the assumption that the within-sample variation is representative of the biological variation among all animals of that group. Variance calculated from technical replicates does not estimate biological variance within each group, and is not an appropriate practice. Technical variation arises from noise due to measurement error and therefore is unrelated to biological variation (Kitchen et al. 2010; Vaux et al. 2012), necessitating the use of multiple pools for any statistical analysis.

Our tendon RNA extraction method is a robust protocol for obtaining high quality RNA for gene expression assays. It decreases the number of mice required for analysis and avoids extra amplification steps, making it straightforward, cost-effective, and easily accessible to researchers new to the tendon field. By providing a means for reproducibly analyzing one Achilles tendon, this method also reduces measurement error associated with pooling tendons from multiple individuals. Moreover, our protocol permits the use of internal comparisons between a limb that has undergone experimental manipulation (e.g., injury or unloading) and the contralateral control limb within the same animal. In addition to facilitating larger-scale RT-qPCR studies, we believe this method will make high dimensional gene expression analysis such as RNA-seq accessible to more researchers studying musculoskeletal tissues, thus opening new frontiers in tendon biology.

CHAPTER 4

A DISTINCT TRANSITION FROM CELL GROWTH TO PHYSIOLOGICAL HOMEOSTASIS IN THE TENDON

This chapter is currently in review at eLife: Grinstein M, Dingwall HL, Zou K, Capellini TD, Galloway JL. A distinct transition from cell growth to physiological homeostasis in the tendon.

4.1 Introduction

Development, growth, and homeostasis rely on the precise regulation of cell proliferation and differentiation to generate and maintain a functioning organism. Frequent cell divisions grow tissues to the proper size, as do modifications to non-cellular tissue properties such as to the extent of extracellular matrix. However, once size is achieved, each tissue maintains its physiological functionality either through stem cell-mediated mechanisms as in the intestine (Simons and Clevers, 2011), the duplication of specialized cell types as in the liver (Miyajima et al., 2014), or in the virtual absence of cell division as in the heart (Senyo et al., 2014). In some cases, the transition from active proliferating to terminally differentiated cells has been attributed to a change in regenerative potential as observed in neonate verses adult mouse hearts (Senyo et al., 2014). Therefore, understanding transitions in cell cycle activity are important for setting the framework for more deeply understanding proliferative-driven growth stages and distinguishing between specific homeostatic renewal mechanisms in the adult. This knowledge is significant in considering therapies for tendon injuries, which can be challenging to treat due to their imperfect healing and propensity for re-injury

(Thomopoulos et al., 2015).

Tendons begin as aggregations of cells that secrete and organize a highly ordered matrix to connect the musculoskeletal system and enable movement. Therefore, tendon growth and maintenance must not only involve its matrix but also the cells that generate and eventually reside within it, making knowledge of the cell cycle transitions important for understanding these processes. In the adult, the tendon matrix contains organized type I collagen fibrils and tenocytes, which are mature tendon cells possessing cellular extensions that project into the matrix (Kalson et al., 2015; Kannus). This mature stellate morphology differs greatly from that of the rounded shape of embryonic and neonatal tenoblasts. During embryogenesis, limb bud mesenchymal cells express the transcription factor, *Scleraxis* (*Scx*), and coalesce into tendon primordia, which organize to connect muscle and bone (Schweitzer et al., 2001). Through the transgenic labeling of cell cycle state using the Fluorescent ubiquitination-based cell cycle indicator (Fucci), robust numbers of mitotic tendon cells have been detected prior to birth (Esteves de Lima et al., 2014). Specific segments of the limb display more cell cycle activity than others (Huang et al., 2015), suggesting that there are localized effects on tendon cell proliferation during embryogenesis. In addition to cell growth, these embryonic stages are marked by an increase in the number of collagen fibrils deposited in the matrix (Kalson et al., 2015). These collagen fibrils grow in length and diameter to grow the tissue (Ezura et al., 2000). Scanning electron microscopy at postnatal stages (P0 and P42) has shown an increase in the diameter of the collagen fibrils rather than an increase in collagen fibril or cell number in the tail tendons of mice (Kalson et al., 2015). These observations have led to a model whereby tendon postnatal growth is primarily driven by expansion of the extracellular matrix (ECM), which results in a reduction in cell density across the whole tissue in growth and aging (Dunkman et al., 2013; Kalson et al., 2015). However, a direct analysis of cell proliferation and the transition in cell cycle activity from birth

to adult and aged stages has not been performed.

Although adult tendons display limited proliferation, some cell cycle activity has been detected *in vivo*, especially in the context of injury. An early study by Lindsay and Birch (Lindsay and Birch, 1964) showed that adult flexor tendons undergo a healing response, which includes marked cell division from cells originating from peritendinous regions rather than from internal tendon cells. Tendon-derived stem/progenitor cells were characterized based on their ex vivo abilities to proliferate, clonally expand, and undergo serial transplantation (Bi et al., 2007). However, the identity and *in vivo* activity of the resident cell population remains unknown. Other studies have reported proliferation in adult tendons during homeostasis and repair (Runesson et al., 2013; Tan et al., 2013). Because these studies were performed without genetic lineage tracing tools, the origin of the cells proliferating in response to injury was unclear. Recent lineage tracing experiments using injured adult mouse Achilles and patellar tendons suggest non-*Scx-GFP* expressing Smooth muscle actin (SMA)-lineage cells are the major source of proliferating cells that mediate adult healing (Dyment et al., 2014; Dyment et al., 2013; Howell et al., 2017). The paratenon is believed to be the source of the cells responding to injury as non-*Scx-GFP* and SMA-lineage cells in the paratenon localize to the injury site and turn on *Scx-GFP* in partial defect injuries (Dyment et al., 2014; Dyment et al., 2013; Sakabe et al., 2018). These results are consistent with earlier studies showing external tendon cell populations exhibited increased proliferative capacity compared with internal tendon cells in culture (Banes et al., 1988). Interestingly, in aging, tendon cell number per unit area decreases, suggesting declining proliferative abilities with age (Dunkman et al., 2013). Consistent with this interpretation, tendon-derived cells from aged mice have reduced proliferative abilities and increased markers of senescence compared with adult-derived tendon cells in culture (Kohler et al., 2013). Together, these studies indirectly indicate that adult and aged tendon cells have reduced cell cycle activity, yet sub-populations of tendon

cells may proliferate in adult tendons under specific injury conditions. Nevertheless, it remains unclear to what extent, if any, there is physiological cell turnover in the adult tendon without injury and how this may differ from cell turnover during periods of active tendon growth.

Therefore, we sought to examine cell turnover rates in limb tendons during growth, adulthood, and aging. Using complementary methods of genetic pulse-chase labeling to trace the cell division history and BrdU/EdU incorporation to detect proliferation, we were able to identify changes in tendon cell cycle activity from birth to the early juvenile period (beginning around 3-4 weeks), with comparisons to adult and aged stages ($\geq$ 3 months and $\geq$ 18 months, respectively). We detect relatively high levels of proliferation during the neonatal period (P0-P7) and a rapid decline by P21. Although proliferation was significantly reduced after one month of age, surprisingly we were able to identify a small population of tendon cells that continued to proliferate in adult and aged mice, albeit at a very low rate. Understanding which cell populations can continue to divide in adults and the mechanisms driving the switch from proliferative to more quiescent stages would greatly benefit clinical approaches to tendon injuries.
##Methods

4.1.1 Animals

We thank Andrew Brack (UCSF) and Konrad Hochedlinger (MGH) for the Doxycycline (Dox) inducible H2B-GFP (*Col1a1:tetO-H2B-GFP; ROSA:rtTA*) heterozygous mice used in these studies. To induce transgene expression, Dox (Sigma D9891, 2 mg/ml, supplemented with sucrose at 10 mg/ml) was added to the drinking water of timed pregnant females from E10 to birth as described (Foudi et al., 2009). Scx-GFP and Scx-Cre mice were provided by the Schweitzer lab (Blitz et al., 2009; Schweitzer et al., 2001). Gt(ROSA)26Sortm9(CAG-tdTomato)Hze (Ai9) were obtained from Jackson

Laboratory (Jax cat# 007909). All experiments were performed according to our protocol approved by the Massachusetts General Hospital Institutional Animal Care and Use Committee (IACUC: 2013N000062).

4.1.2 Flow Cytometry

The tendon cells were isolated from the distal forelimb and hindlimb tendon tissue (Achilles, extensor, deep and superficial flexor tendons) from mice at time points between P0 and 2 years. Limb tendons were enzymatically dissociated in a solution containing 0.2% collagenase II (Worthington Cat# LS004176) in DMEM (Gibco Cat#11956-092) with 1% pen/strep (Corning Cat#30002CL) and 1% Hepes (Gibco Cat#15630-80) for 2 hours at 37°C. Subsequently, a secondary digestion solution containing 0.2% Collagenase I (Gibco Cat# 17100-017) and 0.4% Dispase (Gibco Cat# 1710541) was added and the samples were incubated for an additional for 30 minutes at 37°C. The digested cells were filtered with 30 μm filters (MACS Cat# 130041407) and washed. For the H2B-GFP$^+$ studies, we enriched for tendon cells from H2B-GFP$^+$ mice by excluding for CD31$^+$ and CD45$^+$ cells using FACS prior to analysis (BD Cat#551262, Cat# 557659). For the BrdU analysis, cells were stained with anti-BrdU following tendon tissue dissociation (Biolegend, Cat# 339808), and tendons from *Scx-Cre-TdTom*$^+$ or *Scx-Cre-TdTom*$^+$;*Scx-GFP*$^+$ mice were used to analyze TdTom+ or GFP+ tendon cells. Flow cytometry was performed using 5 ml tubes (BD Biosciences Cat# 352235) on a FACSAria II (BD Biosciences). For each independent experiment, gates were defined by positive and negative control tendon cells from TdTom$^+$/TdTom$^-$ and GFP$^+$/GFP$^-$ cells. For the negative controls for BrdU analysis, BrdU antibody staining was performed on tendon cells isolated from mice that were not administered BrdU. To ensure reproducibility of H2B-GFP emission intensity between different samples and sorting times, the voltage of the

photomultiplier receiving signal from the 488-nm laser was normalized using Green Flow Cytometry Reference Beads prior to every sort (Molecular Probes Cat# C16508).

4.1.3 BrdU, EdU labeling, Tendon Histology, and Imaging

BrdU was injected at a concentration of 150 mg/kg (Sigma Cat#B5002) as described (Magavi and Macklis, 2008). Flow cytometry analysis was performed as described previously. For BrdU immunostaining, sections underwent antigen retrieval and immunostaining using anti-BrdU (1:100; Abcam Cat# 6326). EdU was administered at 20mg/kg as described (Salic and Mitchison, 2008) and tendon sections were stained using the Click-iT EdU kit (Invitrogen Cat# C10337). For histological sections, tendons were fixed overnight in 4% PFA, followed by 5% sucrose for 1 hour, and 30% sucrose overnight before being mounted in OCT. A Leica cryostat (CM3050S) was used to obtain 8-10 μm sections. Pictures were taken with Zeiss AxioImager D2 with (10X and 20X magnification) and prepared using Adobe Photoshop and Illustrator.

4.1.4 Tendon length and cell counting measurements

Tendons from at least 3 mice were measured per stage from the calcaneus to the gastrocnemius muscle, and the data were analyzed using Prism software (Graphpad). For cell counting, at least 10-12 sections of the Achilles tendon were analyzed from each mouse with at least 3 mice per stage examined.

4.1.5 Mathematical modeling

To define the dilution of the GFP signal mathematically, we modeled the change in signal intensity using a logarithmic decay equation (Equation (4.1)). In this formula, we assume that GFP signal intensity decreases through dilution by cell proliferation.

We calculated the constant between populations at different times (k), by comparing the populations' median GFP intensity at particular times ($P(t)$). Assuming that the increase in tendon cell number could be measured by the decrease of the GFP intensity (Figure 4.2B,D), we calculated the dilution of GFP between each time point from P0 to 645 days (Figure 4.2C).

$$P(t) = P(0)e^{-kt} \tag{4.1}$$

4.1.6 RNA extraction and RT-qPCR

Fresh, whole limb tendons (pooled forelimb and hindlimb from a single individual; n = 3 mice per time point) were dissected from mice euthanized via CO2 and immediately placed in cold TRIzol (Invitrogen 15596026). Tendons were roughly chopped with clean microdissection scissors in TRIzol and frozen at -80°C until RNA extraction via TRIzol-chloroform and a proprietary kit. Briefly, the homogenate in TRIzol was thawed on ice, vortexed, and transferred to a clean microcentrifuge tube to remove tissue debris. The traditional TRIzol-chloroform extraction protocol was followed until phase separation. An equal volume of ethanol was added to the upper aqueous phase and the mixture was transferred to a Zymo IIC spin column (Zymo Research C1011) for purification and DNase I treatment using the Zymo Direct-Zol system (Zymo Research R2050, R2060) following the manufacturer's guidelines. RNA quality was examined using spectrophotometry (NanoDrop 2000c, Thermo Scientific) and capillary electrophoresis (2100 Bioanalyzer, Agilent), and concentration was measured via fluorometric quantitation (Qubit HS RNA assay, Invitrogen Q32852). The final RNA product was stored at -80°C.

Total RNA was reverse transcribed using the SuperScript IV first strand synthesis system (Thermo Fisher 18091050). 100 ng total RNA for each sample (n = 3 per

time point) was converted to cDNA using oligo(dT)20 primers. SYBR green assays (Applied Biosystems 4367659) were run in technical triplicate with 1 ng of cDNA template in each 12.5 μl reaction. Samples were amplified for 40 cycles using the LightCycler 480 II real time PCR system (Roche Diagnostics). All targets were normalized to *Gapdh*. Relative expression values were calculated for visualization using the $\Delta\Delta C_T$ method (Livak and Schmittgen) and are shown as $\Delta\Delta C_T \pm$ standard error (Figure 4.4); statistics were performed on ΔC_T values.

4.1.7 Statistics

For the RT-qPCR assays, statistical differences among the six timepoints were investigated via ANOVA and post hoc pairwise comparisons were computed using Tukey's Honestly Significant Difference test on the ΔC_T values (n = 3 biological replicates per time point; alpha = 0.05). R statistical software (R Core Team (2019)) was used for all RT-qPCR calculations and visualizations. Data analysis in R was facilitated using R packages included in the Tidyverse collection (Wickham (2017)) and statistical analysis was performed using the implementations of ANOVA and Tukey's HSD in 'stats' version 3.5.1 (R Core Team (2019)). For each stage analyzed by flow cytometry, least 3 mice were used per group. Statistical differences between time points were calculated using a Welch's t-test.

4.2 Results

4.2.1 H2B-GFP pulse chase experiments demonstrate a shift from high to low proliferation rates in postnatal mice

To characterize cell proliferation in the tendon, we used the doxycycline (Dox) inducible Histone 2B-green fluorescent protein reporter mouse model (*Col1a1-tetO-H2B-GFP;*

ROSA-rtTA, henceforth referred to as H2B-GFP), which has been used to quantify cell proliferation and identify slowly cycling label-retaining cell populations (Chakkalakal et al., 2014; Foudi et al., 2009). After H2B-GFP expression is induced by Dox addition, Dox is removed for the chase period and H2B-GFP signal becomes diluted in proportion with each subsequent cell division (Figure 1A). Therefore, cells cycling more frequently will dilute H2B-GFP expression more quickly and will appear unlabeled earlier in the chase period; more slowly cycling cells will retain H2B-GFP expression longer during the chase. To verify the H2B-GFP system worked efficiently in the tendon, we pulsed mice with Dox from embryonic stage (E) 10 to birth and examined H2B-GFP expression on postnatal day (P) 0. Histological sections of pulsed P0 Achilles (Figure 4.1B-D) and extensor (Figure 4.1E-G) tendons showed widespread expression of H2B-GFP throughout the tendon nuclei. We next confirmed by flow cytometry that more than 90% of the tendon cells were positive for H2B-GFP at P0 (Figure 4.1H), indicating efficient labeling of all tendon cell populations examined. To ensure we were enriching for tendon cells, we only analyzed cells from dissected tendon tissues that were negative for $CD45^-$ and $CD31^-$ to remove blood and endothelial cells, respectively (Sup. Fig. 1). At all stages analyzed, tendon cells were isolated from extensor, deep and superficial flexor, and Achilles tendons in the hindlimbs and extensor, deep and superficial flexor tendons in the forelimbs.

To determine the total cumulative proliferation of tendon cells from birth to aged mice, we next examined H2B-GFP expression in tendons that had been pulsed with Dox at embryonic stages and allowed to chase without Dox for over 18 months (645 days) using section and FACS analysis. Tendons in section appeared to have reduced H2B-GFP^+ nuclei compared to P0 tendons (Figure 4.1I-N). We found that H2B-GFP^+ cells had shifted in the intensity of GFP (Figure 4.1O) with only $20.1 \pm 1.4\%$ of the cells H2B-GFP^+ at 645 days (Figure 4.2A). Previous studies calculate that 7-8 divisions are needed for a cell to fall below the GFP detection threshold (Foudi et al.,

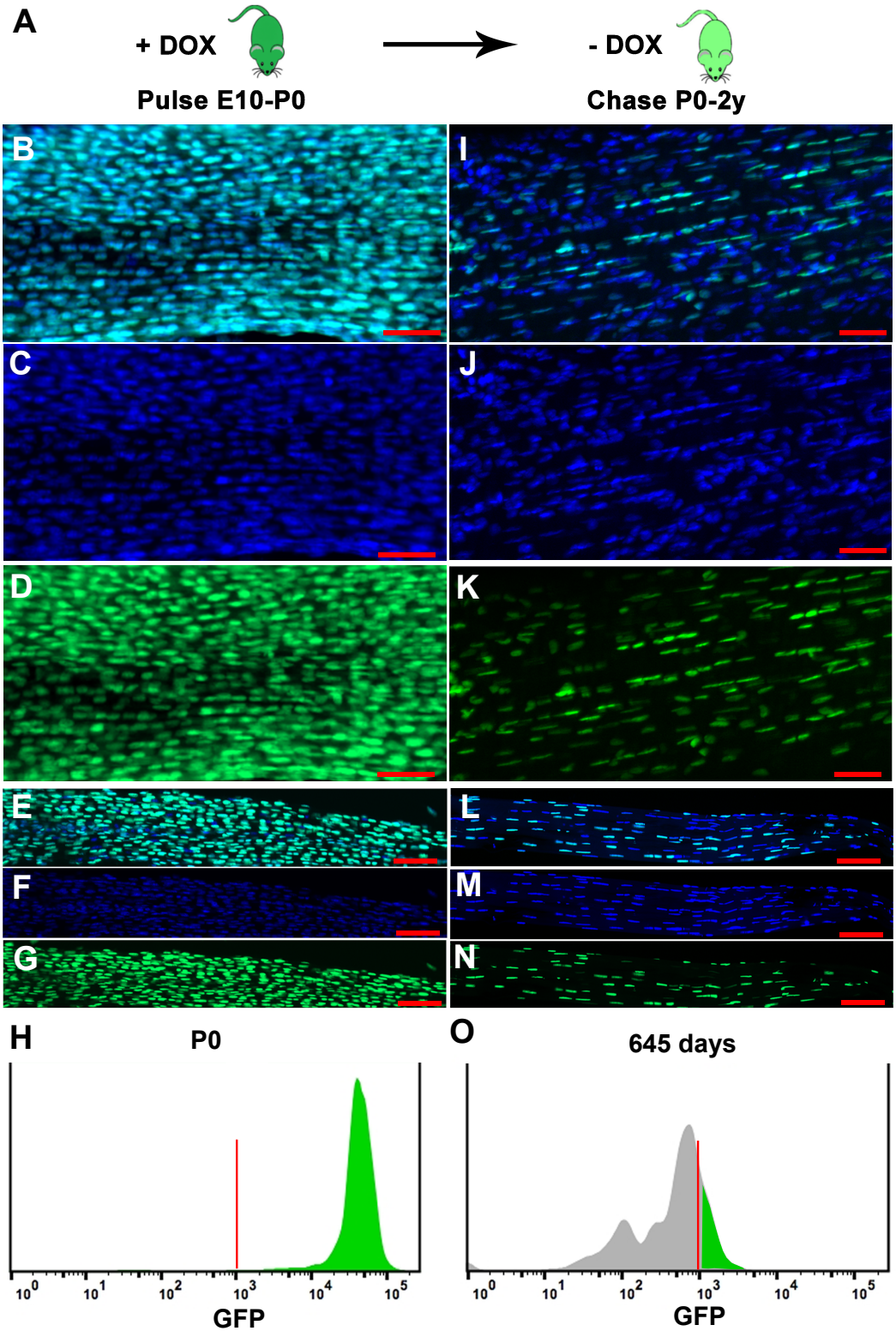


Figure 4.1: H2B-GFP expression is induced upon the addition of Dox to timed pregnant females from E10 to P0 (dark green); Dox is removed for the chase period of 0-2 years and H2B-GFP is diluted (light green) in proportion to cell division.

2009). This would indicate that the H2B-GFP$^+$ population at 645 days proliferated less than 7-8 times, while the H2B-GFP$^{\breve{}}$ population proliferated at a minimum of 7-8 times since birth. Together, these data show that all tendon cells proliferate after birth, but that a subpopulation of the cells display limited cell cycle activity.

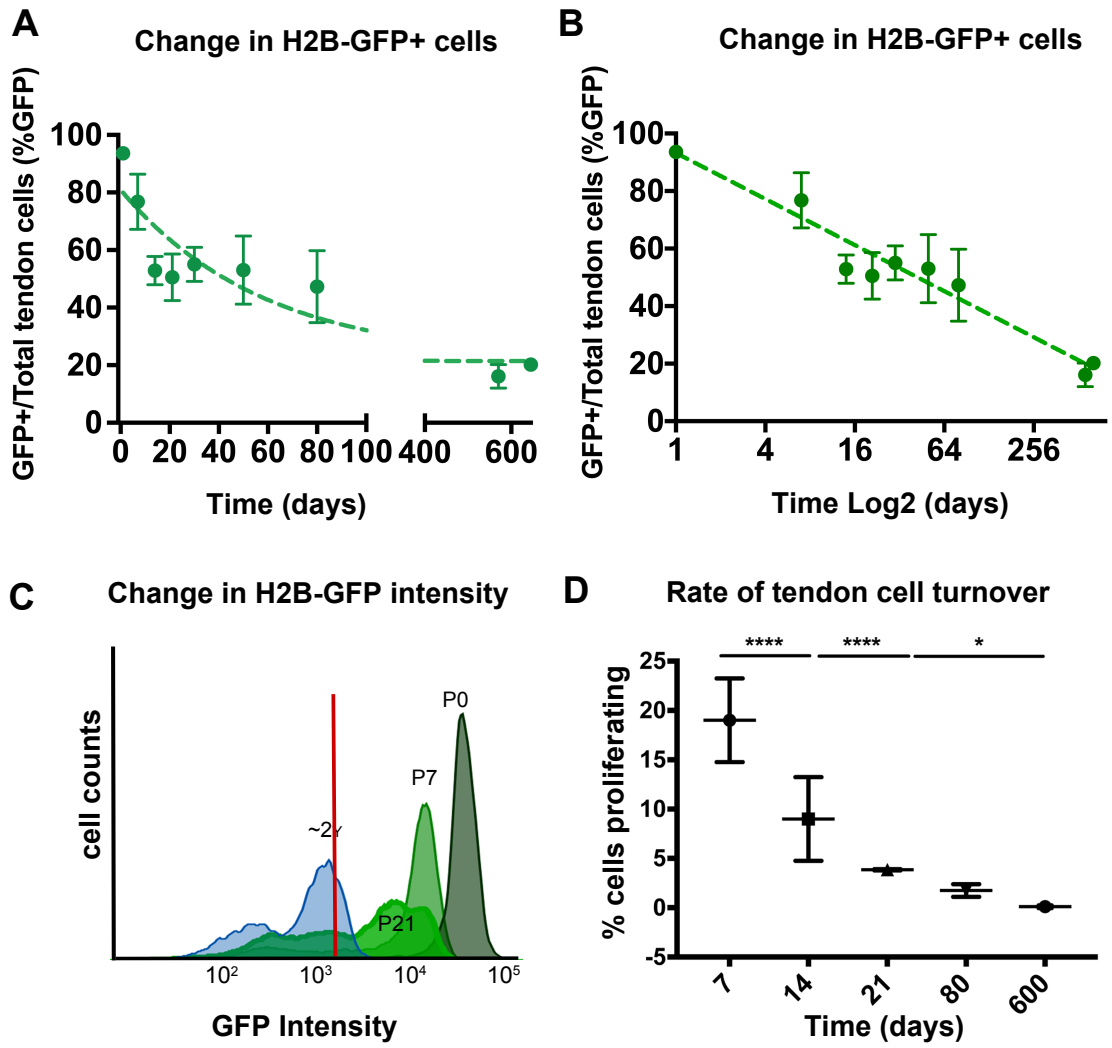


Figure 4.2: Analysis of H2B-GFP$^+$ cells from postnatal to aged stages.

To more deeply assess the dynamic changes of tendon cell proliferation after birth, we analyzed Dox pulsed mice at multiple stages of chase from P0 to P80. We observed a decrease in the total percentage of H2B-GFP$^+$ cells from 93.6 $\pm$ 1.2% at P0 to 76.7 $\pm$ 9.6% at P7 and 52.0% at P14 (Figure 4.2A). However, the percentage of H2B-GFP$^+$

tendon cells remained relatively constant between P14 and P80 with no significant differences among any pair of time points (Figure 4.2A), suggesting limited cell cycle activity from P21-P80. Interestingly, the percentage of H2B-GFP$^+$ cells was further reduced at 645 days to 20.1 ± 1.4 indicating low but detectable amounts of cell division continue in adult and aged mice (Figure 4.2A,B).

As significant changes in the percentage of positive and negative H2B-GFP cells between P14-P80 were not observed, we next examined alterations in H2B-GFP intensity, as this would reveal more subtle changes in cell division that occur. We noted a marked shift in the H2B-GFP$^+$ intensity from 105 at P0 to 103 after 645 days (Figure 1O, 2C). Using a logarithmic decay equation to define the dilution of GFP signal mathematically (see Methods), we also observed increased proliferation at early stages (Figure 4.2D). Our calculations show that tendon cells were dividing at a rate of 19 ± 4.2% per day from P0 to P7 and 9 ± 4.25% from P7 to P14 (Figure 4.2D). Proliferation rates decreased to 3.85 ± 0.07% per day between P14 and P21 and 1.75 ± 0.64% per day from P21 to P80. After P80, tendon cell proliferation was markedly decreased to 0.1 ± 0.13% per day by P600 (Figure 4.2D). Together, these proliferation rates derived from mathematical modeling of H2B-GFP decay and the absolute loss of H2B-GFP over time from our flow cytometry analysis indicate that there are relatively high levels of proliferation at the early postnatal stages. In addition, this cell cycle activity is greatly diminished after one month of age, but not extinguished in adult or aged tendons.

4.2.2 BrdU incorporation analysis identifies a postnatal transition from high to low cell cycle activity

To complement our mathematical model of H2B-GFP decay, we used flow cytometry to quantify the percentage of tendon cells that had incorporated Bromodeoxyuridine

(BrdU), a thymidine analog that incorporates into replicating DNA, for different BrdU administration lengths and stages. We performed intraperitoneal (IP) injection of BrdU and harvested tendons to determine the number of $BrdU^+$ cells after 24 hours. For flow cytometry analysis, highly proliferative organs (gastrocnemius muscle) were used as positive controls, tendon tissues from mice that had not received BrdU treatment were used as negative controls, and *Scleraxis (Scx)-Cre;Rosa:TdTomato*$^+$ mice were used to analyze Scx-descendent tendon cells (Blitz et al., 2009) We found that BrdU injection at P0 resulted in 76 ± 13.8% $BrdU^+$ tendon cells at P1, while at P8 and P22, 26 ± 6.5% and 8.4 ± 3.4% of the tendon cells were $BrdU^+$, respectively (Figure 4.3A, B). In adult mice, we observed that less than 1% of the cells were $BrdU^+$ (P60 = 0.4 ± 0.2%, P370 = 0.5 ± 0.1%). To verify these findings in tissue sections, we injected EdU at P1 and P59 and examined *Scx;Cre;Rosa-TdTom*$^+$ and EdU^+ tendon cells in section one day later. Consistent with our BrdU and H2B-GFP results, we observed more *Scx;Cre;TdTom*$^+$/EdU^+ cells in the Achilles tendon at P2 compared with P60 mice (Figure 4.3C). Interestingly, we also observed noticeable doublets of EdU^+ cells in rows along on the longitudinal axis of the Achilles tendon (Figure 4.3C, B'). This indicates that cells divided and retained their relative position in channels along the long axis of the tendon and is consistent with prior work noting an increase in cells along the longitudinal axis at postnatal stages (Kalson et al., 2015). These results show a high rate of proliferation immediately following birth, and a decrease in the first weeks of postnatal life, specifically after P21, which is consistent with our H2B-GFP mathematical model. However, the low percentage of $BrdU^+$ cells at P60 and P370 suggests minimal turnover in adult tendons. To more accurately quantify the amount of cell cycle activity in adults, we administered BrdU continuously in the drinking water of *Scx-GFP;Scx-Cre;Rosa-TdTom* mice for 90 to 100 days. We found that after long periods of BrdU administration, 4 month old mice had incorporated BrdU into 2.35 ± 1.2% of the *Scx-Cre;Rosa-TdTom*$^+$ cells and 2.75 ± 2.9% of the

Scx-GFP$^+$ cells (Figure 4.3D, E; Sup. Fig. 1B), using flow cytometry. Quantification of BrdU stained tendon sections showed a similar percentage of 2.38 ± 2% BrdU$^+$ cells (Figure 4.3F), further supporting a low, but detectable rate of turnover in adult mouse tendons. In mice older than 1 year of age, 90 days administration of BrdU yielded 0.48 ± 0.26% of BrdU$^+$ tendon cells (Figure 4.3D), however, this decrease was not statistically significant between 4, 6 and 13 month stages.

4.2.3 Dynamic gene expression changes occur during the transition in cell cycle state

Since we have determined that there is a transition in cell cycle state during the first postnatal month, we also predict that there are dynamic gene expression changes occurring during this period, especially for genes important for proliferation and matrix production. We performed RT-qPCR assays on RNA isolated from whole distal limb tendon homogenate for a small set of transcripts. These assays provide further information about cell proliferation (*Ki67*), tendon cell identity and differentiation (*Scx*, *Mkx*), and matrix production and assembly (*Col1a2*, *Col3a1*, *Fmod*), during tendon growth. An analysis of variance (ANOVA) on ΔC_T values for each gene demonstrated a significant change in expression of all genes across the developmental range ($p < 0.05$). Tukey's Honestly Significant Difference (HSD) post hoc tests revealed the specific pairs of time points for which relative expression is significantly different.

For many of the genes, relative expression levels decreased during the first month of age. Although KI-67 protein expression is commonly used as a marker of proliferating cells, *Ki67* mRNA expression has been shown to correlate with protein levels and the number of KI-67 positive cells seen in histological sections (Prihantono et al.; Schleifman et al.). Based on this, we examined *Ki67* transcript levels as another independent way to assess the number of mitotically active cells. *Ki67* gene expression

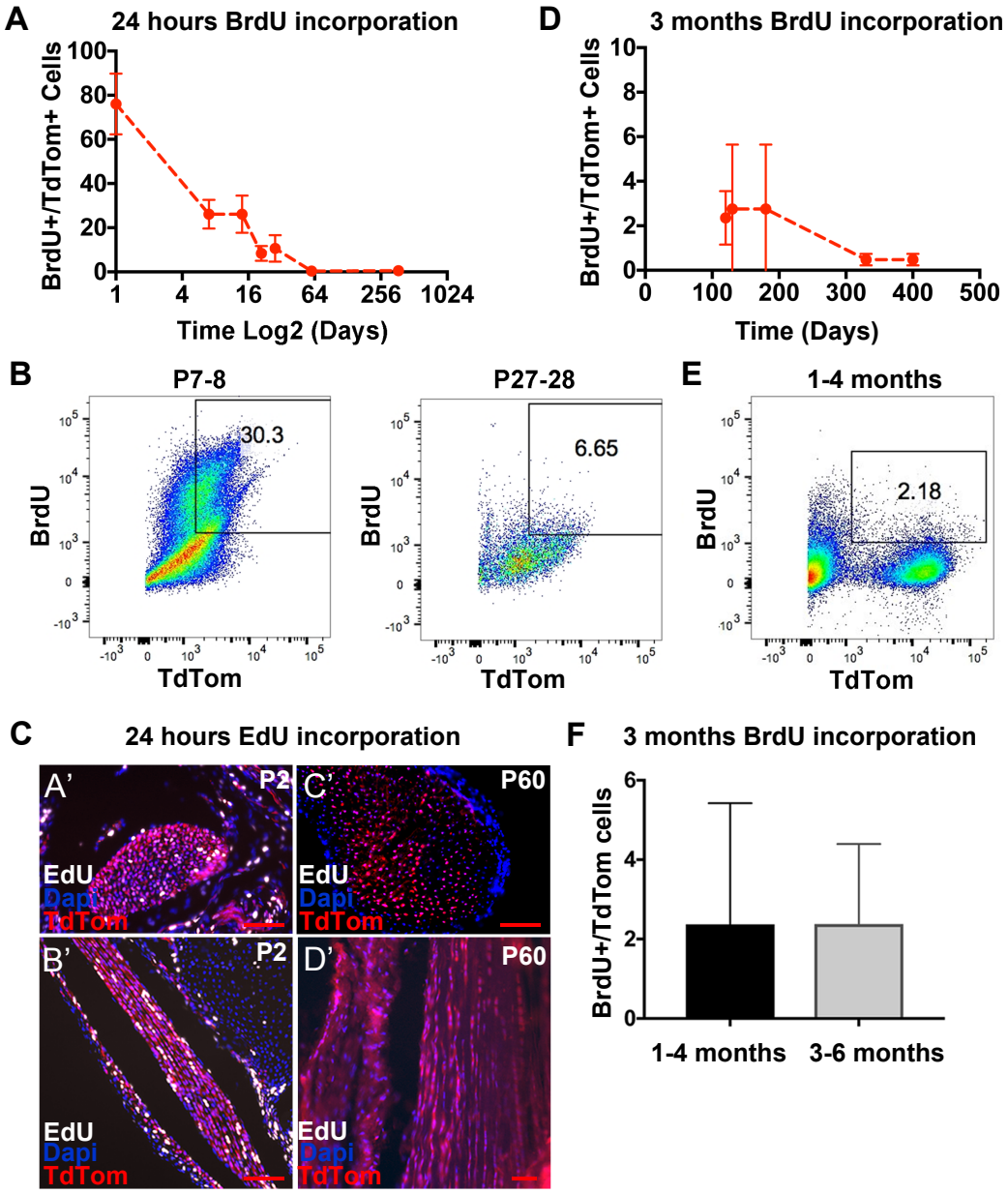


Figure 4.3: Analysis of tendon cell proliferation using short and long pulses of BrdU and EdU.

was highest during the first week after birth (P0 to P7), and no significant differences were observed between P0, P7, and P14 (all $p > 0.8$; Figure 4.4). By P21, however, the relative amount of *Ki67* mRNA present in the tendon became significantly reduced compared to earlier timepoints (P0, P7, and P14, all $p < 0.05$) and remained low throughout the rest of the time series. By P35, *Ki67* expression levels approached the lower limit of detection for our RT-qPCR assays (C_T values ~35). Therefore, these results suggest that the number of proliferating cells is highest during the first week after birth, but by P35 most tendon cells are no longer mitotically active. The expression of *Scx*, *Mkx*, *Col1a2*, and *Col3a1* measured via RT-qPCR also decreased by P35 compared with P0, while *Scx* alone shows significantly increased expression at P14 relative to birth and later stages (Figure 4.4). *Fmod* expression follows a different pattern, however, with higher transcript measurements at all timepoints from P7 to P28 compared to P0; however, none of these differences in *Fmod* expression achieved statistical significance during post hoc testing (Figure 4.4).

4.2.4 Tendon cell density and tendon length undergo dynamic changes during early postnatal stages

To understand how tendon cell number changes relative to matrix expansion during growth, we also quantified tendon cell density during the first postnatal month. Using transverse sections from Achilles and plantaris tendons, we counted cells at P0, P14, and P28. As has been previously reported (Kalson et al., 2015), we observed a decrease in cell density per unit area, with 76 ± 25, 50 ± 17, and 20 ± 7 cells per 0.01 mm^2 at P0, P14, and P28, respectively (Figure 4.5A). These results suggest that matrix expansion outpaces cell proliferation, at least for the cross-sectional area of the tendon. We also noticed a larger variability in cell density at earlier stages compared to P28, which could indicate that natural variability in growth rates during

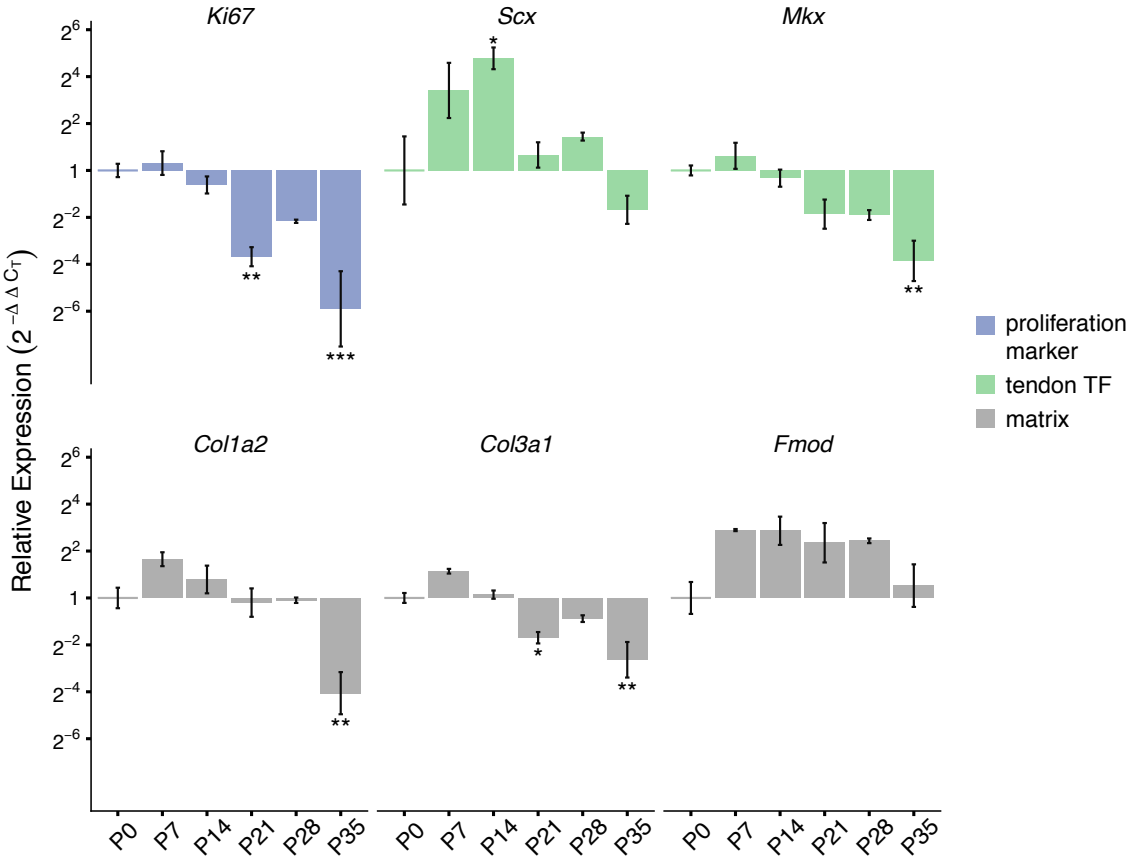


Figure 4.4: Expression of tendon and matrix related genes changes during the transition in cell proliferative potential.

early postnatal stages becomes refined by P28. To understand how cell cycle activity compares with longitudinal tendon growth, we measured the Achilles tendon length from the enthesis to its connection with the gastrocnemius muscle at postnatal and adult stages. Strikingly, we observed rapid growth in the early postnatal stages with the Achilles tendon increasing from 12.71 ± 1.9 mm at P0 to 35 ± 1.7 mm at P14, and to 43.6 ± 1.8 mm at P21. However, the length of the Achilles tendon did not change significantly between P21 to P28 (Figure 4.5B), and only increased modestly from 49.6 ± 1.0 mm at P28 to 54 ± 0.7 mm after P30 (P30-P270) (Figure 4.5B). Overall, the time periods where we observed significant increases in Achilles tendon length correspond directly with our observations of periods of active tendon cell cycle activity. This suggests the interesting possibility that, in parallel with matrix expansion, cell proliferation during the first two weeks after birth may in some way contribute to longitudinal growth or result from mechanical or chemical changes that occur during this dynamic longitudinal growth period.

4.3 Discussion

Defining the transition from developmental growth to adult homeostasis is important for understanding functional tissue physiology. Adult tissues range from high self-renewal activity driven by stem cell populations, such as in the blood and intestine, to low or even no self-renewal as has been reported for the liver and heart, respectively. The tendon presents an intriguing case as growth and maintenance involve both its highly organized matrix and the cells that reside within it. Many studies have highlighted the changes that tendon matrix undergoes in growth, adulthood, and aging. However, the activity of the cells as the matrix transitions from growth to maturation is less well understood. Previous work has suggested that cell proliferation in adult tendons is limited (Runesson et al., 2013), but it is unclear when and to what

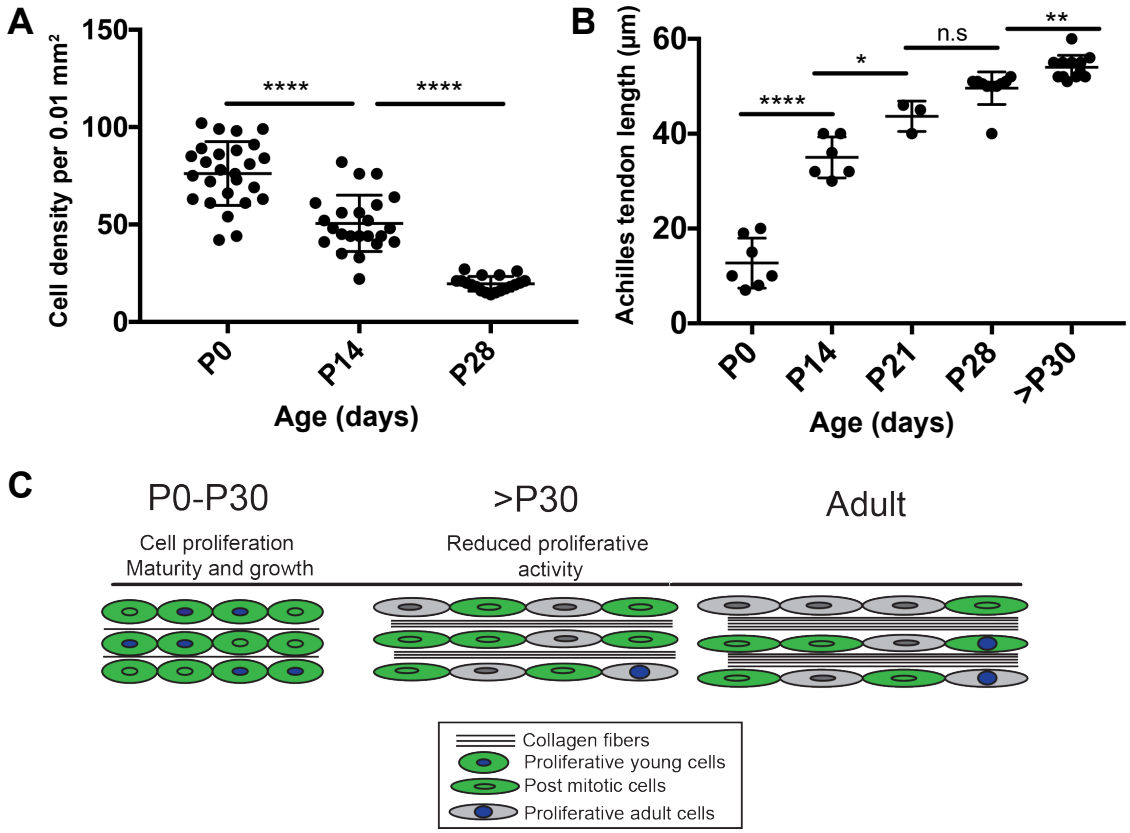


Figure 4.5: Tendon growth measurements and model of postnatal tendon development.

extent this decline in cell cycle activity occurs. Identifying the shift from proliferative growth to homeostasis is also important for properly defining cellular growth periods and understanding self-renewal mechanisms in the adult tendon.

During postnatal development, the tendon ECM undergoes increases in collagen fibril diameter, collagen content, and mechanical properties (Ansorge et al., 2011). In our study, we sought to define the changes that occur to the cells within the tendon during the same periods of growth and homeostasis. Using the H2B-GFP system and BrdU/EdU labeling, we detected significant cell proliferation prior to one month of age. In addition to loss of H2B-GFP, the intensity of H2B-GFP expression decreased demonstrating that all tendon cells divide at least once during postnatal life. The decrease in H2B-GFP intensity across the time series is best described by a logarithmic decay model, which yields proliferation rates similar to those measured via BrdU labeling. Although there were some discrepancies between BrdU labeling and our H2B-GFP mathematical model at P21, these differences were modest and could be attributed to differences in BrdU incorporation into the tendon or a low level (<1%) leakiness of the H2B-GFP system (Sup. Fig. 1C). Despite the potential drawbacks from each method, we obtained similar results from these complementary approaches further strengthening our conclusions. In relation to the growth of the tissue and consistent with others (Dunkman et al., 2013; Kalson et al., 2015), we have observed decreased cell density in postnatal Achilles and plantaris tendons as the mice mature from P0 to P28. This corresponds with rapid elongation of the Achilles tendon in the early postnatal stages with little change occurring from P30 to adulthood. Taken together, our analyses show there is significant proliferation even as the tendon cells are reduced in density. Although this indicates that matrix expansion outpaces cell growth, it also points towards a possible co-regulation of proliferation and matrix expansion during early postnatal stages, which could have interesting implications for how cells regulate, or respond to, ECM expansion and changes in biochemical and

mechanical signals.

Our gene expression analysis also demonstrates interesting changes in the first month of this transition from growth to homeostasis. Relative expression of *Ki67* is significantly downregulated by P21 compared to the earlier time points. In later stages of the time series *Ki67* transcripts are reduced to nearly undetectable levels (CT ~35). Concurrently, the relative expression patterns of tendon transcription factors (*Scx* and *Mkx*), and pro-collagen genes (*Col1a2* and *Col3a1*) largely match that of *Ki67*. Both Scx (Murchison et al.; Schweitzer et al.; Shukunami et al.) and Mkx (Ito et al., 2010; Liu et al., 2014) are involved in tenocyte differentiation, as well as matrix organization via interactions with Smad3 (Berthet et al.). Our findings on the coordinated downregulation of *Scx*, *Mkx*, *Col1a2*, and *Col3a1* after P14 fits within this established framework and suggests that the period from P0 to P14 is a key window of postnatal tendon development. Although matrix genes *Col1a2* and *Col3a1* are down-regulated at P35 relative to P0, elevated *Fmod* expression persists through this time point. This expression pattern is concordant with previous studies of extracellular matrix proteins during the postnatal period in mice (Ezura et al.), indicating that collagen fibril formation slows early, but fibril growth, mediated by *Fmod*, continues into the juvenile period ($>$ 1 month).

The tendon has also been shown to undergo regenerative healing during fetal and early postnatal periods (Ansorge et al., 2011; Favata et al., 2006; Howell et al.). The timing in these studies is reminiscent of the other organ systems such as the heart (Bassat et al., 2017), which demonstrate more regenerative potential at neonatal compared to adult stages. In mice, tendons injured prior to one week of life undergo regenerative healing, with mechanical properties of the healed tendon nearly matching those of the uninjured controls; injured tendons of mice older than 3 weeks of age healed imperfectly through scar formation (Ansorge et al., 2011; Howell et al., 2017).

Previous work has also demonstrated that the regenerative abilities of injured fetal sheep tendons are not affected by transplantation into an adult environment (Favata et al., 2006), suggesting that the regenerative properties of developing tendons are intrinsic. Interestingly, neonatal cardiac regeneration has been attributed to the ability of cardiomyocytes to proliferate during the first 1-2 weeks of postnatal life (Bassat et al., 2017) . It is interesting to speculate that the swift decline in tendon cell cycle activity that we observed at 3 weeks of age may also underlie the shift in regenerative to reparative healing in the tendon.

In addition to defining distinct postnatal periods of cell proliferation, our work also establishes the presence of cell cycle activity in tendon cells at adult and aged stages. Although the levels of proliferation are low, we detected BrdU incorporation in both *Scx*-lineage and *Scx-GFP*$^{+}$ cells in adults. Although our current understanding of the self-renewal mechanisms in the tendon are limited, studies have shown that tendon-derived stem/progenitor cells divide readily and are multipotent when isolated and expanded in culture (Bi et al., 2007). These cells can also form tendon-like tissues upon transplantation (Bi et al., 2007), but how this activity reflects that of resident cells in their native environment is unclear. In the context of injury, recent studies have shown contributions to the healing tissue from *Scx-GFP*-negative cells originating from tendon sheath regions (Dyment et al., 2014; Wang et al., 2017). However, it is unknown if these previously identified cells are responsible for the homeostatic proliferation detected in adults. Other studies have used Carbon-14 (C14) isotope analysis to infer human tendon tissue turnover rates because of known changes in atmospheric C14 levels originating from atomic bomb tests. These studies show that the majority of the tendon core mass is formed by adolescence (Heinemeier et al., 2013). Consistent with this previous study, our results show that most cell cycle activity in the tendon occurs prior to the juvenile stage. However, our work also indicates continued low levels of cell cycle activity in adults. As the previous C14 studies were

performed with tissue samples, which are predominantly matrix, it is unclear, as the authors also note, if they could detect low rates of turnover by a small population of cells. Therefore, even though the C14 results indicate very little tissue turnover after adolescence, they do not exclude the possibility of a slowly cycling tendon cell population in humans. Interestingly, further C-14 analysis of collagen isolated from tendinopathy samples showed evidence of collagen turnover after adolescent periods (Heinemeier et al., 2018). Although it is unclear if the collagen turnover is a cause or effect of the tendinopathy, these results suggest that adult tendon cells can be active at adult stages to remodel their matrix significantly.

In summary, by using complementary genetic and chemical labeling methods, we have gained a comprehensive understanding of the dynamic cell proliferation rates in the tendon from birth to aging. We show that limb tendon cells remain proliferative throughout early postnatal stages (P0-P21) and that mitotic activity declines significantly in juvenile periods with a small population of cells continuing to divide from one month and 1-2 years of life. The timing of these changes in cell turnover appears to be correlated with the timing at which the tendon matrix is undergoing expansion and maturation, as well as when the tendon cells are changing morphology from rounded to stellate. These changes in cell cycle activity also correlate with the transition from regenerative to reparative healing that has been documented in murine tendons. These findings are important to consider in studying tendon growth, maturation and self-renewal mechanisms, and have implications in identifying and characterizing self-renewal mechanisms of a tissue.

CHAPTER 5

INTEGRATIVE GENOMICS REVEALS DISCRETE MODULES INVOLVED IN POSTNATAL TENDON GROWTH IN A MOUSE MODEL

5.1 Introduction

Tendons are crucial components of the musculoskeletal system that transmit muscular forces to bones, thus enabling movement. In adults, tendons are prone to injury, from acute rupture to chronic degeneration. Mature mammalian tendons are incapable of regenerative healing after such injuries; instead, the response to tendon injury is characterized by slow, scar-mediated healing resulting in imperfect tissue repair (Montgomery (1989)). Despite advances in surgical intervention for tendon injuries, the repaired tissue remains mechanically compromised due to this form of healing, yielding a high failure rate after surgery (Beredjiklian et al. (2003); Gomoll et al. (2004)). However, immature mammalian tendons are capable of regenerative healing without the formation of a fibrotic scar, which has been demonstrated in both fetal (Beredjiklian et al. (2003); Favata et al. (2006)) and early postnatal tendon (Ansorge et al. (2011); Howell et al. (2017)). This regenerative ability is retained even when the immature tendon is wounded after transplantation into the adult environment (Beredjiklian et al. (2003); Favata et al. (2006)) suggesting the regenerative potential is intrinsic to the tendon. Regenerative ability in the tendon appears to be restricted to fetal and early postnatal stages, declining significantly with maturation (Beredjiklian et al. (2003); Howell et al. (2017)).

Currently, the molecular factors and cellular states that allow for regenerative healing in mammalian tendon at fetal and early postnatal stages are not well understood.

In fact, little is known about the molecular control of the growth and maturation of healthy postnatal tendon itself. Although some studies have implicated differing expression patterns of specific TGF-β isoforms in fetal vs adult healing (Beredjiklian et al. (2003); Favata et al. (2006)), approaches to activate and inhibit TGF-β have yielded inconclusive results (Alberton et al. (2015); Loiselle et al. (2015)), further obscuring its role in regenerative healing. In mice, it has been shown that the first 1-2 weeks of postnatal life is a window of significantly improved healing potential when compared with 3 week old (Ansorge et al. (2012)) and 4-5 month old mice (Howell et al. (2017)). A similar decline in regenerative ability has been observed in other organs, such as the heart. Cardiac muscle can regenerate after injury prior to 7 days of neonatal development, but no regeneration is observed at later stages (Porrello et al. (2011)); this transition coincides with the developmental stage at which the cardiomyocytes lose their ability to proliferate (Xin et al. (2013)).

We have recently demonstrated a similar phenomenon in mouse limb tendons (Grinstein et al. in review). It has long been appreciated that during development, embryonic tendons proliferate readily as the tissue is forming (Murchison et al. (2007)), but adult tendon cells exhibit minimal or no cell turnover (Messier and Leblond (1960); Ruchti et al. (1983)). Our recent work showed that there is significant proliferation in early postnatal days (P) 0 to 14. However, proliferation decreases significantly by P21 and continues at an extremely low rate from 1 month to 1-2 years of life (Grinstein et al. in review). As in the heart, this drop in tendon cell proliferative activity appears to correlate with the transition from regenerative to reparative healing in mouse tendons around 3 weeks of age (Ansorge et al. (2012)). This decline in regenerative potential of the tendon may be directly related to the decline in tendon cell proliferative activity.

In addition to this marked shift in proliferative and regenerative potential, postnatal tendon growth is characterized by structural and compositional changes to the extra-

cellular matrix (ECM) that produce a highly organized, matrix-rich mature tissue (Ansorge et al. (2011); Connizzo, Yannascoli, and Soslowsky (2013); Kalson et al. (2015)). Both linear and lateral growth in the tendon continue beyond the highly proliferative postnatal phase indicating an important role for ECM expansion during postnatal tendon growth (Connizzo, Yannascoli, and Soslowsky (2013); Ezura et al. (2000)). Thus the transition in cell cycle activity, and regenerative potential, may also correlate with the timing at which tendon cells mature and predominantly function in matrix secretion.

Quite a bit is known about embryonic tendon development. Previous work has identified several transcription factors (TF) and signaling factors that are vital to proper tendon formation. Scleraxis (Scx) is a basic helix-loop-helix TF that has been identified as a marker of tendon cell fate (Schweitzer et al. (2001)). Along with Scx, Mohawk (Mkx) and the early growth response-like (Egr) TFs Egr1 and Egr2 have also been demonstrated to be important for embryonic tendon specification and differentiation (Lejard et al. (2011); Huanhuan Liu et al. (2014)). All four of these TFs are capable of regulating the production of major collagens (Col1a1, Col1a2, Col3a1) and tenomodulin (Tnmd), another tendon marker (Brandau et al. (2001); Shukunami, Oshima, and Hiraki (2001); Lejard et al. (2011); Huanhuan Liu et al. (2014); reviewed in Subramanian and Schilling (2015)). TGF-β signaling also plays an important role in embryonic tendon formation through the regulation of collagen and other ECM proteins (Ignotz and Massagué (1986); Montesano and Sciences (1988); Pryce et al. (2009)) and the induction of Scx (Maeda et al. (2011); C. Mendias et al. (2012); Havis et al. (2016)).

Despite the rich literature on embryonic tendon specification and development, postnatal development has not received as much attention. While postnatal changes to the ECM have been documented in the tendon (Ansorge et al. (2011)), little work has

examined the molecular changes taking place in the tendon cells at postnatal stages. The identification of gene regulatory programs that are specifically controlled during this process would be of great significance towards understanding the mechanisms by which tendon cells mature and tendon tissue grows. Few transcriptomics studies have been performed on tendon cells, with most focusing on the active transcriptional programs in embryonic tendon cells (Havis et al. (2014); Liu et al. (2015)). Other work has examined gene expression differences in isolated postnatal tendon progenitor cells after expansion under *in vitro* culture conditions (Kohler et al. (2013)), which is known to alter the behavior of tendon cells. Expression profiling studies of postnatal tendon have been performed, but these focused specifically on the adult transcriptome to identify tendon-specific genes at a single time point (Jelinsky et al. (2010)) and genes involved in aging (Peffers et al. (2015)). Furthermore, the link between tendon cell growth and healing potential is likely impacted by non-coding regulation of tendon cell differentiation and proliferation. But among all studies to date, none have attempted to identify *cis*-regulatory regions potentially involved in regulating this transition.

In this study, we use an integrative genomics approach to characterize tendon cells as they transition from highly proliferative to relatively quiescent. Specifically, we employ RNA sequencing (RNA-seq) and Assay for Transposase Accessible Chromatin and sequencing (ATAC-seq) to identify expressed genes and open chromatin regions, respectively, in mouse tendon cells as they mature from neonatal stages (P0) to early adulthood (P35). ATAC-seq uses an engineered transposase (Tn5) to simultaneously cut and tag open chromatin for sequencing (J. Buenrostro et al. (2013)). "Open chromatin" refers to regions of the genome that are not tightly wrapped around histone proteins, making them more available to DNA-binding proteins, like transcription factors. Thus, a non-coding genomic locus within a region of open chromatin has the potential to regulate the expression of nearby genes in *cis*. Changes in chromatin accessibility throughout tendon growth can help identify temporally specific putative

cis-regulatory elements involved in the coordination of the transcriptome dynamics involved in the shift in cell proliferative potential. By integrating these transcriptomic and epigenomic signatures, we are able to identify key genes and signaling pathways that mediate the observed shift in tendon cell proliferative potential. This research applies innovative techniques to the study of a poorly understood process, providing new insight and candidate pathways to investigate in the context of tendon growth, maturation, and injury repair.

5.2 Materials and Methods

5.2.1 Experimental Model and Subject Details

5.2.1.1 Animals

Scx-Cre mice were provided by the Schweitzer lab (Blitz et al. (2009); Schweitzer et al. (2001)). *Gt(ROSA)26Sor*$^{tm9(CAG\text{-}tdTomato)Hze}$ (Ai9) were purchased from the Jackson Laboratory (Jax 007909). Male *Scx-Cre*$^{+}$ mice were mated to female mice positive for the Ai9 Cre reporter allele to generate the *Scx-Cre;Rosa:TdTomato* mouse line (henceforth referred to as *Scx-Cre;TdTm*) to mark *Scx*-lineage cells. An equal number of male and female mice were used for all experiments. All mice used in this study were housed at the Center for Comparative Medicine at Massachusetts General Hospital (MGH) and experiments were approved by the MGH Institutional Animal Care and Use Committee (protocol #2013N000062).

5.2.2 Method Details

See (Figure 5.1) for an overview of the RNA isolation and library preparation methods for RNA-seq and ATAC-seq.

5.2.2.1 RNA Isolation

Extraction of intact total RNA from whole tendons was performed as previously described (Grinstein et al. (2018); Grinstein et al. in review). Briefly, distal hindlimb and forelimb tendons were dissected from mice at weekly time points between P0 and P35 and submerged in cold TRIzol (Invitrogen 15596026) immediately following euthanasia. Multiple tendons from a single animal were collected in the same 1.5 mL tube containing 500 μl to 1mL of TRIzol and high impact zirconium 1.5 mm beads (Benchmark D1032-15). The volume of TRIzol used was dependent on the size of the sample. Tendons were first roughly chopped with clean microdissection scissors and then homogenized in two 180-second bouts of bead beating at 50 Hz in a BeadBug microtube homogenizer (Benchmark). Following homogenization, samples were stored at -80°C until extraction using a combination of the TRIzol-chloroform method (Rio et al. (2010)) and Zymo Direct-Zol system (Zymo Research R2050) with an on-column DNaseI digestion (Zymo Research E1010). RNA purity and quality were evaluated using spectrophotometry (NanoDrop 2000c, Thermo Scientific) and capillary electrophoresis (2100 Bioanalyzer and TapeStation 2200, Agilent), respectively. Concentration of each sample was measured via fluorometric quantitation (Qubit HS RNA assay, Invitrogen Q32852) and the final RNA product was stored at -80°C.

5.2.2.2 RNA-seq library preparation and sequencing

RNA samples were excluded from RNA-seq experiments if RIN < 6.7 (threshold determined empirically), sample purity measures (260/280 and 260/230) were poor, and/or if the sample contained insufficient RNA for optimal library preparation. Between 7 and 9 biological replicates (i.e., tendon RNA from independent mice) per time point passed these quality measures, yielding a total of 46 samples spanning 6 time points. To minimize batch effects, all RNA-seq library preparation steps were

completed in a single batch using the Apollo 324 NGS library prep system (IntegenX) and the PrepX mRNA library protocol (Takara Bio) in the Harvard Bauer Core Facility. First, mRNA was isolated from 1 μg total RNA by polyA selection (PrepX polyA 48, Takara Bio 640098) and checked for rRNA contamination using the mRNA 2100 Bioanalyzer chip and protocol (Agilent). The remaining mRNA for each sample was then reverse transcribed and purified using PrepX chemistry (PrepX mRNA 48, Takara Bio 640097). The resulting cDNA libraries were uniquely barcoded and amplified with 11 PCR cycles and then multiplexed into three pools of 16 samples for sequencing. Single end 75 bp reads were sequenced on an Illumina NextSeq 500 using a High-Output 75-cycle kit in the Harvard Bauer Core Facility. Sequenced libraries that achieved at least 20 million reads were used for downstream analyses.

5.2.2.3 Tendon cell isolation and FACS sorting for ATAC-seq

Distal hindlimb and forelimb tendons were dissected from *Scx-Cre;TdTm*$^+$ mice at weekly time points from P0 to P35. Tendons were collected in a digestion buffer containing 0.2% collagenase type II (Worthington L5004176) in DMEM (Gibco 11965-092), coarsely chopped, and incubated in a shaking water bath at 37C. Midway through the incubation the digestion media was spiked with 0.2% collagenase type I (Gibco 17100-017) and 0.4% dispase (Gibco 17105-041). The digested tendons were then gently manually dissociated with a 20G needle and passed through a 30 μm pre-separation filter (MACS 130-041-407) to collect the cells. Tendon cells were washed with 10% horse serum in Hams F10 media (Gibco 11550-043) and finally filtered through the cell strainer cap of a FACS tube (Falcon 352235) for flow cytometry and cell sorting. Tendon cell suspensions were incubated with DAPI immediately before sorting for live/dead exclusion. TdTomato (TdTm) FACS gates were defined based on control TdTm$^+$ and TdTm$^-$ tendon cells that were collected and processed in parallel with the samples of interest. 5,000 TdTm$^+$ cells were collected per replicate in 5%

FBS/PBS. Control samples of 50 ng of naked DNA (i.e., DNA free of histones and DNA-binding proteins) from mice were also subjected to tagmentation and ATAC-seq library preparation in parallel with the tendon samples.

5.2.2.4 ATAC-seq library preparation and sequencing

Cells were pelleted and washed in clean 1x PBS. Transposition was performed using the Fast-ATAC protocol (Corces et al. (2017)) with the following changes: 2 μl of Tn5 transposase (TDE1 from Illumina FC-121-1030) was used in the transposition reaction and the reactions were incubated at 37°C in a shaking water bath for 35 minutes. Transposed DNA was purified using an Omega MicroElute Cleanup kit (Omega D6296) and eluted in 15 μl nuclease free water. The optimal number of PCR cycles for each library was determined via a qPCR side reaction (see Buenrostro et al. (2015)). The purified transposed fragments were then PCR amplified and barcoded as described in Buenrostro et al. (Buenrostro et al. (2015)), followed by bead purification (Omega M1386-01). Library quality was examined via capillary electrophoresis (2100 Bioanalyzer and TapeStation 2200, Agilent) and libraries were quantified using the KAPA library quantification system for qPCR (KAPA KK4824). ATAC libraries were multiplexed and paired-end 42 bp reads were sequenced on an Illumina NextSeq 500 using a High-Output 75-cycle kit in the Harvard Bauer Core Facility.

5.2.2.5 RT-qPCR

Tendons were collected from *Scx-Cre;TdTm*$^+$ mice (n = 4 per time point; 20 total) and RNA was extracted as described above. Because minimal differences were found between P28 and P35 in the transcriptomics analysis, we stopped collection at P28 for RT-qPCR validation. 1 μg of total RNA from each sample was reverse transcribed using the SSIV first strand synthesis system with oligo(dT)20 primers (Thermo Fisher 18091050). SYBR green (PowerUp SYBR, Applied Biosystems A25742) qPCR assays

were conducted in technical triplicate using 10 ng cDNA per reaction (final cDNA concentration = 0.8 ng/μl). Samples were amplified with target-specific primers (see Supp. Table 1 for primer sequences) using a LightCycler 480 II real time PCR system (Roche Diagnostics) as previously described (Grinstein et al. in review).

5.2.3 Quantification and Statistical Analysis

5.2.3.1 RNA-seq Data Processing and Differential Expression Analysis

Sequenced RNA-seq reads were demultiplexed, followed by quality filtering and trimming using TrimGalore (Andrews et al. (2012)). After quality filtering, the final sample size was between 6 and 9 biological replicates per time point. A Salmon transcript index was built from the Ensembl mouse transcriptome (GRCm38v91) and transcripts were then quantified from quality-trimmed reads using Salmon (Patro et al. (2017)) in mapping-based mode with the `--seqBias` and `--gcBias` flags enabled. When used, these options allow the Salmon algorithm to learn and correct for sequence specific and GC biases, respectively, in the data. Gene-level counts were then calculated with TxImport (Soneson, Love, and Robinson (2015)). Genes with consistently low counts were excluded from the data set using automatic independent filtering implemented in DESeq2 (see Bourgon, Gentleman, and Huber (2010) and Love, Huber, and Anders (2014)). Differential expression analyses were conducted using DESeq2 (Love, Huber, and Anders (2014)). We defined two negative binomial generalized linear models of gene expression using RIN as a blocking factor:

$$log(\mu_{i,j}) = \beta_i^0 + (\beta_i^{RIN} x_j^{RIN}) + (\beta_i^{TP} x_j^{TP}) \tag{5.1}$$

$$log(\mu_{i,j}) = \beta_i^0 + (\beta_i^{RIN} x_j^{RIN}) \tag{5.2}$$

where $i = gene, j = sample$. The full model (Equation (5.1)) includes both RIN and time point as predictor variables whereas RIN is the only predictor in the reduced model (Equation (5.2)) . Using a likelihood ratio test (LRT) we compared the two models to identify significant genes that are explained by time point, but not RIN, in order to computationally account for any effects of RNA integrity on the results. The significance threshold for differential expression was set at p adj. < 0.05.

5.2.3.2 ATAC-seq Data Processing and Differential Accessibility Analysis

Demultiplexed paired end reads were trimmed using NGmerge (Gaspar (2018)) and aligned to the mouse genome using Bowtie2 (Langmead and Salzberg (2012)). Read pairs mapping to the mitochondrial genome were removed using removeChrom (https://github.com/jsh58/harvard) and PCR duplicates were filtered with the `dedup` tool in bamUtil (Jun et al. (2015)). Peaks were called with MACS2 in paired end BAM mode with the options `-f BAMPE --nolambda --keep-dup all`. Any peaks falling in the ENCODE defined blacklist for mm10 (Davis et al. (2017)) and peaks called from the naked DNA control were deemed spurious and removed using bedtools (Quinlan and Hall (2010)). Two successful biological replicates were obtained for all time points except P21, which was excluded from all analyses.

We defined a consensus peak set using the DiffBind R package (Ross-Innes et al. (2012)) based on stringent requirements. In order for a peak to be included in the consensus set it had to be present in both biological replicates for at least one time point. Peaks that did not meet this criterion were deemed irreproducible and were excluded from all downstream analyses. We also constrained consensus peak widths to 500 bp. ATAC-seq peaks were annotated to genomic features in R using the ChIPseeker package (Yu, Wang, and He (2015)) and were assigned to the nearest gene within a 100 kb window, defined as $\pm$ 50 kb from the consensus peak summit. DiffBind was used to count reads in these consensus peaks and compute peak differential accessibility

(DA) using the underlying DESeq2 framework. The significance threshold was set at p adj. < 0.05.

5.2.3.3 Clustering and Functional Enrichment

Prior to all clustering analyses, counts of significantly DE genes and DA peaks were normalized using the DESeq2 framework (median of ratios method) and scaled to library size to produce log2 counts per million (CPM). These counts were quantile-normalized and z-transformed. Pearson distance matrices were then calculated for the entire gene or peak set. Clusters were computed in R (R Core Team (2019)) based on this distance matrix and the normalized/transformed counts using the Partitioning Around Medoids (PAM) algorithm (Kaufman and Rousseeuw (1990)) implemented in the ClusterR package (Mouselimis (2019)) with the `fuzzy` option enabled. Heatmaps of clusters were constructed using the ComplexHeatmap package (Gu, Eils, and Schlesner (2016)) and genes (rows) within each PAM module were subclustered via average linkage hierarchical clustering to aid visualization. Functional enrichment analyses of peak and gene clusters were conducted in R using the clusterProfiler package (Yu et al. (2012)) with significance thresholds set at p adj. < 0.01 and q < 0.05. The Benjamini Hochberg method was used for p value adjustment.

5.2.3.4 Integrative Analyses

In order to compare the patterns of gene expression and chromatin accessibility, DA peaks were clustered as described above (Methods, Clustering and Functional Enrichment) into five modules. Next, a matrix of normalized counts for each peak's assigned gene was generated and this matrix was aligned with the matrix of normalized DA peak counts. We computed the Pearson correlation coefficient for each peak-gene pair using the 'lineup' R package (Broman et al. (2015)). For co-clustering of expression and accessibility patterns, the data set was filtered for a Pearson correlation

coefficient > 0.5. This matrix of positively correlated of peak-gene pairs were re-clustered into five modules with the PAM algorithm. Gene lists for the final integrated clusters were used for functional analyses as described above (Methods, Clustering and Functional Enrichment). For each integrative cluster, known motifs were identified within peaks and putative target gene promoters using the HOMER motif finding algorithm (Li et al. (2017)). Per cluster motif enrichment calculations based these on these motif predictions were facilitated by the 'marge' (Amezquita (2019)) and 'valr' (Riemondy et al. (2017)) packages for R.

5.2.3.5 RT-qPCR statistics

Cycle threshold (C_T) values for all targets were normalized to *Gapdh* and the $\Delta\Delta C_T$ method (Livak and Schmittengen 2001) was used to calculate relative gene expression for visualization. Relative expression is visualized as $2^{\Delta\Delta C_T} \pm$ standard deviation. Statistics were computed on the ΔC_T values (Supp. Tables ##). Statistical differences among the time points were investigated using a Kruskal-Wallis rank sum test followed by a Dunn test with Benjamini-Hochberg correction to test for specific differences among pairs of time points. Statistics were performed on the ΔC_T values (normalized to *Gapdh*) for all target genes and samples (n = 4 biological replicates per time point). R statistical software (R Core Team (2019)) was used for all RT-qPCR data analysis, statistics, and visualization. Statistical analysis was performed using the implementation of the Kruskal-Wallace test in 'stats' (version 3.5.3) (R Core Team (2019)) and the Dunn test from 'FSA' (version 0.8.25; Ogle, Wheeler, and Dinno (2019)).

5.2.3.6 Software and Computing Environment

Processing and large-scale analysis of sequencing data was performed on the Harvard Odyssey computing cluster (centOS7). Python programs were run in Python version

2.7.12. R programs were run in R version 3.5.3 (R Core Team (2019)) and RStudio version 1.0.143 (RStudio Team (2015)). Data analysis and visualization in R was assisted by R packages included in the Tidyverse collection (Wickham (2017)), 'ggplot2' (Wickham (2016)), and 'viridis' (Garnier (2018)).

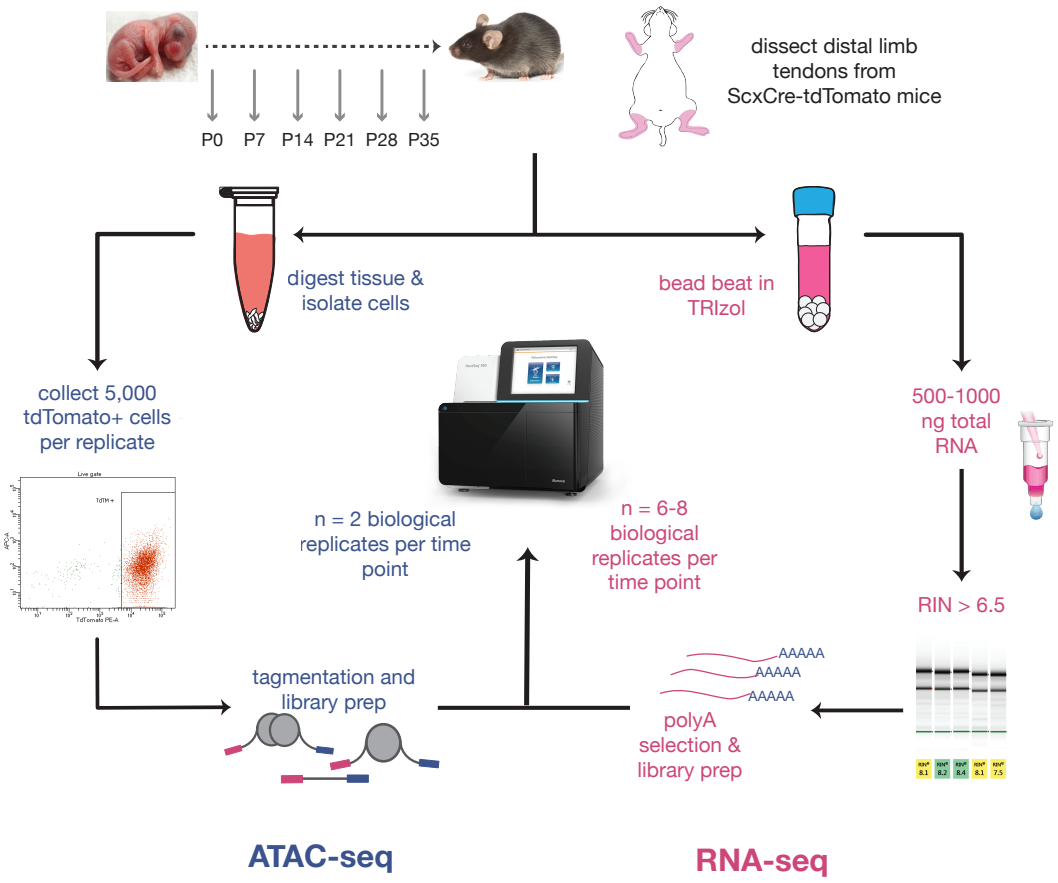


Figure 5.1: Overview of method details for RNA-seq and ATAC-seq.

5.3 Results

5.3.1 Transcriptome dynamics during postnatal tendon growth and development

To identify genes involved in the postnatal transition from a proliferation-based growth program to one driven by ECM, we performed RNA-seq on mRNA isolated from

whole tendon homogenate collected weekly at six consecutive time points during the early postnatal period (P0 to P35). A Principal Components Analysis (PCA) on normalized gene counts shows separation of P0, P7, and P14 samples along PC1, while samples from mice at P21 and older do not separate clearly from one another (Figure 5.2). It is also notable that P35 has higher within-group variability than the other time points, which could at least partially explain the inability of this time point to separate from P28. Importantly, the PCA also shows that sequencing pool does not significantly influence sample clustering (Figure 5.2) indicating that the group in which a sample was sequenced (sequencing pool) accounts for negligible, if any, sample variance. A likelihood ratio test on gene counts using the DESeq2 framework (Love, Huber, and Anders (2014)) found that approximately 22% of detected genes were differentially expressed (DE) between at least two time points in the 5 week time series (p adj. < 0.05). This approach, in which two user-defined models of gene expression are assessed for their goodness of fit, allowed us to test the effects of the time component of the model on gene expression independent of RIN (see Methods), as well as investigate differences in expression among all time points simultaneously. Because our previous work on tendon cell proliferation dynamics (Grinstein et al. in review) points to the second to third postnatal week as an important transitional period during postnatal tendon development, we first investigated the specific pairwise differences in transcriptome wide gene expression from P0 to P35 and found that many genes are differentially expressed between these three time points (see Figure 5.2). Given the assumptions in the tendon literature that mature tendon cells exhibit relatively low metabolic activity, we were quite surprised by the large number of genes (2,508) that were up-regulated from P0 to P35. Comparing the number of differentially expressed genes between each pair of time points shows few genes that are differentially expressed between sequential time points at older stages (i.e., after P14) compared to younger stages (Figure 5.2). This suggests that the first two postnatal weeks are

a period of rapid, dynamic transcriptomic change, which slows after P21 mirroring tendon cell proliferation dynamics (Grinstein et al. in review).

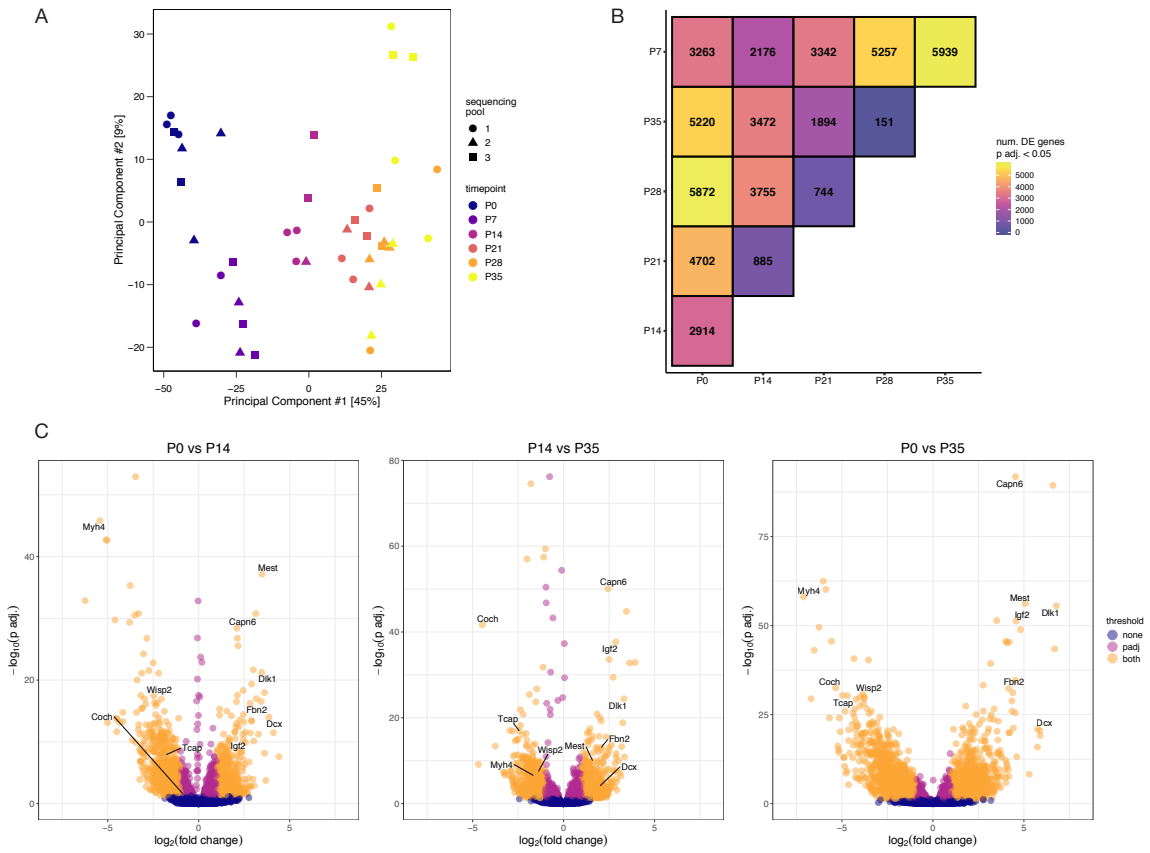


Figure 5.2: RNA-seq identifies differentially expressed genes at key time points during postnatal growth.

5.3.1.1 Expression of known genes involved in tendon development

Next we defined a set of "tendon" and ECM genes (i.e., genes that code for matrix proteins) based on the current literature to investigate their expression within this data set. Because so few genes are known to be directly involved in tendon development, this set is small enough to examine the expression of each gene individually (Figure 5.3). *Scx*, *Mkx*, *Egr1*, and *Egr2* area all expressed postnatally, but none are significantly DE (Figure 5.3A; p adj. > 0.05). *Scx* expression is highly variable among the individuals within a sample; each time point contains distinct high- and low-expressers of *Scx*

(Figure 5.3A). While we have observed higher than expected within-sample variance in *Scx* expression before (see Grinstein et al. in review), the cause remains unclear.

We also investigated the expression of TGF-β ligands and receptors, which have been well studied in the tendon and have been demonstrated to be required for normal tendon development (Pryce et al. (2009)). We found that *Tgfb1* is differentially upregulated at P7 only (p adj. < 0.05), while *Tgfb2* and *Tgfb3* are not DE at any point (Figure 5.3). TGF-β receptor expression varies: *Tgfbr1* is intermittently DE, *Tgfbr2* is not DE, and *Tgfbr3* is differentially upregulated from P0 to P35 (p adj. $<$ 0.05; Figure 5.3B).

All ECM related genes in this gene set are significantly DE at some point during the time series, however (Figure 5.3; p adj. < 0.05). *Decorin* (*Dcn*) and *Fibromodulin* (*Fmod*) expression increases monotonically from P0 to P35; *Col2a1*, *Col14a1*, *Col3a1*, and *Tenomodulin* (*Tnmd*) expression decreases steadily from P0 to P35; and *Col1a1* and *Col1a2* expression peaks around P14 after which it is downregulated. Expression of the matrix gene *Biglycan* (*Bgn*) appears to be partitioned into two phases: it is more highly expressed from P0-P14 followed by downregulation by P21 (p adj. $<$ 0.01), after which point expression remains low through P35 (Figure 5.3C).

5.3.1.2 Clustering analysis and expression module identification

In order to gain a more detailed understanding of the temporal gene expression changes that occur between P0 and P35, we performed an unsupervised cluster analysis using the partitioning around medoids (PAM) algorithm on normalized gene counts. This allowed us to discover six differential expression modules – cohorts of genes that tended to be co-expressed over the course of the six time points in the study (Figure 5.4). These co-expression modules reveal 3 broad patterns of differential expression within our data: downregulation over time (Clusters 1 and 4); upregulation

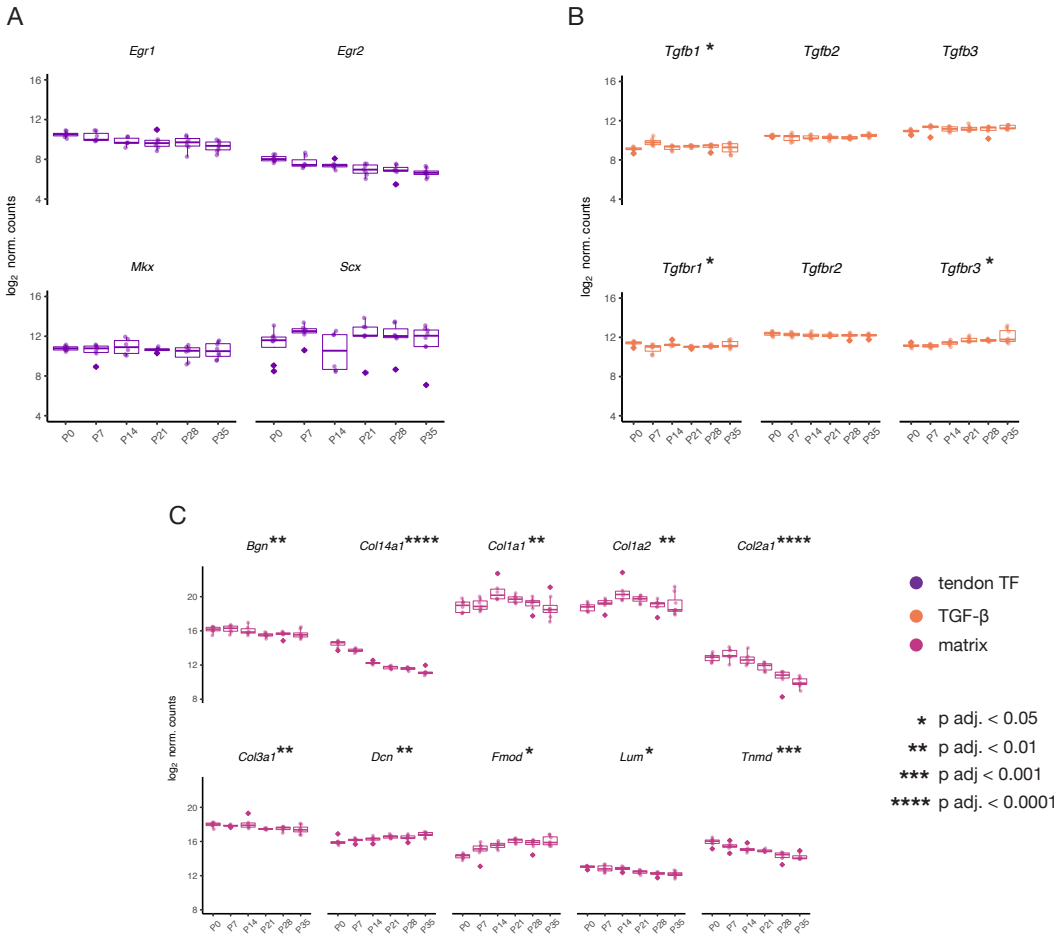
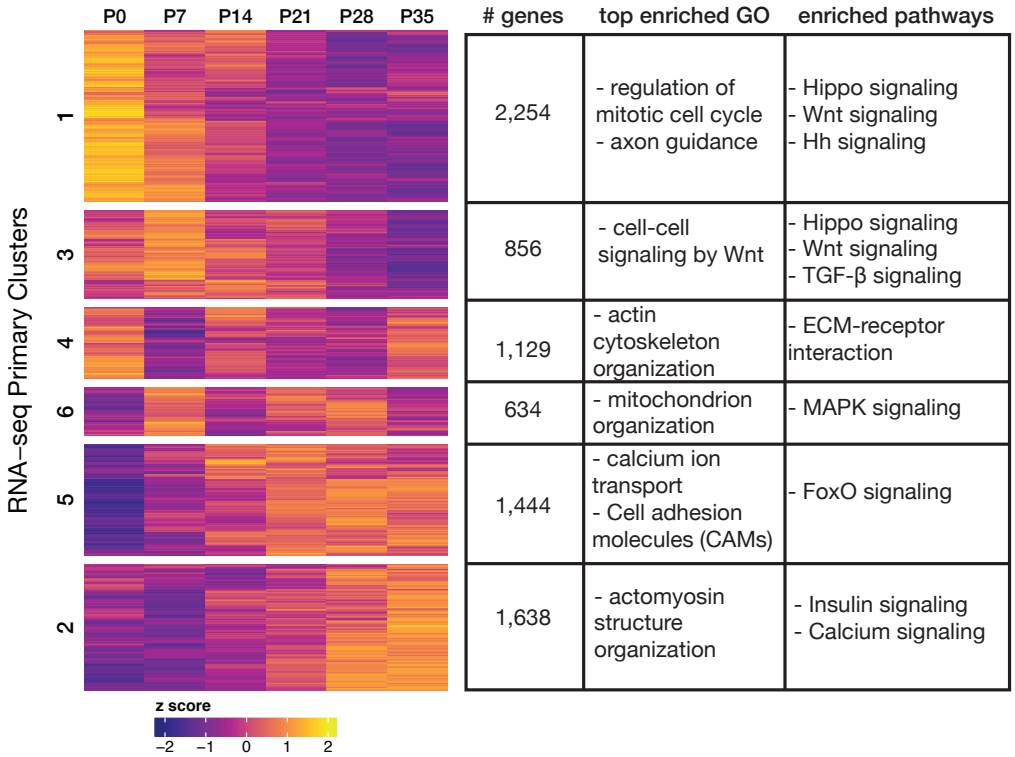


Figure 5.3: Expression of known "tendon" and ECM genes measured with RNA-seq.

over time (Clusters 2 and 5); and intermittent expression (Clusters 3 and 6). The key difference between each pair of expression modules that fit a pattern is the exact timing of the expression change(s). For example, many of the genes grouped into cluster 5 are more strongly upregulated about one week earlier (~P21) than those in cluster 2 (~P28). Gene ontology (GO) enrichment analyses on each expression module suggest that biological processes involved in cell proliferation and differentiation dominate the earlier time points, while cell communication and cytoskeleton organization become more important later during postnatal development (Figure 5.4).



# genes	top enriched GO	enriched pathways
2,254	- regulation of mitotic cell cycle - axon guidance	- Hippo signaling - Wnt signaling - Hh signaling
856	- cell-cell signaling by Wnt	- Hippo signaling - Wnt signaling - TGF-β signaling
1,129	- actin cytoskeleton organization	- ECM-receptor interaction
634	- mitochondrion organization	- MAPK signaling
1,444	- calcium ion transport - Cell adhesion molecules (CAMs)	- FoxO signaling
1,638	- actomyosin structure organization	- Insulin signaling - Calcium signaling

Figure 5.4: RNA-seq identifies modules of differentially co-expressed genes during postnatal tendon growth.

Cluster 1 is highly enriched for GO terms related to mitosis, cell cycle, and DNA replication (p. adj $< 2 \times 10^{-18}$, q $< 5 \times 10^{-16}$). Additionally, a pathway analysis on

Cluster 1 found an enrichment of genes involved in the Wnt and Hippo signaling pathways, both of which are known to be involved in regulating cell proliferation in multiple tissues. Cluster 3, which contains genes that are expressed intermittently throughout the five-week time series and highest at P0, is also enriched for genes involved in Wnt signaling, as well as Smoothened (Smo) signaling, a key component of the Hedgehog (Hh) signaling pathway. This expression module is also enriched for muscle cell proliferation and differentiation. Like Cluster 1, the genes belonging to Cluster 4 are also most highly expressed at earlier time points, however most of the genes in this module exhibit peak expression around P7 or P14 instead of P0. Interestingly, enrichment analyses found that the genes comprising Cluster 4 are highly enriched for processes related to regulation of the actin cytoskeleton, ECM organization, and collagen biosynthesis, as well as small GTPase mediated signal transduction.

Both modules that are characterized by a pattern of steadily increasing expression from P0 to P35 (Clusters 2 and 5) show enrichment for biological processes involved in calcium ion transport, regulation, and signaling. Clusters 2 and 5 are also enriched for muscle related GO terms, although they differ in their specifics. Cluster 2 genes, which are upregulated ~P28, are specifically enriched for muscle developmental and differentiation processes, as well as actomyosin structure organization. Meanwhile, Cluster 5, which contains genes that are upregulated earlier, is enriched for GO terms related to muscle contraction and structure, in addition to processes involved in calcium ion transport (*S100a1*). Multiple GO terms associated with the mitochondria are enriched in the gene set comprising Cluster 6, which is discontinuously upregulated, first at P7 and then from P21 to P28. We also found that Clusters 5 and 6 are enriched for GO terms associated with various metabolic processes and other homeostatic functions, suggesting that the middle of this time series represents the beginning of a shift from growth to homeostasis and a change in cell metabolism.

5.3.1.3 Expression of "muscle" genes in the growing tendon

Somewhat surprisingly, half of the co-expression modules were enriched for muscle related GO terms and functions (Figure 5.4). Figure 5.5 illustrates the specific expression patterns of each category of muscle-associated genes identified in Clusters 1, 2, and 5. Interestingly, there are subsets of muscle development genes that are differentially expressed in either direction throughout the time series. Both Clusters 1 and 2 show GO enrichments related to muscle development and differentiation (Figure 5.4), but on closer examination this signal appears to be driven by different gene families. The muscle development signal in Cluster 2 appears to be largely driven by the expression of genes in the MEF2 family (*Mef2c*, *Mef2d*), while the Cluster 1 genes contributing strongly to this signal are myogenic regulatory factors (*Myf5*, *Myod1*, *Myog*). Because these genes are not widely studied in the tendon, and were not expected to be expressed, we sought to replicate these results using alternative techniques for assessing gene expression. RT-qPCR on *Myf5* and *Myod1* using tendon RNA collected from a new cohort of *Scx-Cre;TdTm* mice, the same strain used to generate the RNA-seq libraries (see Methods; Figure 5.5B,C) replicated this result showing downregulation of both genes from early to late time points. Unlike Clusters 1 and 2, Cluster 5 is enriched for genes involved in regulating the assembly of contractile elements in striated muscle cells, specifically actin (e.g., *Acta1*, *Actn3*, *Tmod1*, *Tmod4*) and titin (e.g., *Tcap*). Although not enriched for muscle-specific processes, Cluster 4 does contain many genes involved in the regulation of the actin cytoskeleton (e.g., *Acta2*).

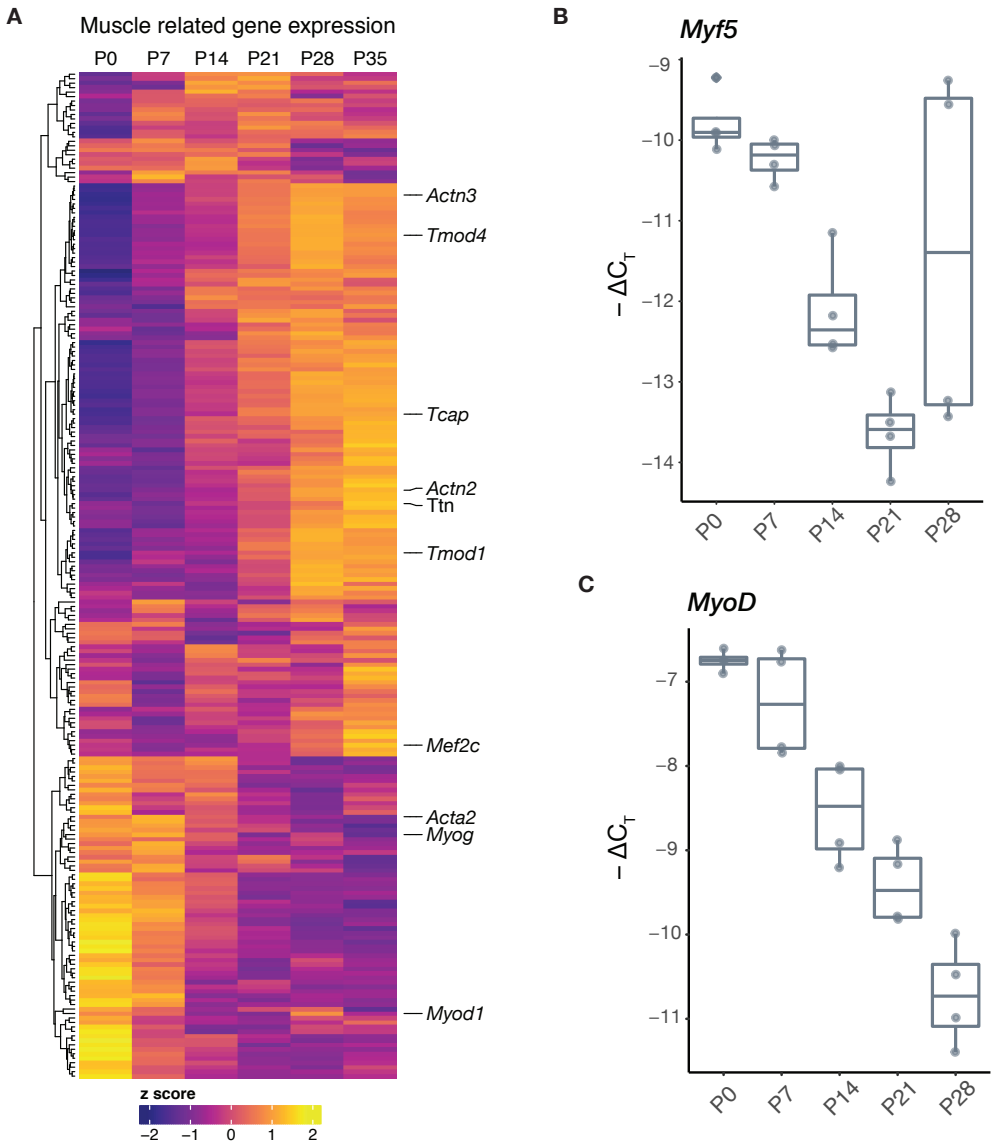


Figure 5.5: Expression of "muscle" genes in the tendon during early postnatal development.

5.3.2 Chromatin Accessibility

5.3.2.1 ATAC-seq identifies accessible chromatin regions in tendon cells

To map chromatin accessibility and putative TF-binding events throughout postnatal tendon development, we performed ATAC-seq on 5,000 FACS-sorted Scx-lineage (Scx-Cre;TdTm$^+$) cells collected from tendons at the same six time points described above. Due to insufficiencies in sequenced library quality, we had to discard both P21 replicates from the analysis. Because Tn5 transposase is known to preferentially cut certain sequences (Green et al. (2012)), we performed a control ATAC assay on naked, or genomic, DNA and sequenced this library along with those from each time point. We controlled for Tn5 sequence preference and other library preparation artifacts by filtering reads found in the control library from the data set for all downstream analyses (see Methods). All sequenced libraries are enriched for insert sizes $<$ 250 bp, which indicates a large number of nucleosome free ($<$100 bp) and mono-nucleosomal (180-247 bp) regions (J. Buenrostro et al. (2013)). Additionally, biological replicates for each time point are well correlated (Pearson correlation $>$ 0.8).

Localized regions of accessible chromatin (peaks) were identified for each replicate using the MACS2 peak calling algorithm (Y. Zhang et al. (2008)) and a consensus peak set for all replicates was defined using the DiffBind package (Ross-Innes et al. (2012)) in R yielding 67,438 consensus peaks that are accessible in both replicates of at least one time point. To determine differential accessibility of peaks over time, reads in peaks were counted for each biological replicate (see Methods) and the DESeq2 framework was implemented in DiffBind (see Methods). A PCA shows that the time points separate well along PC1, which explains 81% of the variance in the data set (Figure 5.6A). However, similar to the transcriptomic data discussed above, P28 and P35 fail to fully separate along PC1 suggesting minimal differences between these two time points. This is supported by the differential accessibility analysis, which finds

only 8 differentially accessible (DA) peaks between P28 and P35. Despite the relative similarities between the two latest time points, 8,358 peaks were found to be DA at some point between P0 and P35 (adj. $p < 0.05$). The majority of these DA peaks are located distal to known transcription start sites (TSS) in non-coding regions of the genome: ~15% of DA peaks are found within the promoter of a gene and ~34% fall in a distal intergenic region, while less than 2% of DA consensus peaks are located within a known exon (Figure 5.6B,C). These results indicate that ATAC-seq identifies dynamic chromatin accessibility in non-coding regions of the mouse genome during postnatal tendon development.

5.3.2.2 Integration of ATAC-seq and RNA-seq data

To test whether any of these DA regions could harbor functionally relevant *cis*-regulatory elements, we first examined the relationship between genome-wide chromatin accessibility and gene expression measured via RNA-seq across postnatal tendon development. ATAC-seq peaks were assigned to the nearest gene (see Methods) and Pearson correlation was computed between peak accessibility and gene expression for each peak-gene pair. Of the 6,193 DA peak-gene pairs for which gene expression was detected throughout the time series, 2,180 (~35%) were found to exhibit a positive correlation between accessibility and expression (Pearson correlation > 0.5) indicating potential transcriptional activation (enhancer) activity by these DA peaks (Figure 5.7). Meanwhile, 1335 (~21%) demonstrated a negative correlation (Pearson correlation $<$ -0.5), potentially indicating that these non-coding regions may behave as repressive elements (Figure 5.7). Although we are interested in general transcriptional regulation throughout postnatal growth, the relationship between transcriptional activators and their targets is more straightforward. Because little is known about *cis*-regulatory control during tendon growth in general, we chose to focus on potential enhancer regions in downstream analyses. We applied the PAM algorithm (Kaufman and Rousseeuw

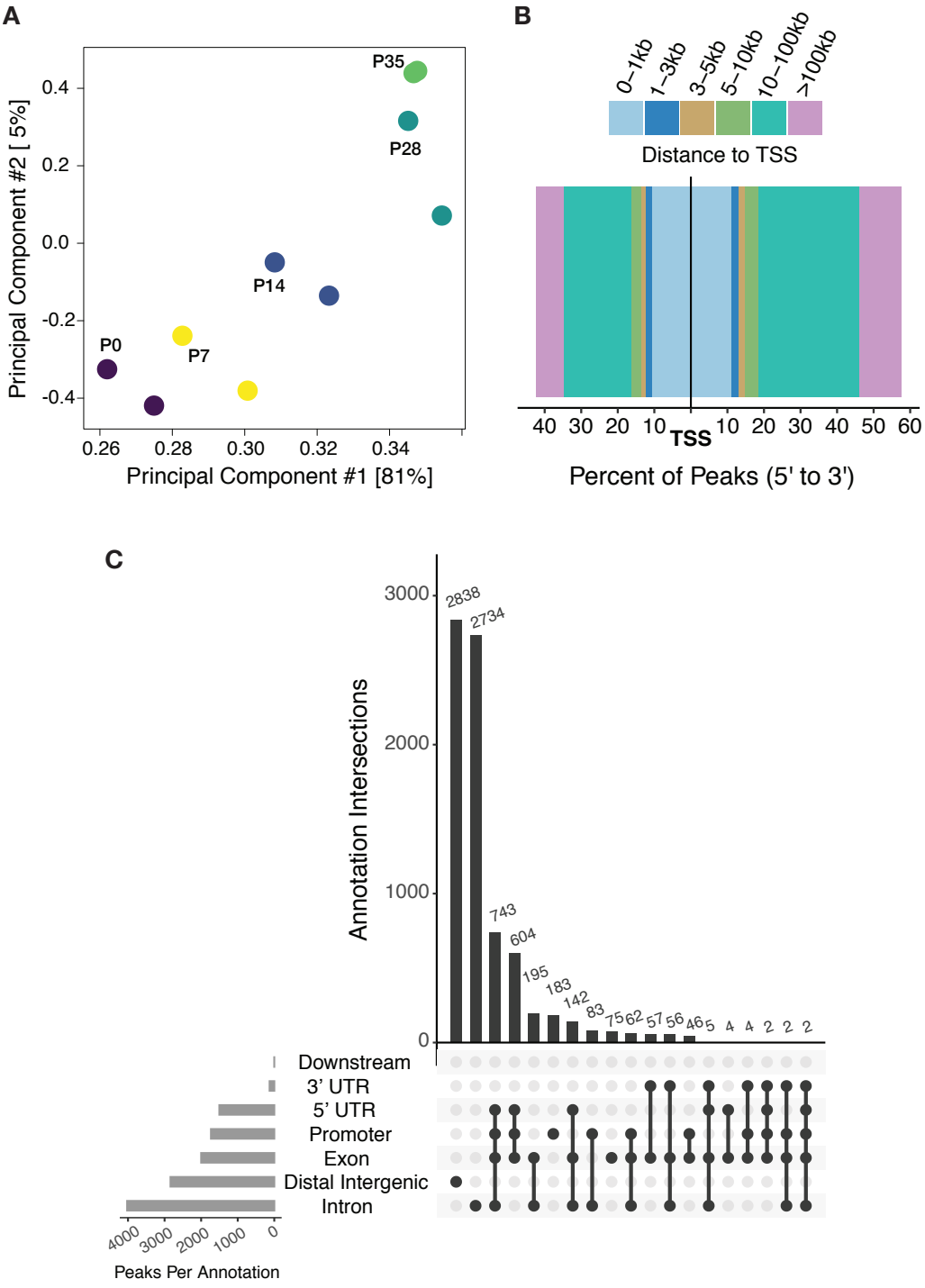


Figure 5.6: ATAC-seq on 5,000 sorted *Scx*-lineage cells detects differential chromatin accessibility during postnatal tendon development.

(1990)) to the subset of positively correlated peak-gene pairs based on their accessibility and expression over time, which revealed three primary patterns of accessibility partitioned into five main clusters (Figure 5.7). Similar to the transcriptomic clustering results, the majority of these peaks and their assigned genes show either monotonic increasing or decreasing accessibility and expression over time. Interestingly, this co-clustering analysis also found a group of peak-gene pairs that exhibit a specific, coordinated increase at P7 (Figure 5.7), similar to the pattern reflected in Cluster 4 from the transcriptomic analysis (Figure 5.4).

Next, we performed enrichment analyses on the integrated modules to determine whether the genes assigned to these putative enhancer regions are involved in similar functions identified based on gene expression alone. Much like the transcriptomic results discussed above, the two modules that characterize a coordinated decrease in chromatin accessibility and gene expression over time (Clusters 1 and 2) are both enriched for peaks associated with Wnt signaling related genes (*Wnt2*, *Fzd7*, *Lgr4*, *Wisp1*, *Wls*; p adj. < 0.005), as well as mesenchymal cell differentiation, proliferation, and organ growth (e.g., *Yap1*, *Igf1*, *Dlk1*, *Jag1*). Interestingly, Cluster 1 is also enriched for peak-gene pairs associated with TGF-β signaling (e.g., *Bmpr1a*, *Bmp2*, *Jun*; Figure 5.7), as is Cluster 5.

The two modules of integrated genomic data that represent upregulation over time (Clusters 3 and 4) are both enriched for muscle-related GO terms, in keeping with the findings of the transcriptomic analyses, as well as MAPK signaling and cell migration (Figure 5.7). The earlier of these modules to become active, Cluster 3, is also enriched for peaks near genes involved in fat and immune cell differentiation, hematopoiesis, and the regulation of reactive oxygen species (e.g., *Runx1*, *Foxo1*, *Foxo3*, *Cebpb*), with quite a bit of overlap between these GO terms. Cluster 4 is more specifically enriched for cell adhesion, integrin binding, and heart morphogenesis-related terms. Among the

genes that appear to be driving this cardiac signal include *Sav1* and *Smad6*, which are known negative regulators of Hippo signaling (through Yap) and TGF-β signaling, respectively. Cluster 5 is also enriched for peaks associated with genes involved in TGF-β signaling (Figure 5.7).

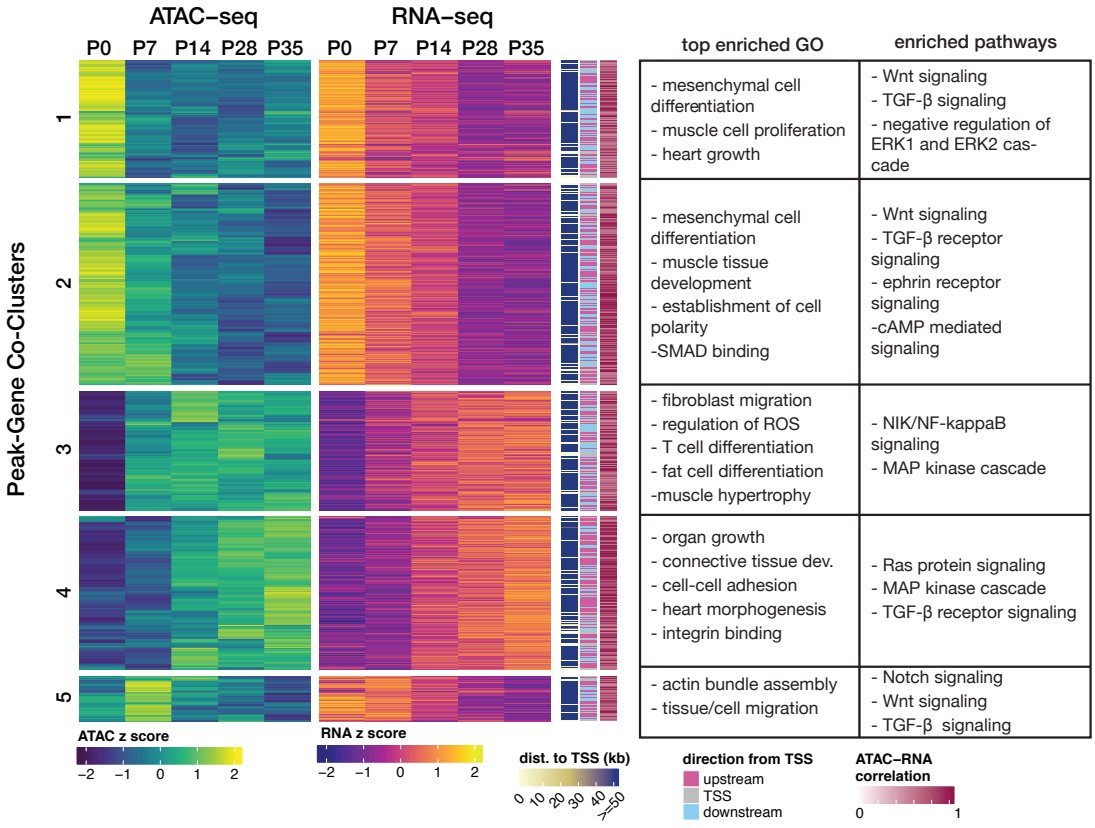


Figure 5.7: Integrative analysis of transcriptomics and chromatin accessibility reveals putative regulatory about the postnatal transition in tendon cells.

5.3.2.3 Motif analysis

Coordinated changes in accessibility and expression are suggestive of potential cis-regulatory activity at these non-coding loci, which is ultimately controlled by TF binding activity within these regions. Thus we sought to identify which TFs may be capable of modulating the activity of these putative enhancers by searching for known DNA motifs within these peaks using the HOMER motif-finding algorithm (Li et al. (2017); see Methods). Multiple motifs were identified in each peak, with slight

variation in the distribution of the number of motifs found per peak across clusters (Figure 5.8B); the majority of motifs were found within 100 bp of the peak summit (Figure 5.8A). Next we performed enrichment analyses to identify motifs associated with specific accessibility and expression patterns in each co-cluster of DA peaks ($q < 0.05$). We found a number of motif enrichments that are shared between modules that change in the same direction with different timing (i.e., Clusters 1 and 2; Clusters 3 and 4), but very little overlap between peaks that become more accessible (Clusters 3 and 4) and those that become less accessible from P0 to P35 (Clusters 1, 2, and 5; Figure 5.8C). The enrichment patterns of "muscle" transcription factor motifs mirrors what was found in the transcriptomic analyses: Myog and Myod motifs are enriched in peaks that are more accessible early, while Mef2a motifs are enriched in peaks that become accessible later (Figure 5.8C). This provides further evidence that "muscle" genes and proteins may play a functional role in postnatal tendon development.

Motifs for Smad3 are also enriched in all modules characterized by decreasing accessibility over time. Cluster 1 is uniquely enriched for several Hox motifs, as well as Cdx2 motifs, another homeodomain protein (Figure 5.8C). However, one of the genes (Hoxb13) does not appear to be expressed in the tendon at any of these time points. The peaks in Clusters 2 and 5, which maintain a high level of accessibility through P7, are enriched for c-JUN and JunD motifs. These proteins are known participants in Wnt signaling (Kan and Tabin (2013); Mann et al. (1999)) and have been shown to be involved in cell growth (Castellazzi et al. (1991)). These clusters are also enriched for other AP-1 family member motifs, including Atf1, Atf2, Atf7. Interestingly, Clusters 3 and 4 are enriched for three other AP-1 family motifs (Atf3, Fosl2, and Fra1) in addition to the motif for AP-1 heterodimers suggesting that similar pathways are being targeted to those in modules showing coordinated downregulation and decreasing accessibility of putative regulatory regions.

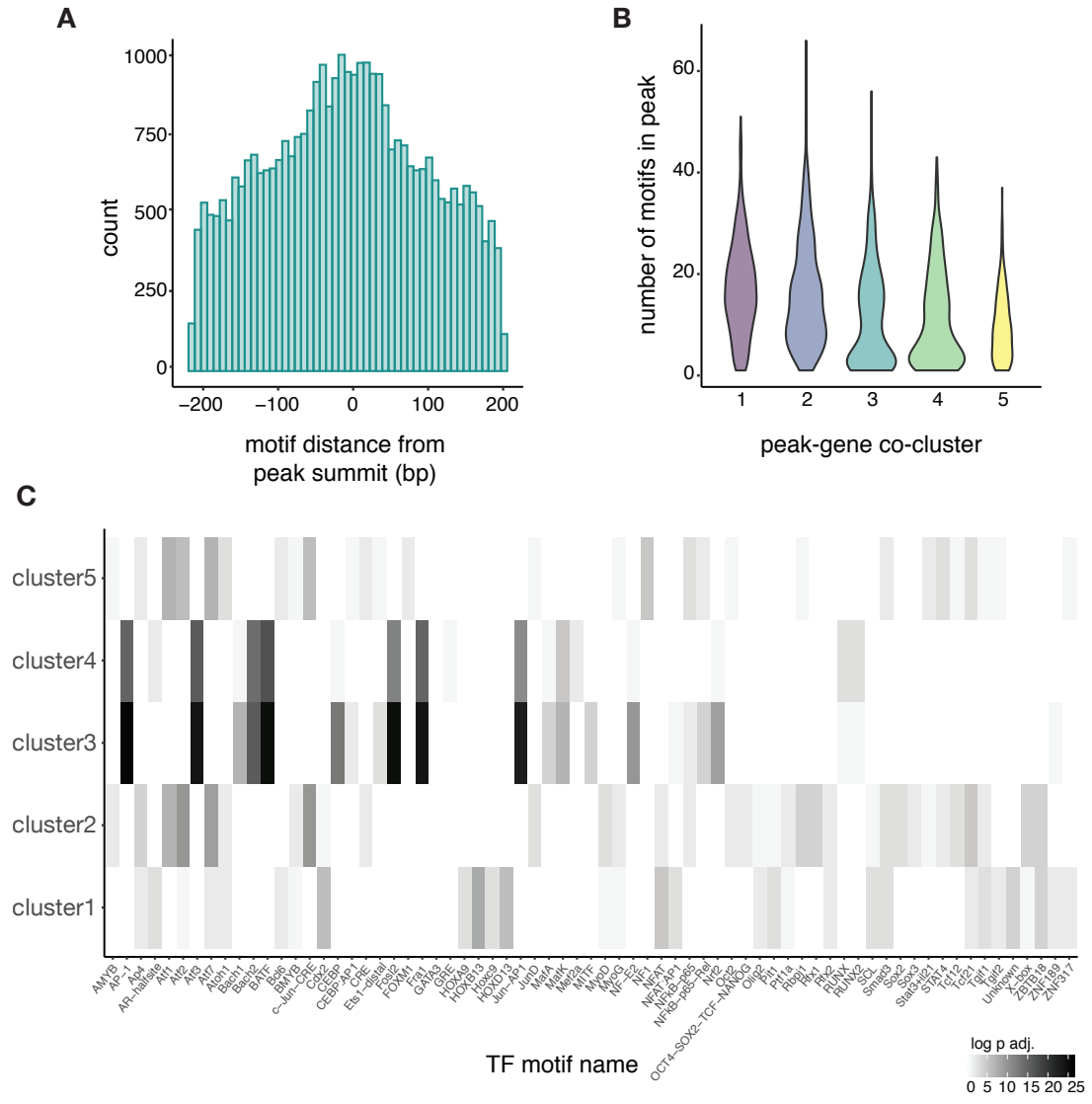


Figure 5.8: Enrichment of transcription factor binding motifs within differentially accessible chromatin regions.

Because the primary model of enhancer function requires interactions between the proteins bound at both the enhancer and promoter (Zabidi and Stark 2016), it is also important to know which TFs are capable of binding the putative target promoter of a given cis-regulatory region. To that end, we identified TF binding motifs in the annotated proximal promoter regions of the genes assigned to these putative enhancers (positively correlated DA peak-gene pairs) and tested whether similar TF motifs, and/or those with known interactions, were enriched in both (Figure 5.9). Overall we found fewer specific motifs enriched for each cluster ($q < 0.05$), but they largely reflected enrichment of the same protein families if not the same TFs. For example, AP-1 family member motifs are, again, enriched in Cluster 1 (Atf4, c-Jun, JunD), as is the Hoxd13 motif – this directly mirrors the results from the distal enhancer motif enrichment analyses. Cluster 2 enhancers, however, are enriched for Smad3 motifs whereas the proximal promoters are enriched for Smad4 motifs. Similarly, Cluster 4 enhancers contain a significant enrichment of Mef2c motifs (Figure 5.8) while the promoters are enriched for Mef2b binding sequences (Figure 5.9). Although no members of the TEAD family motifs are enriched within any of the enhancers in these co-clustered modules, we have identified enrichments of binding motifs for Tead2 (Cluster 4) and Tead4 (Clusters 1, 2, 4) in promoter-proximal regions of the putative target genes of these potential enhancers.

An RT-qPCR analysis of *Yap* and *Taz* expression showed that both of these central signaling factors are significantly downregulated at P28 compared to P0 (Figure 5.10C). This assay also found that transcripts of Tead2, a co-factor of Yap/Taz and an effector of Hippo signaling, is also downregulated throughout this period although the baseline expression is lower than that of *Yap* and *Taz* (Figure 5.10C).

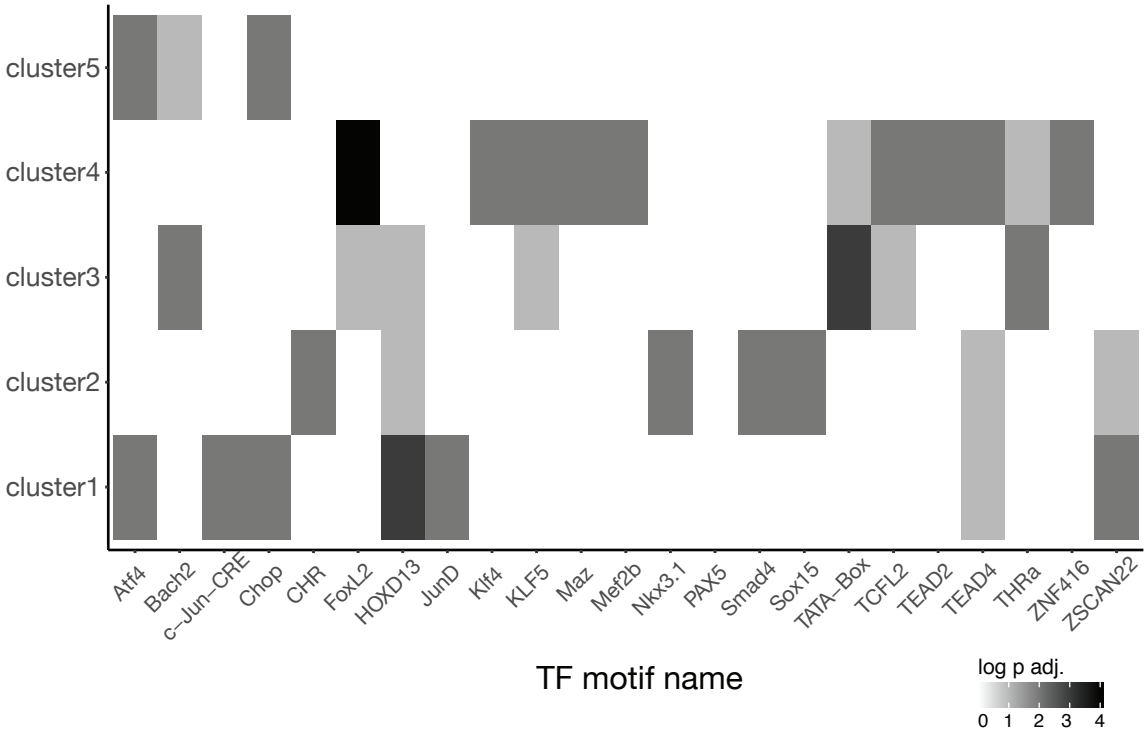


Figure 5.9: Enrichment of transcription factor binding motifs within the promoter regions of putative enhancer target genes.

5.3.3 Wnt signaling during early postnatal tendon development

The Wnt signaling pathway was continually demonstrated by the analyses described above to be over-represented during early postnatal development, i.e., during the period of high levels of tendon cell proliferation. Because Wnt signaling is well known to regulate cell proliferation in many tissues and systems, we chose to probe this pathway more deeply in our data set. Using the GO framework, we defined a cohort of genes known to be involved in Wnt signaling based on annotations in the Gene Ontology (The Gene Ontology Consortium 2019) – the resulting gene list included Wnt ligands, receptors, and known targets. Then we extracted these genes and their assigned ATAC-seq peaks from our integrated accessibility/expression data set and filtered out non-DA peaks and non-DE genes. We then co-clustered the Wnt-specific peak-gene pairs and identified three co-accessibility/co-expression modules (Figure 5.10A). Notably, of the 14 Wnt ligands that are expressed in the postnatal tendon, 8 are DE (Figure 5.10) but only *Wnt2* and *Wnt2b* are both DE and have at least one associated DA peak. *Wnt11* also has an associated DA peak and is expressed at all time points, with slight upregulation at P7, but does not achieve significance after correction for multiple testing (p adj. = 0.09; Figure 5.10A). We performed RT-qPCR to validate expression of *Wnt2*, *Wnt2b*, and *Wnt11* from P0 to P28. These experiments replicated the findings that *Wnt2* and *Wnt11* are significantly downregulated at P28 compared to P0 (p adj. < 0.05; Figure 5.10B). They also showed that *Wnt2b* is significantly downregulated at P14 compared to P0 (p adj. < 0.05), but not at P21 or P28 (Figure 5.10B) although this is likely due to low power due to a small sample size (n = 4). Interestingly, the RT-qPCR assays found that *Axin2*, a negative regulator of Wnt signaling typically used as a read-out of Wnt signaling, is not significantly DE at all throughout this time series (Figure 5.10B).

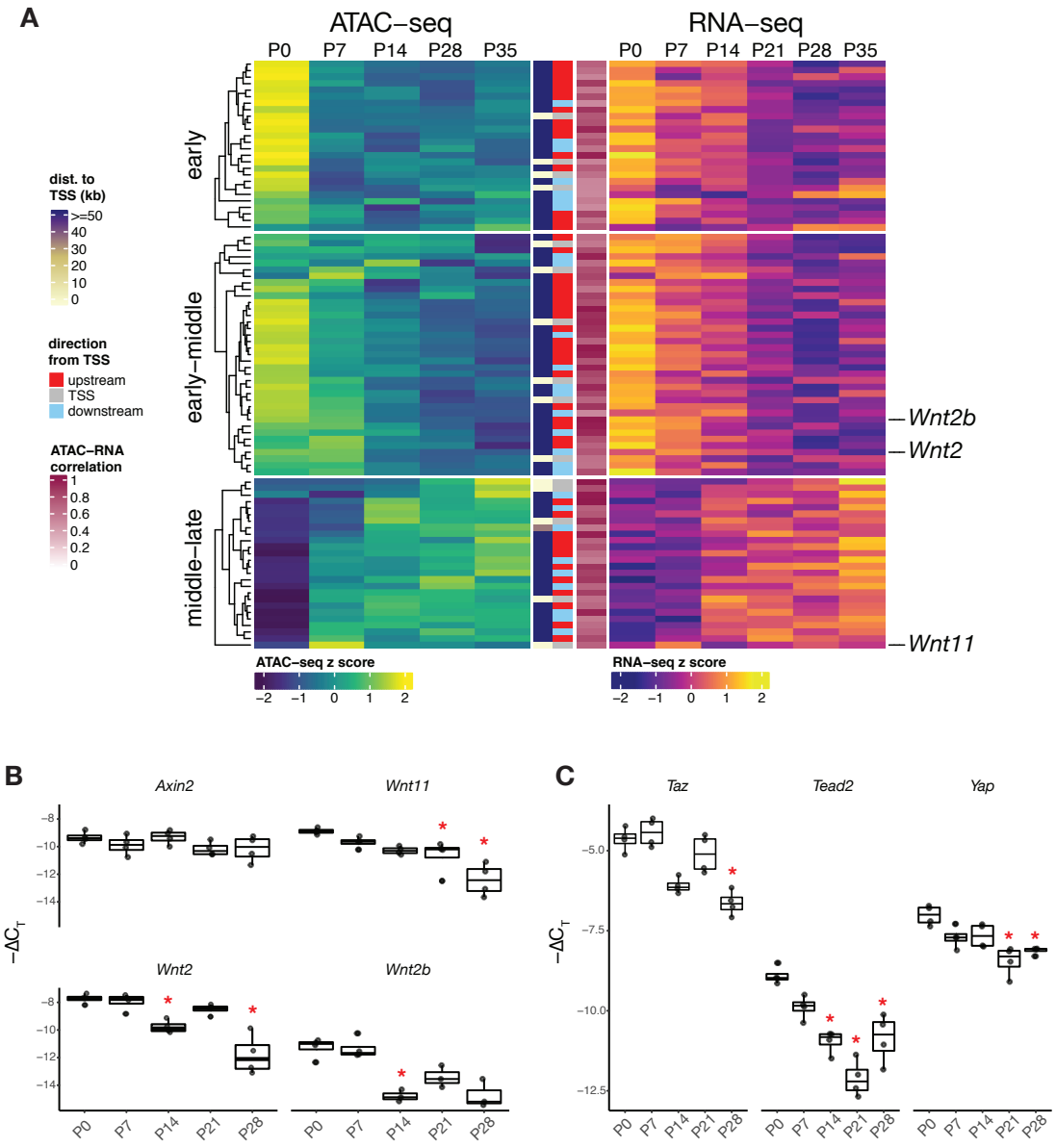


Figure 5.10: Integrative genomics suggests activity of Wnt signaling pathways correlates with period of high cell turnover in postnatal tendon.

5.4 Discussion and Conclusions

5.4.1 Transcriptomics identifies differential expression of known musculoskeletal genes

Previously, little was known about the molecular changes that occur within the tendon during postnatal growth. Using a transcriptomic screen, we identified genes involved in the postnatal transition from a proliferation-based growth program to one driven by ECM expansion. RNA-seq was performed on mRNA isolated from murine tendons collected weekly from birth (P0) to adolescence (P35). Although ours is the first study to examine transcriptome-wide gene expression, some previous work has examined the expression of selected target genes during this developmental period. Therefore, we can compare our results with those limited data to assess the validity of our findings.

5.4.1.1 ECM genes

Most studies of postnatal tendon growth in mice have focused on changes to the ECM due to its known role in overall tendon tissue growth (Connizzo, Yannascoli, and Soslowsky (2013); Ezura et al. (2000); Kalson et al. (2015)). The ECM protein collagen makes up a large portion of the dry mass of tendon (65-80%), most of which is type I collagen (Kannus (2000)). Collagenous matrix production occurs in three phases: collagen molecule secretion and assembly into fibrils (fibrillogenesis); end-to-end fibril assembly to increase length; and lateral assembly to increase fibril diameter. We found that expression of the procollagen genes *Col1a1* and *Col1a2* peaks around P14 in accordance with our previous findings from RT-qPCR (Grinstein et al. in review). Expression of procollagen genes is indicative of collagen molecule production and secretion, which is known to occur after birth (Birk, Nurminskaya, and Zycband (1995); Zhang et al. (2005)). It has also been previously shown that *Col3a1*, *Lum*, and *Bgn* are involved in fibril assembly and the initial stages of growth and are downregulated with age (Ansorge et al. (2011); Birk and Mayne (1997); Ezura et al.

(2000)), which is supported by our data as well (Figure 5.3). In contrast, we found that *Dcn* is upregulated from P0 to P35 (p adj. < 0.01) and *Fmod* is upregulated from P0 to P21 (p adj. < 0.05), at which point expression appears to plateau (Figure 5.3). These genes code for proteoglycans that bind fibrillar collagens and aid in regulation of lateral fibril growth (Vogel, Paulsson, and Heinegård (1984); Hedbom and Heinegaard (1989); Rada, Cornuet, and Hassell (1993)), and the expression patterns shown here largely reflect those previously reported (Ezura et al. (2000); Ansorge et al. (2011); Grinstein et al. in review).

5.4.1.2 "Tendon" genes

In contrast to the dynamic changes seen among ECM genes from P0 to P35, genes known to be important for embryonic tendon development do not change significantly during postnatal development. Although they are all expressed throughout postnatal development, none of the embryonic tendon TF genes (*Scx*, *Mkx*, *Egr1*, *Egr2*) are significantly DE at any point during this time series (p adj. > 0.05). The fact that they are not DE does not preclude their involvement in transcriptional activation during postnatal growth, it simply shows that they are not differentially regulated themselves. These results contradict our own previous findings. We showed via RT-qPCR that *Scx* is significantly upregulated at P14 relative to P0 and P35, and *Mkx* is significantly downregulated by P35 compared to P0, P7, and P14 (Grinstein et al. in review). This discrepancy could be caused by multiple factors. First, RT-qPCR and RNA-seq measurements represent two different classes of data: RT-qPCR measures expression relative to some housekeeping gene(s) while RNA-seq measures absolute counts of transcript abundance and uses a model of transcriptome-wide expression variance to correct for biases in the data (Love, Huber, and Anders (2014)). Second, the RT-qPCR assay was performed on a smaller sample size (n = 3 per time point) than the RNA-seq (n = 6-8 per time point), so it is possible that variation within this small sample

was not representative of the true population variance. For this reason, it is notable that the RNA-seq *Scx* measurements have far higher within-group variability than other genes (see Figure 5.3). High and low *Scx* expressing samples do not partition by sex, and none of the mice contain a *Scx* knock-out or knock-in allele in their genetic background. If such large variation in *Scx* expression is characteristic of postnatal tendon, the discrepancy between the RNA-seq results and those from the previous *Scx* RT-qPCR assay could be due to sampling bias from a highly variable population.

5.4.1.3 "Muscle" genes

The transcriptomics analyses of tendon growth described above yielded a surprising number of differentially expressed genes that are typically associated with muscle. In fact, the size of these "muscle" gene cohorts in three expression modules (Clusters 1, 2, and 5) was substantial enough to influence the GO enrichment analyses (Figure 5.4). Although RNA-seq Clusters 1 and 2 represent differential expression in opposite directions (downregulated vs upregulated, respectively), they are both enriched for muscle development and differentiation-related processes and functions. However, it seems that the direction of DE is partitioned by gene family. Genes in the Mef2 family of muscle TFs (*Mef2c*, *Mef2d*) are upregulated from P0 to P35 (Cluster 2), while the myogenic regulatory factors *Myf5*, *Myod1*, and *Myog*, which are members of the bHLH TF family, are downregulated throughout postnatal development (Cluster 1). Meanwhile Cluster 5 is specifically enriched for muscle-related genes that are involved in contractile element regulation.

Myf5 and Myod1 play a vital role in specifying the myogenic lineage during early embryonic development (Rudnicki et al. (1993)) and regulate the pool of muscle stem cells during adulthood. Myf5 is highly expressed in these satellite cells, allowing them to proliferate after injury (Ustanina et al. (2007)), but Myod1 represses stem cell self-renewal and thus is not expressed in quiescent satellite cells (Asakura et al.

(2007)). In adult mice Myf5 is also expressed by muscle spindles, stretch receptors within the muscle (Zammit et al. (2004)). While *Myf5* gene expression in non-muscle tissues has been reported (e.g., brown preadipocytes, Timmons et al. (2007)), *Myod1* expression is believed to be restricted to the myogenic lineage. To date, neither *Myod1* nor *Myf5* expression has been reported in the tendon. More work will be necessary to determine whether these TFs do indeed serve a functional role in the postnatal tendon.

Despite their association with muscle development, Mef2 genes are expressed in a variety of tissue types. Transcripts of these genes have been detected in developing muscle (both cardiac and skeletal), embryonic chondrocytes, injured smooth muscle, and various parts of the brain (Pon and Marra (2016)). *Mef2c* has been shown to promote smooth muscle cell proliferation (Lin et al. (1998)). It also can regulate the differentiation of cardiomyocytes in concert with Hh signaling (Voronova et al. (2013)), neurons (Li et al. (2008)), and hematopoietic cells (Canté-Barrett, Pieters, and Meijerink (2014)). Both Mef2c and Mef2d have roles in regulating cytoskeletal proteins (M. J. Potthoff and Olson (2007)). The specific source of *Mef2c* and *Mef2d* transcripts in our study is unclear. While they could originate from tendon cells and indicate an upregulation of genes involved in cytoskeletal organization, it is also possible that the increasing Mef2 signal with age is due to angiogenesis and blood vessel invasion of the tendon.

5.4.2 Shift in proliferative potential is correlated with transcriptome dynamics

Unsupervised clustering of significantly DE genes identified six gene co-expression modules that describe three broader patterns of expression. Cluster 1, which contains genes that are downregulated early (~P14-P21), is highly enriched for genes involved in cell cycle. This provides transcriptomic support for our previous findings that

tendon cell proliferative potential declines significantly from P0 to P21 (Grinstein et al. in review). Genes involved in Wnt (Clusters 1 and 3), Hippo (Cluster 1), and Hh signaling (Cluster 3) are also overrepresented in early time points; all three of these pathways are capable of regulating cell proliferation and differentiation, and crosstalk among these signaling pathways has been previously described (reviewed in M. Kim and Jho (2014a); McNeill and Woodgett (2010); Zhao et al. (2010)).

These pathways have been previously investigated for their role in tendon progenitor cell specification and differentiation during embryonic limb development (Zhu et al. (2012)). Exposure of early limb progenitor cells to Wnt maintains their proliferative abilities and increases the expression of soft connective tissue ECM genes such as *collagen 1* (*Col1a1*, *Col1a2*), *tenascin C* (*TnC*), and *Dcn* (D. ten Berge et al. (2008)). Others have also shown that Wnt signaling is sufficient to inhibit chondrogenic differentiation of mesenchymal cells within the developing limb (Hartmann and Tabin (2001)). However, the Wnt ligands at these early embryonic time points appear to be supplied solely by the ectoderm, not the tendon cells themselves, as is shown here. Hh signaling via Sonic hedgehog (Shh) is known to be instrumental in regulating axial tendon progenitor specification along with FGFs (Brent, Schweitzer, and Tabin (2003)), as well as regulating the expression of *Six1* in embryonic limb tendon (Bonnin et al. (2005)). Meanwhile Hippo signaling via Yap has been studied in the context of mechanobiology due to its role in mechanotransduction. Recent work has shown that mechanical factors can influence cell fate, proliferation, and organ size control via Yap and Hippo signaling (Driscoll et al. (2015); Egerbacher et al. (2008); Low et al. (2014); Schiele, Marturano, and Kuo (2013)). However this pathway has not been studied previously during tendon postnatal growth and it is currently unclear whether the rules governing its regulation in other tissues are applicable within the context of postnatal tendon. At the very least, our results are indicative of strong signatures of cell proliferation persisting into the early postnatal period that may be regulated by

some combination of Wnt and Hippo signaling.

5.4.3 Changes in chromatin accessibility reflect transcriptomics and identify putative cis-regulatory regions

Overall, the results from the differential chromatin accessibility analyses reflect those from the transcriptomic analyses alone, including the myogenic signatures. As seen in the differential co-expression modules (Figure 5.4), DNA motifs for myogenic regulatory factors (e.g., Myod1) are significantly enriched within peaks that are more accessible early, while Mef2 motifs (e.g., Mef2a) are more prevalent in regions that become progressively more open during postnatal development (Figure 5.8 and Figure 5.9). While this evidence from DA chromatin does not definitively indicate binding of these myogenic TFs, it is suggestive that they may indeed play a functional role in postnatal tendon development and maturation. Because the RNA-seq was performed on bulk tissue, there was a possibility that these myogenic signatures originated from non-tendon tissue that contaminated the sample. However, ATAC-seq was performed on a sorted population of *Scx*-lineage cells, minimizing the possibility that these results are contamination driven.

The integrative analysis reinforces the finding that Wnt signaling early in postnatal development is correlated with the proliferative period within the tendon (Figs. 7 and 10). Peaks that are more accessible early in postnatal development tend to be associated with Wnt signaling genes (Figure 5.7) and also contain AP-1 family binding motifs, including c-Jun and JunD, which are involved in the Wnt signaling pathway and cell growth (Figure 5.8) (Castellazzi et al. (1991); Kan and Tabin (2013); Mann et al. (1999)). The genes associated with these putative enhancers also contained a significant enrichment of AP-1 binding motifs in their promoter-proximal regions. Genes that code for two specific Wnt ligands, *Wnt2* and *Wnt2b*, are both DE and are

associated with DA peaks (Figure 5.10). *Wnt11* is also associated with a DA peak, but based on the RNA-seq data it was not significant DE at any point during the time series. Thus it is likely that the primary ligands involved in Wnt signaling during early postnatal development are Wnt2 and Wnt2b.

Peaks near TGF-β signaling factors were also found to be enriched in Clusters 1 and 5, which represent early downregulation and P7-specific upregulation respectively (Figure 5.7). Smad3 motifs are also significantly enriched in these peaks that become progressively less accessible over time (Clusters 1, 2, and 5) (Figure 5.8). Smad3 is one of two receptor-regulated Smad proteins and is a key transducer of canonical TGF-β signaling (Massague (2012)). Increasing accessibility/expression of peak-gene pairs that negatively regulate Yap1 and TGF-β from P0 to P35 provides further evidence that these signaling pathways are specific to the proliferative period from P0 to P14, and that they may play a role in regulating this proliferation.

5.4.4 Potential for signaling cross-talk during postnatal tendon growth

Taken together, these data suggest that some level of co-ordination between Wnt, TGF-β, and Hippo signaling is involved in the transition from a highly proliferative to relatively quiescent tendon cell program. Cross-talk between the Wnt and TGF-β signaling pathways has been demonstrated in multiple cell types. In vascular smooth muscle cells, TGF-β/Smad3 stimulates secretion of several Wnt ligands (Wnt2b, 4, 5a, and 9a), which promote proliferation by stabilizing β-catenin (DiRenzo et al. (2016)). Other work in mesenchymal stem cells has shown that Tgfb1 stimulates expression of Wnt2, 4, 5a, 7a, and 10a (Tuli et al. (2003); Zhou (2011)) during chondrogenesis, and that Smad3 and Smad4 interact with β-catenin to activate the Wnt/β-catenin pathway in chondrogenesis (Zhang et al. (2010)) and osteogenic differentiation (Jian et al. (2006)). It has also been shown in chondrocytes that TGF-β inhibits expression

of Axin1 and Axin2 (negative regulators of Wnt) via Smad3, which inhibits TGF-β and promotes Wnt/β-catenin signaling (Dao et al. (2007)).

Additionally, previous work has shown that Smad3 can regulate tendon ECM through physical interactions with Scx and Mkx (Berthet, Chen, Butcher, Schneider, et al. (2013a)), both of which are stably expressed during postnatal development (Figure 5.3). Thus, it is possible that the TGF-β signaling we have observed during early postnatal stages is indirectly regulating cell proliferation via matrix regulation through Smad3. Yap, a key protein in the Hippo pathway, is responsive to changes in matrix mechanical properties and thus may serve as the bridge between TGF-β control of the ECM and the regulation of cell proliferation. Furthermore, Yap and its cofactor Taz can interact with various Wnt transcriptional activators to regulate their cytoplasmic retention in a Wnt-dependent manner. And in the context of fibrosis, Yap/Taz can form complexes with Smad2/3 and Tead proteins, and the concentration of Taz influences nucleocytoplasmic shuttling of Smad proteins (reviewed in Piersma, Bank, and Boersema (2015)). Thus Yap/Taz appear to have the ability to act as orchestrating molecules in the coordination of multiple signaling pathways. Many of these reported mechanisms governing signaling cross-talk are likely cell-type and/or context specific, but they are important demonstrations that these pathways have various ways of influencing each other. Our analyses show that genes involved in the Hippo signaling pathway are differentially expressed throughout the postnatal growth period, and also found that chromatin regions that become progressively less accessible during this time are enriched for Smad3 binding motifs (Figure 5.8). Tead2 motifs, however, are enriched in the putative target promoters of peaks that become more accessible from P0 to P35 (Figure 5.9); our RT-qPCR assays found that *Tead2* expression decreased from P0 to P28. Thus the direction of *Tead2* expression change measured by RT-qPCR is the inverse of the change in Tead2 motif accessibility in these promoters, indicating the need for further experimentation. Of course, it is possible

that some of these results are spurious and will not be replicated in future studies. Assuming, however, that this inverse pattern of gene expression and motif accessibility holds, there are many possibilities for why this may be. For example, perhaps despite the decreased expression of *Tead2* transcripts from P0 to P35, other processes are upregulated to improve Tead2 protein stability and/or its ability to bind Yap/Taz, making higher levels of transcription unneccessary. Because we cannot assume that gene and protein expression exist in a one-to-one ratio, and because post-translational protein modifications are an important part of Hippo signaling regulation, future work to quantify the amount of Hippo effectors and describe their forms at these time points will be paramount to the study of Hippo signaling in tendon development.

5.4.5 Limitations

This study is not without its limitations. First, RNA-seq was performed on bulk tissue, not a purified cell population that was sorted based on a genetic marker. Although tendon tissue was carefully microdissected and cleaned of all visible non-tendon tissue, it is still likely that other cell types were present in the samples given that tendon is not composed of one homogeneous cell population. ATAC-seq was performed on sorted Scx-lineage cells, however, and shows relatively good agreement with the transcriptomics results, giving us more confidence in the bulk RNA-seq.

The ATAC-seq itself is also imperfect. Due to the relatively hypocellular nature of tendons, especially at later stages, we were limited to samples of 5,000 cells per ATAC-seq biological replicate. This constrains library diversity and increases the proportion of duplicate sequenced reads, thereby limiting our power to detect more minor changes in chromatin accessibility. Ideally, each replicate would have achieved at least 50,000 cells, but the intrinsic properties of tendon tissue precluded this without pooling multiple animals. We ultimately decided that, because pooling individuals into

a "biological replicate" obscures true biological variation, it would be more appropriate to perform these assays on samples that originated from a single animal. Secondly, our biological sample size for the ATAC-seq assay is very small (n = 2 per time point). Tendon tissue is difficult to dissociate without harming the cells or changing their behavior. Additionally, more mature tendon cells do not fare well outside of their matrix environment. While a larger sample size (e.g., n = 4) would have been preferable, it was simply not feasible due to the high proportion of lost samples during the dissociation and/or sorting process. Previous studies have performed ATAC-seq on as few as 2 biological replicates and reported reproducible results (J. Buenrostro et al. (2013); Gehrke et al. (2019)), indicating that good quality data are possible from a limited sample. To minimize spurious results, we constructed the consensus peak set using stringent criteria requiring a peak to be present in both biological replicates to be considered for differential accessibility; if a peak was only present in one replicate for a time point it was excluded from all downstream analyses. This likely inflates the number of false negatives in our results, but it is preferred to inflating false positive results.

5.4.6 Conclusions and Future Directions

Given our findings that Wnt, Hippo, and TGF-β signaling appear to be downregulated in a coordinated fashion throughout postnatal development, it is possible that all three of these pathways play a role in the transition from a highly proliferative tendon cell program to a largely quiescent one. We found that *Wnt2* and *Wnt2b* are downregulated from P0 to P35; this expression change is also correlated with a change in chromatin accessibility in the proximal promoter regions of both genes. *Ctnnb1* (β-catenin; the canonical Wnt signal transducer), *Smad3* (a key TF involved in TGF-β signaling), and the Hippo-related TFs *Tead2* and *Tead3* are also consistently downregulated over

the course of postnatal development. *Tgfb1*, a cytokine involved in TGF-β signaling, is significantly upregulated at P7 relative to all other time points. This is particularly intriguing given the subsequent decline in genes related to Wnt and Hippo signaling.

Functional studies targeting specific components of these pathways will be vital for untangling the complicated network of interactions among these signaling factors during tendon postnatal growth. Additionally, further investigation into the putative TF binding sites in the promoter-proximal regions of *Wnt2* and *Wnt2b* would improve our understanding of the upstream regulators of these ligands during postnatal tendon development. It would also be informative to identify putative targets of Wnt signaling during this time frame. Using our genome-wide information about changes in chromatin accessibility, we can identify DA regions that contain motifs for known binding partners of β-catenin and investigate the expression dynamics of their putative target genes.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

6.1 Primary Findings

The studies that comprise this dissertation are among the first to comprehensively characterize the cellular and molecular dynamics of postnatal growth in the tendon. Although embryonic tendon development has been reasonably well studied, investigation of the molecular mechanisms governing postnatal growth and development have been hindered by the difficult nature of tendon tissue. Of the few studies that have investigated postnatal development, they have predominantly focused on changes to the ECM (e.g., Ansorge et al. (2011)). This is likely due to the abundance of ECM proteins compared to cells, especially later in postnatal life, and the relative ease of studying the acellular components. The very features of the tendon that allow it to behave as a viscoelastic spring during human locomotion – a strong but extensible matrix made of hierarchically organized, cross-linked type I collagen fibers embedded within a hydrophilic, proteoglycan-rich ground substance – inhibit the extraction of intact RNA and the isolation of healthy tendon cells from the tissue, which are necessary for performing molecular biology experiments. This becomes even more difficult as the matrix and cells mature, rendering comparisons between early and late time points nearly impossible.

To address this methodological gap, Chapter 3 establishes a novel method for RNA isolation from fresh tendon tissue, which is vital for the investigation of molecular mechanisms through gene expression analyses. In this study, we also demonstrate the dangers of pooling tissue from multiple individuals into a single biological replicate,

a common practice in tendon biology due to the low yield of genetic material from tendons (e.g., Bell et al. (2013); Trella et al. (2017)).

Chapter 4 finds that tendon cells do indeed proliferate during postnatal life. Prior to this work, existing tendon literature generally assumed that healthy postnatal cells are not proliferative due to the observation that tendons become more hypocellular from birth to maturity (Dunkman et al. (2014); Kalson et al. (2015)). Thus, it was believed that postnatal tendon growth was driven by expansion of the ECM, not cell proliferation. Using both genetic and chemical *in vivo* cell labeling methods in mice, we assayed for cell turnover across the tendon growth period from birth to the early juvenile period. In this study, we show that tendon cells are highly proliferative during the neonatal period, but that proliferative potential declines significantly by the beginning of the juvenile period (P21). This suggests that cell growth likely contributes to overall tendon tissue growth early in postnatal life (P0 to P14 in mice) and that ECM expansion takes over the bulk of growth during the juvenile period. These results indicate a transition in the tendon cell program during postnatal development that is likely under genetic control.

Using an integrative genomics approach, Chapter 5 identifies key expression and regulatory modules that are correlated with the changing proliferation dynamics identified in Chapter 4. We identified several key pathways that may be involved in the regulation of tendon cell proliferation during postnatal development, including Wnt, Hippo, and TGF-β signaling. These three pathways have been shown to interact in several other tissue types, suggesting the potential for signaling crosstalk among them in the tendon. Finally, by integrating expression and chromatin accessibility data on a genomic scale, Chapter 5 provides a rich data set for discovery of *cis*-regulatory modules that can aid in the identification of the upstream factors responsible for regulating these coordinated changes in multiple signaling pathways. Together, the

components of this dissertation expand our understanding of the molecular networks controlling postnatal tendon cell growth and maturation.

6.2 Discussion and Future Directions

Although this dissertation was motivated by questions on an evolutionary scale, not enough was known about tendon growth and development to adequately address hypotheses about the ultimate mechanism behind human Achilles tendon elongation. Before we can directly test hypotheses about the evolution of the Achilles tendon in primates, we need to better understand how a normal, healthy tendon grows and develops. While this dissertation provides a step towards building that understanding, our current picture of the molecular mechanisms of tendon growth and development is far from complete. On a fundamental level, we have not yet demonstrated functionality of the pathways and putative enhancers discovered in Chapter 5 either *in vitro* or *in vivo*. And while we have replicated the expression patterns of some key genes identified in Chapter 5 via RT-qPCR, this method suffers from similar limitations to RNA-seq, namely a lack of spatial resolution. Validating expression of both the genes (e.g., using *in situ* hybridization) and the proteins they code for (using immunohistochemistry) will be paramount for understanding the spatial distribution of these molecules within the tendon, as well as validating that these molecular signatures are indeed native to tendon cells. Another key question that remains to be answered is whether the putative enhancers whose accessibility is correlated with postnatal changes in tendon cell proliferation are capable of driving gene expression. A first step towards answering this question will require isolating these potential regulatory sequences and testing their ability to drive expression of a reporter gene (i.e., luciferase) in cells *in vitro*, followed by further assays to determine which transcription factor(s) facilitate enhancer activity and whether the predicted downstream target is correct.

On a broader level, the findings of this dissertation indicate several interesting avenues for future long term study. First, the degree to which Wnt signaling is involved in regulating the shift in proliferative potential during the early postnatal period is unknown. Several tools exist in mouse genetics to help dissect the role of this pathway in postnatal development. A tendon-specific inducible knockout of a core component of the Wnt pathway (e.g., Dishevelled or β-catenin) could reveal the degree to which this pathway is necessary for the progression of normal postnatal proliferation dynamics (see reviews of mouse models in Amerongen and Berns (2006) and Grigoryan et al. (2008)). Because of the functional redundancy of Dishevelled proteins, a double-knockout of *Dvl1* and *Dvl2* may be necessary to fully inhibit Wnt signaling in the tendon (see Amerongen and Berns (2006)). Investigating the effects of β-catenin loss-of-function would be beneficial in clarifying whether β-catenin is indeed the primary transducer of Wnt signaling during tendon postnatal development. Second, the potential for crosstalk between the Wnt and TGF-β signaling pathways is particularly intriguing, especially given the well-established importance of TGF-β signaling in tendon development, healing, and maintenance (Chan et al. (2008); Pryce et al. (2009); Maeda et al. (2011); Berthet, Chen, Butcher, Schneider, et al. (2013b); Havis et al. (2016)) and its ability to promote tenogenesis in both avian and mammalian cell culture (Hoffmann et al. (2006); Lorda-Diez et al. (2009); Kapacee et al. (2010); Barsby and Guest (2013); Yin et al. (2016)). An instrumental first step in determining whether there is a significant effect of TGF-β signaling on Wnt signaling in the tendon would be *in vivo* and/or *in vitro* manipulations of TGF-β signaling via *Tgfbr2* expression, or *Tgfb1* specifically, and assaying for expression of Wnt ligands previously shown to be affected by TGF-β signaling in mesenchymal cells (e.g., *Wnt2*) (Tuli et al. (2003); Zhou (2011)).

Although deciphering the molecular dynamics among three signaling pathways is a complicated endeavor and requires detailed biochemical analysis, the question

of how Hippo signaling may interact with both Wnt and TGF-β signaling during tendon development is a fascinating one. Currently, little is known about the role of Yap and Hippo signaling in the tendon. Tendon cells are highly responsive to mechanical loading (Banes et al. (1999); reviews (2004); Heinemeier and Interact (2011); Zhang and Wang (2013)) and Hippo signaling through Yap is an important mechanotransduction pathway in many tissues, allowing the cells to respond to their loading environment (Dupont et al. (2011)). Future experiments over- and under-expressing the protein Yap *in vivo* during postnatal development would help shed light on the role of Hippo signaling during tendon growth. This could be done with genetic mouse models in a tendon-specific inducible manner as well. However, much more work on mechanotransduction via Yap and Hippo signaling during tendon development, homeostasis, and/or injury repair would be necessary before we could begin to study its potential for crosstalk with other pathways in earnest.

6.3 Implications for Human Tendon Evolution

The findings presented in this dissertation are important for improving our understanding of the evolution of the primate Achilles tendon. We have identified a distinct postnatal shift in modes of growth and some potential pathways that may control this transition. Although all experiments were performed in a mouse model, mammalian tendon development is believed to be a relatively conserved process (REF). Assuming mechanisms of mouse postnatal growth are analogous to those of primates, it is plausible that components of the pathways identified here could have served as the molecular targets of selection during the evolution of a longer Achilles tendon in the human lineage.

Given the theory that adaptive genetic mutations underlying morphological change are more likely to be found in non-coding regulatory sequences (King and Wilson (1975);

Carroll (2008); Jones et al. (2012)), the putative *cis*-regulatory regions identified in Chapter 5 can serve as a candidate list of evolutionarily relevant non-coding elements. We can begin to interrogate these putative enhancers for markers of selection and accelerated evolution as a first step toward testing the hypothesis that a long Achilles tendon was selected for during human evolution. To preliminarily demonstrate such an application of these data, we queried the differentially accessible chromatin regions identified in Chapter 5 for loci that overlap with previously published human accelerated regions (HARs) (Pollard et al. (2006); Prabhakar et al. (2006); Bird et al. (2007); Gittelman et al. (2015)) and lineage specific accelerated regions (linARs) (Kostka, Holloway, and Pollard (2018)). This sort of analysis is imperfect and rather superficial, but it can quickly indicate genomic loci of potential evolutionary relevance (Figure 6.1). Compared to the full consensus ATAC-seq peak set, we identified very few accelerated regions (HARs and/or human linARs) that overlap with peaks that are differentially accessible during postnatal growth (Figure 6.1A,B), yielding a tractable set of loci for further investigation. One of the differentially accessible peaks was found to overlap with a human linAR located just downstream of the gene *BHLHE40* (GRCh37/hg19 human genome assembly) (Figure 6.1C). This locus also overlaps with multiple markers of functional regulatory elements and open chromatin in several cell types (E. P. Consortium (2012)) and contains four human specific single nucleotide changes (Figure 6.1C). While this by no means is a conclusive finding of selection on a regulatory locus, it does provide an interesting candidate for future study and provides an example of how the data from Chapter 5 can be applied to evolutionary questions.

An important consideration for researchers interested in the evolution of Achilles tendon size is the potential role of muscle in the development and evolution of this structure. Previous work has shown that primates with longer Achilles tendons tend to have shorter muscle fibers in their gastrocnemius and soleus muscles, and vice versa

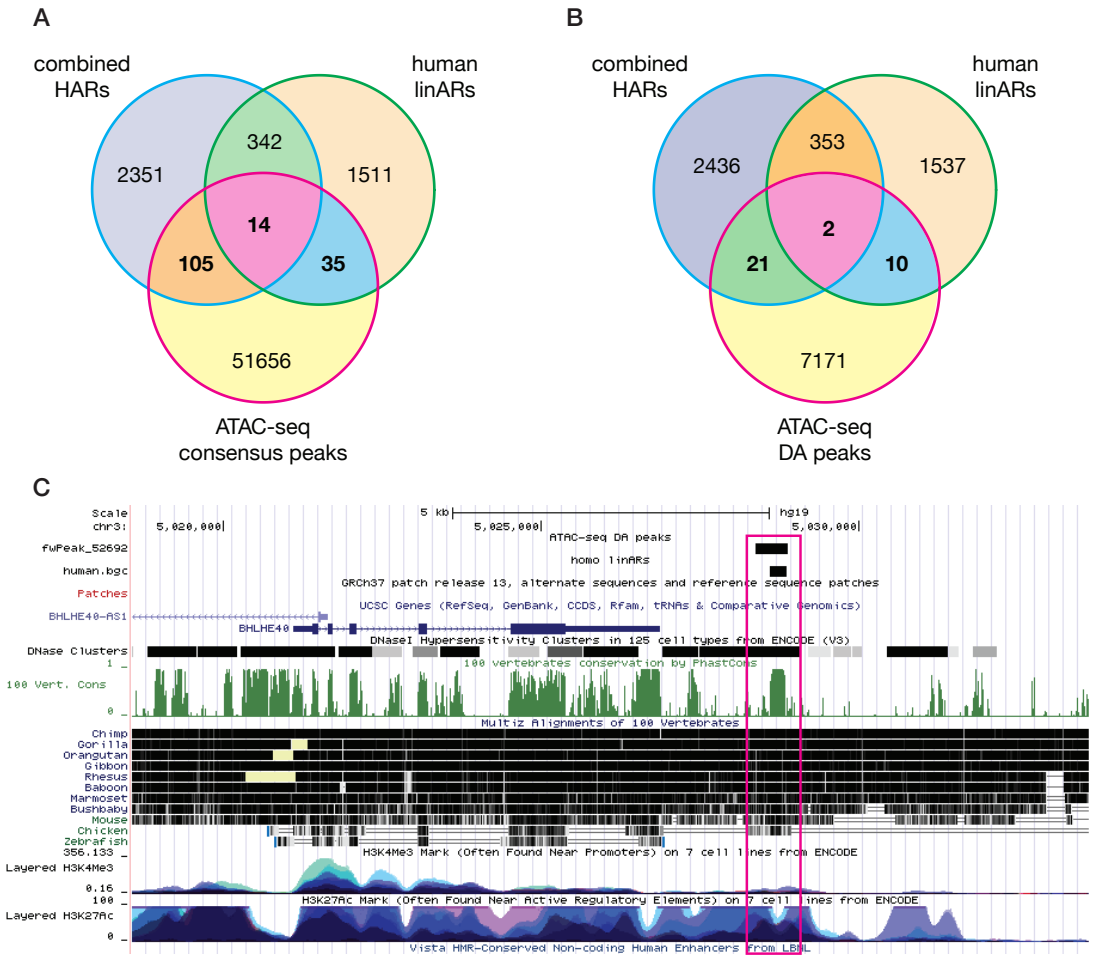


Figure 6.1: Postnatal tendon development ATAC-seq peaks overlap with human accelerated regions (A,B) and an example of a DA ATAC-seq peak that overlaps with one of these accelerated regions (C).

(Rauwerdink (1991)). This suggests that these traits may be linked developmentally and evolutionarily. Currently nothing is known about the relationship between muscle and tendon during postnatal development. It is clear that these two tissues must grow in a coordinated manner to achieve normal adult morphology, but how this occurs and which tissue is the driver of this growth is unknown. This will be an important area of study for both developmental and evolutionary biologists given the important implications for tendon evolution. It is possible that the tendon receives most of its growth cues from the attached muscle in the form of mechanical force, secreted factors, etc. Manipulations to these programs may have pleiotropic effects that influence both muscle and tendon, thus future work may need to investigate regulatory changes to the muscle as well as the tendon during this growth period. Galagos provide an interesting exception to this inverse relationship between tendon and muscle fiber lengths. They exhibit long tendons and average muscle fiber length compared to other primates, which is presumed to be an adaptation for leaping (Rauwerdink (1991)). This indicates that these traits can be at least partially decoupled during development and evolution, and that other animals adapted for leaping may also exhibit similar characteristics. Comparative studies of tendon development on such animals may shed light on the nature of this apparent link between muscle and tendon during development.

If tendon growth is primarily driven by signals from the muscle, it may be that what we see as selection for a long Achilles is actually selection for shorter muscle fibers that, because of the strong developmental link between the tissues, results in both phenotypes. However, it should be noted that studies of early tendon development suggest that tendon patterning and growth are independent of other elements of the musculoskeletal system (Hurle et al. (1990); Brent, Braun, and Tabin (2005)), and that it is actually the tendon and other connective tissue that direct the patterning of muscle (Chevallier and Kieny (1982)). Currently there is simply not enough known

about postnatal tendon growth to know whether this relationship is maintained throughout postnatal development.

This dissertation contributes to the growing literature on molecular mechanisms of tendon development. This work has introduced a new method for isolation of genetic material from postnatal tendon tissue (a previously intractable tissue for RNA analysis), described tendon cell proliferation dynamics during postnatal development, and identified candidate pathways that may control these observed shifts in proliferative potential and growth modes. Finally, these results can help evolutionary biologists formulate more targeted hypotheses about primate tendon evolution and provides a way forward for identifying the molecular targets of selection on the musculoskeletal system.

REFERENCES

Aerts, Peter, D'Août, Kristiaan, Susannah Thorpe, Gilles Berillon, and Evie Vereecke. 2018. "The Gibbon's Achilles Tendon Revisited: Consequences for the Evolution of the Great Apes?" *Proc Royal Soc B* 285 (1880): 20180859. doi:10.1098/rspb.2018.0859.

Akiyama, H., J. E. Kim, K. Nakashima, and others. 2005. "Osteo-Chondroprogenitor Cells Are Derived from Sox9 Expressing Precursors." *Proceedings of the National Academy of Sciences of the USA* 102 (41): 14665–70.

Alberton, Paolo, Sarah Dex, Cvetan Popov, Chisa Shukunami, Matthias Schieker, and Denitsa Docheva. 2015. "Loss of Tenomodulin Results in Reduced Self-Renewal and Augmented Senescence of Tendon Stem/Progenitor Cells." *Stem Cells Dev* 24 (5): 597–609. doi:10.1089/scd.2014.0314.

Alexander, R. 2002. "Tendon Elasticity and Muscle Function." *Comp Biochem Physiology Part Mol Integr Physiology* 133 (4): 1001–11. doi:10.1016/S1095-6433(02)00143-5.

Alexander, R. McNeill. 1991a. "Elastic Mechanisms in Primate Locomotion." *Zeitschrift Für Morphologie Und Anthropologie* 78 (3): 315–20.

———. 1991b. "Energy-Saving Mechanisms in Walking and Running." *J Exp Biology* 160: 55–69.

Alexander, RMcNeill. 1984. "Elastic Energy Stores in Running Vertebrates." *American Zoologist*.

Alexander, RMN, and Bennet-Clark - HC Nature. 1977. "Storage of Elastic Strain Energy in Muscle and Other Tissues." *Nature*. doi:10.1038/265114a0.

Alfieri, Christina, Jonathan Cheek, Santanu Chakraborty, and Katherine Yutzey. 2010. "Wnt Signaling in Heart Valve Development and Osteogenic Gene Induction." *Dev*

Biol 338 (2): 127–35. doi:10.1016/j.ydbio.2009.11.030.

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

Amerongen, Renee van, and Anton Berns. 2006. "Knockout Mouse Models to Study Wnt Signal Transduction." *TRENDS in Genetics* 22 (12). Elsevier: 678–89.

Amezquita, Robert A. 2019. *Marge: API for Homer in R for Genomic Analysis Using Tidy Conventions*.

Analysis, Chimpanzee Consortium. 2005. "Initial Sequence of the Chimpanzee Genome and Com- Parison with the Human Genome." *Nature* 437 (7055): 69–87.

Andersson, D., N. Akrap, D. Svec, T. E. Godfrey, M. Kubista, G. Landberg, and A. Ståhlberg. 2015. "Properties of Targeted Preamplification in Dna and CDNA Quantification." *Expert Review of Molecular Diagnostics* 15: 1085–1100.

Andrews, Simon, Felix Krueger, Anne Segonds-Pichon, Laura Biggins, Christel Krueger, and Steven Wingett. 2012. "FastQC." Babraham, UK: Babraham Institute.

Ansorge, Heather L, Sheila Adams, David E Birk, and Louis J Soslowsky. 2011. "Mechanical, Compositional, and Structural Properties of the Post-Natal Mouse Achilles Tendon" 39 (7): 1904–13. doi:10.1007/s10439-011-0299-0.

Ansorge, Heather, Jason Hsu, Lena Edelstein, Sheila Adams, David Birk, and Louis Soslowsky. 2012. "Recapitulation of the Achilles Tendon Mechanical Properties During Neonatal Development: A Study of Differential Healing During Two Stages of Development in a Mouse Model." *J Orthopaed Res* 30 (3): 448–56. doi:10.1002/jor.21542.

Asakura, Atsushi, Hiroyuki Hirai, Boris Kablar, Shigeru Morita, Jeff Ishibashi, Bryan Piras, Amanda Christ, Mayank Verma, Karin Vineretsky, and Michael Rudnicki.

2007. "Increased Survival of Muscle Stem Cells Lacking the MyoD Gene After Transplantation into Regenerating Skeletal Muscle." *Proc National Acad Sci* 104 (42): 16552–7. doi:10.1073/pnas.0708145104.

Auton, A., and G. McVean. 2012. "Estimating Recombination Rates from Genetic Variation in Humans." *Methods in Molecular Biology* 856: 217–37.

Aziz, A., Q. C. Liu, and F. J. Dilworth. 2010. "Regulating a Master Regulator: Establishing Tissue- Specific Gene Expression in Skeletal Muscle." *Epigenetics* 5: 8.

Banes, AJ, P Weinhold, X Yang, and Tsuzaki - M (1976-2007 .... 1999. "Gap Junctions Regulate Responses of Tendon Cells Ex Vivo to Mechanical Loading." *... Research (1976-2007 ....*

Banes, Albert J, Kathy Donlon, G William Link, Yancy Gillespie, A Griswald Bevin, Hugh D Peterson, Donald Bynum, Susan Watts, and Laurence Dahners. 1988. "Cell Populations of Tendon: A Simplified Method for Isolation of Synovial Cells and Internal Fibroblasts: Confirmation of Origin and Biologic Properties." *Journal of Orthopaedic Research* 6 (1). Wiley Online Library: 83–94.

Barrett, R. D., and H. E. Hoekstra. 2011. "Molecular Spandrels: Tests of Adaptation at the Genetic Level." *Nature Reviews Genetics* 12 (11): 767–80.

Barsby, Tom, and Debbie Guest. 2013. "Transforming Growth Factor Beta3 Promotes Tendon Differentiation of Equine Embryo-Derived Stem Cells." *Tissue Engineering Part A* 19 (19-20). Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA: 2156–65.

Bassat, Elad, Yara Eid Mutlak, Alex Genzelinakh, Ilya Y Shadrin, Kfir Baruch Umansky, Oren Yifa, David Kain, et al. 2017. "The Extracellular Matrix Protein Agrin Promotes Heart Regeneration in Mice." *Nature* 547 (7662). Nature Publishing

Group: 179.

Bataillon, T., J. Duan, C. Hvilsom, and others. n.d. "2015." *Inference of Purifying and Positive Selection in Three Subspecies of Chimpanzees Pan Troglodytes) from Exome Sequencing* 7 (4): 1122–32.

Beason, D. P., A. F. Kuntz, J. E. Hsu, K. S. Miller, and Soslowsky LJ. 2012. "Development and Evaluation of Multiple Tendon Injury Models in the Mouse." *Journal of Biomechanics* 45: 1550–3.

Behringer, R., Vintersten Nagy, G. M., and K. Nagy. 2014. *Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition*. New York, NY: Cold Spring Harbor Press.

Bell, R., J. Li, D. J. Gorski, A. K. Bartels, E. F. Shewman, R. W. Wysocki, B. J. Cole, B. R. Bach Jr, K. Mikecz, and Sandy JD. 2013. "Controlled Treadmill Exercise Eliminates Chondroid Deposits and Restores Tensile Properties in a New Murine Tendinopathy Model." *Journal of Biomechanics* 46: 498–505.

Benito-Sanz, S., M. Aza-Carmona, A. Rodriguez-Estevez, and others. n.d. "2012." *Identification of the First PAR* 20 (1): 125–27.

Beredjiklian, Pedro, Michele Favata, Jeffrey Cartmell, Colleen Flanagan, Timothy Crombleholme, and Louis Soslowsky. 2003. "Regenerative Versus Reparative Healing in Tendon: A Study of Biomechanical and Histological Properties in Fetal Sheep." *Ann Biomed Eng* 31 (10): 1143–52. doi:10.1114/1.1616931.

Berge, Derk, Samantha Brugmann, Jill Helms, and Roel Nusse. 2008. "Wnt and FGF Signals Interact to Coordinate Growth with Cell Fate Specification During Limb Development." *Development* 135 (19): 3247–57. doi:10.1242/dev.023176.

Berge, Derk ten, Samantha A Brugmann, Jill A Helms, and Roel Nusse. 2008. "Wnt and Fgf Signals Interact to Coordinate Growth with Cell Fate Specification During

Limb Development." *Development* 135 (19). The Company of Biologists Ltd: 3247–57.

Bernstein, B. E., M. Kamal, K. Lindblad-Toh, and others. n.d. "2005." *Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse* 120 (2): 169–81.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, and others. n.d. "2006." *A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells* 125 (2): 315–26.

Berthet, Ellora, Carol Chen, Kristin Butcher, Richard A Schneider, Tamara Alliston, and Mohana Amirtharajah. 2013a. "Smad3 Binds Scleraxis and Mohawk and Regulates Tendon Matrix Organization." *Journal of Orthopaedic Research* 31 (9). Wiley Online Library: 1475–83.

Berthet, Ellora, Carol Chen, Kristin Butcher, Richard Schneider, Tamara Alliston, and Mohana Amirtharajah. 2013b. "Smad3 Binds Scleraxis and Mohawk and Regulates Tendon Matrix Organization." *J Orthopaed Res* 31 (9): 1475–83. doi:10.1002/jor.22382.

Bi, W., J. M. Deng, Z. Zhang, R. R. Behringer, and B. de Crombrugghe. 1999. "Sox9 Is Required for Cartilage Formation." *Nature Genetics* 22 (1): 85–89.

Bi, Yanming, Driss Ehirchiou, Tina M Kilts, Colette A Inkson, Mildred C Embree, Wataru Sonoyama, Li Li, et al. 2007. "Identification of Tendon Stem/Progenitor Cells and the Role of the Extracellular Matrix in Their Niche." *Nature Medicine* 13 (10). Nature Publishing Group: 1219.

Biewener, AA. 2008. "Tendons and Ligaments: Structure, Mechanical Behavior and Biological Function." In *Collagen*, 269–84. Springer.

Biewener, AA, and Roberts - TJ Rev. 2000. "Muscle and Tendon Contributions to Force, Work, and Elastic Energy Savings: A Comparative Perspective." *Exerc Sport*

Sci Rev.

Biewener, Andrew, and Sheila Patek. 2018. *Animal Locomotion.* Oxford University Press.

Biology, RMN of. 1991. "Energy-Saving Mechanisms in Walking and Running." *Journal of Experimental Biology.*

Bird, Christine P, Barbara E Stranger, Maureen Liu, Daryl J Thomas, Catherine E Ingle, Claude Beazley, Webb Miller, Matthew E Hurles, and Emmanouil T Dermitzakis. 2007. "Fast-Evolving Noncoding Sequences in the Human Genome." *Genome Biology* 8 (6). BioMed Central: R118.

Birk, David E, and Richard Mayne. 1997. "Localization of Collagen Types I, Iii and V During Tendon Development. Changes in Collagen Types I and Iii Are Correlated with Changes in Fibril Diameter." *European Journal of Cell Biology* 72 (4): 352–61.

Birk, DE, MV Nurminskaya, and EI Zycband. 1995. "Collagen Fibrillogenesis in Situ: Fibril Segments Undergo Post-Depositional Modifications Resulting in Linear and Lateral Growth During Matrix Development." *Dev. Dyn.* 202 (3): 229–43. doi:10.1002/aja.1002020303.

BL., Welch. 1947. "The Generalization of Student's Problem When Several Different Population Variances Are Involved." *Biometrika* 34: 28–35.

Blitz, Einat, Sergey Viukov, Amnon Sharir, Yulia Shwartz, Jenna L Galloway, Brian A Pryce, Randy L Johnson, Clifford J Tabin, Ronen Schweitzer, and Elazar Zelzer. 2009. "Bone Ridge Patterning During Musculoskeletal Assembly Is Mediated Through Scx Regulation of Bmp4 at the Tendon-Skeleton Junction." *Developmental Cell* 17 (6). Elsevier: 861–73.

Bonnin, Marie-Ange, Christine Laclef, Régis Blaise, Eloy-Trinquet, Sophie, Frédéric Relaix, Pascal Maire, and Delphine Duprez. 2005. "Six1 Is Not Involved in Limb Ten-

don Development, but Is Expressed in Limb Connective Tissue Under Shh Regulation." *Mech Develop* 122 (4): 573–85. doi:10.1016/j.mod.2004.11.005.

Bourgon, Richard, Robert Gentleman, and Wolfgang Huber. 2010. "Independent Filtering Increases Detection Power for High-Throughput Experiments." *Proceedings of the National Academy of Sciences* 107 (21). National Acad Sciences: 9546–51.

Bowen, M. E., U. M. Ayturk, K. C. Kurek, W. Yang, and M. L. Warman. 2014. "SHP2 Regulates Chondrocyte Terminal Differentiation, Growth Plate Architecture and Skeletal Cell Fates." *PLoS Genetics* 10: 5.

Bramble, Dennis M, and Daniel E Lieberman. 2004. "Endurance Running and the Evolution of Homo." *Nature* 432 (7015). Nature Publishing Group: 345.

Brandau, Oliver, Alfons Meindl, Reinhard Fässler, and Attila Aszódi. 2001. "A Novel Gene, Tendin, Is Strongly Expressed in Tendons and Ligaments and Shows High Homology with chondromodulin-I." *Dev Dynam* 221 (1): 72–80. doi:10.1002/dvdy.1126.

Brent, Ava E, Thomas Braun, and Clifford J Tabin. 2005. "Genetic Analysis of Interactions Between the Somitic Muscle, Cartilage and Tendon Cell Lineages During Mouse Development." *Development* 132 (3). The Company of Biologists Ltd: 515–28.

Brent, Ava E, Ronen Schweitzer, and Clifford J Tabin. 2003. "A Somitic Compartment of Tendon Progenitors." *Cell* 113 (2). Elsevier: 235–48.

Broman, Karl W., Mark P. Keller, Aimee Teo Broman, Christina Kendziorski, Brian S. Yandell, Śaunak Sen, and Alan D. Attie. 2015. "Identification and Correction of Sample Mix-Ups in Expression Genetic Data: A Case Study." *G3* 5: 2177–86. doi:10.1534/g3.115.019778.

Brown, D, D Wagner, X Li, J Richardson, and E Olson. 1999. "Dual Role of the Basic Helix-Loop-Helix Transcription Factor Scleraxis in Mesoderm Formation and

Chondrogenesis During Mouse Embryogenesis." *Dev Camb Engl* 126 (19): 4317–29.

Buchwalow, I. B., and W. B"ocker. 2014. *Immunohistochemistry: Basics and Methods*. New York, NY: Springer.

Buenrostro, J. D., P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, Dna- Binding Proteins and Nucleosome Position." *Nature Methods* 10 (12): 1213–8.

Buenrostro, Jason D, Beijing Wu, Howard Y Chang, and William J Greenleaf. 2015. *ATAC-seq: a Method for Assaying Chromatin Accessibility Genome-Wide*. Vol. 109. wiley. doi:10.1002/0471142727.mb2129s109.

Buenrostro, Jason, Paul Giresi, Lisa Zaba, Howard Chang, and William Greenleaf. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-binding Proteins and Nucleosome Position." *Nat Methods* 10 (12): 1213–8. doi:10.1038/nmeth.2688.

Campbell, B., C. Agarwal, and J. Wang. 2004. "TGF-B1, TGF-B3, and PGE2 Regulate Contraction of Human Patellar Tendon Fibroblasts." *Biomech Model Mechan* 2 (4): 239–45. doi:10.1007/s10237-004-0041-z.

Campeau, E., and S. Gobeil. 2011. "RNA Interference in Mammals: Behind the Screen." *Briefings in Functional Genomics* 10 (4): 215–26.

Canavez, F., N. T. Young, L. A. Guethlein, and others. n.d. "2001." *Comparison of Chimpanzee and Human Leukocyte Ig-Like Receptor Genes Reveals Framework and Rapidly Evolving Genes* 167 (10): 5786–94.

Canté-Barrett, K, R Pieters, and JPP Meijerink. 2014. "Myocyte Enhancer Factor 2C

in Hematopoiesis and Leukemia." *Oncogene* 33 (4). Nature Publishing Group: 403.

Capellini, T. D., Di Giacomo, and Salsi V. G. 2006. "Pbx1/Pbx2 Requirement for Distal Limb Patterning Is Mediated by the Hierarchical Control of Hox Gene Spatial Distribution and Shh Expression." *Development* 133 (11): 2263–73.

Capellini, T. D., G. Vaccari, and E. Ferretti. 2010. "Scapula Development Is Governed by Genetic Interactions of Pbx1 with Its Family Members and with Emx2 via Their Cooperative Control of Alx1." *Development* 137 (15): 2559–69.

Carlstedt, Carl Anders. 1987. "Mechanical and Chemical Factors in Tendon Healing: Effects of Indomethacin and Surgery in the Rabbit." *Acta Orthopaedica Scandinavica* 58 (sup224). Taylor; Francis: 1–75.

Carroll, S. B. 2008. "Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution." *Cell* 134 (1): 25–36.

Castellazzi, Marc, Giannis Spyrou, Nathalie La Vista, Jean-Pierre Dangy, Fabrice Piu, Moshe Yaniv, and Gilbert Brun. 1991. "Overexpression of c-Jun, JunB, or JunD Affects Cell Growth Differently." *Proceedings of the National Academy of Sciences* 88 (20). National Acad Sciences: 8890–4.

Chakkalakal, Joe V, Josef Christensen, Wanyi Xiang, Mathew T Tierney, Francesca S Boscolo, Alessandra Sacco, and Andrew S Brack. 2014. "Early Forming Label-Retaining Muscle Stem Cells Require P27kip1 for Maintenance of the Primitive State." *Development* 141 (8). Oxford University Press for The Company of Biologists Limited: 1649–59.

Chan, Kai-Ming, Sai-Chuen Fu, Yim-Ping Wong, Wun-Chun Hui, Yau-Chuk Cheuk, and Margaret Wan-Nar Wong. 2008. "Expression of Transforming Growth Factor B Isoforms and Their Roles in Tendon Healing." *Wound Repair and Regeneration* 16 (3).

Wiley Online Library: 399–407.

Chan, Y. F., M. E. Marks, and F. C. Jones. 2010. "Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer." *Science* 327 (5963): 302–5.

Chau, M., J. C. Lui, E. B. Landman, and others. 2014. "Gene Expression Profiling Reveals Similarities Between the Spatial Architectures of Postnatal Articular and Growth Plate Cartilage." *PLoS ONE* 9: 7. doi:e103061.

Chen, Y., Y. Zheng, Y. Kang, and others. 2015. "Functional Disruption of the Dystrophin Gene in Rhesus Monkey Using Crispr/Cas9." *Human Molecular Genetics* 24 (13): 3764–74.

Cheng, Y., Z. Ma, B. H. Kim, and others. 2014. "Principles of Regulatory Information Conservation Between Mouse and Human." *Nature* 515 (7527): 371–75.

Chevallier, Alain, and Madeleine Kieny. 1982. "On the Role of the Connective Tissue in the Patterning of the Chick Limb Musculature." *Wilhelm Roux's Archives of Developmental Biology* 191 (4). Springer: 277–80.

Chiang, C., Y. Litingtung, E. Lee, and others. 1996. "Cyclopia and Defective Axial Patterning in Mice Lacking Sonic Hedgehog Gene Function." *Nature* 383 (6599): 407–13.

Christian, M., T. Cermak, E. L. Doyle, and others. 2010. "Targeting Dna Double-Strand Breaks with Tal Effector Nucleases." *Genetics* 186 (2): 757–61.

Christians, J. K., D. R. de Zwaan, and S. H. Fung. 2013. "Pregnancy Associated Plasma Protein A2 Papp-A2) Affects Bone Size and Shape and Contributes to Natural Variation in Postnatal Growth in Mice." *PLoS ONE* 8: 2. doi:e56260.

Church, V., T. Nohno, C. Linker, C. Marcelle, and P. Francis-West. 2002. "Wnt

Regulation of Chondrocyte Differentiation." *Journal of Cell Science* 115 (24): 4809–18.

CM, McNeilly, AJ Banes, and Benjamin - M of anatomy. 1996. "Tendon Cells in Vivo Form a Three Dimensional Network of Cell Processes Linked by Gap Junctions." *Journal of Anatomy*.

Codina-Sola, M., B. Rodriguez-Santiago, and A. Homs. 2015. "Integrated Analysis of Whole-Exome Sequencing and Transcriptome Profiling in Males with Autism Spectrum Disorders." *Molecular Autism* 6: 21.

Colosimo, P. F., K. E. Hosemann, S. G. Balabhadra, and others. 2005. "Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles." *Science* 307 (5717): 1928–33.

Connizzo, Brianne, Sarah Yannascoli, and Louis Soslowsky. 2013. "Structure–function Relationships of Postnatal Tendon Development: A Parallel to Healing." *Matrix Biol* 32 (2): 106–16. doi:10.1016/j.matbio.2013.01.007.

Consortium, 1000 Genomes Project. 2012. "An Integrated Map of Genetic Variation from 1,092 Human Genomes." *Nature* 491 (7422): 56–65.

Consortium, Encode Project. 2012. "An Integrated Encyclopedia of Dna Elements in the Human Genome." *Nature* 489 (7414): 57–74.

Consortium, Roadmap Epigenomics Project. 2015. "Integrative Analysis of 111 Reference Human Epigenomes." *Nature* 518 (7539): 317–30.

Cooper, K. L., S. Oh, Y. Sung, and others. 2013. "Multiple Phases of Chondrocyte Enlargement Underlie Differences in Skeletal Proportions." *Nature* 495 (7441): 375–78.

Corces, M Ryan, Alexandro E Trevino, Emily G Hamilton, Peyton G Greenside, Nicholas A Sinnott-Armstrong, Sam Vesuna, Ansuman T Satpathy, et al. 2017. "An Improved Atac-Seq Protocol Reduces Background and Enables Interrogation of Frozen

Tissues." *Nature Methods* 14 (10). Nature Publishing Group: 959.

Cotney, J., J. Leng, J. Yin, and others. 2013. "The Evolution of Lineage-Specific Regulatory Activities in the Human Embryonic Limb." *Cell* 154 (1): 185–96.

Crawford, G. E., I. E. Holt, J. C. Mullikin, and others. 2004. "Identifying Gene Regulatory Elements by Genome-Wide Recovery of Dnase Hypersensitive Sites." *Proceedings of the National Academy of Sciences of the USA* 101 (4): 992–97.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, and others. 2010. "Histone H3k27ac Separates Active from Poised Enhancers and Predicts Developmental State." *Proceedings of the National Academy of Sciences of the USA* 107 (50): 21931–6.

Cserjesi, P, D Brown, K Ligon, G Lyons, N Copeland, D Gilbert, N Jenkins, and E Olson. 1995. "Scleraxis: A Basic Helix-Loop-Helix Protein That Prefigures Skeletal Formation During Mouse Embryogenesis." *Dev Camb Engl* 121 (4): 1099–1110.

Dao, Debbie Y, XUE Yang, Chen Di, Michael Zuscik, and REGIS J O'KEEFE. 2007. "Axin1 and Axin2 Are Regulated by Tgf-B and Mediate Cross-Talk Between Tgf-B and Wnt Signaling Pathways." *Annals of the New York Academy of Sciences* 1116. NIH Public Access: 82.

Datta, S., L. Malhotra, R. Dickerson, and others. 2015. "Laser Capture Microdissection: Big Data from Small Samples." *Histology and Histopathology* 30 (11): 1255–69.

Davidson, E. 2001. *Genomic Regulatory Systems: Development and Evolution*. San Diego, CA: Academic Press.

Davis, Carrie A, Benjamin C Hitz, Cricket A Sloan, Esther T Chan, Jean M Davidson, Idan Gabdank, Jason A Hilton, et al. 2017. "The Encyclopedia of Dna Elements (Encode): Data Portal Update." *Nucleic Acids Research* 46 (D1). Oxford University

Press: D794–D801.

Decker, R. S., E. Koyama, M. Enomoto-Iwamoto, and others. 2014. “Mouse Limb Skeletal Growth and Synovial Joint Development Are Coordinately Enhanced by Kartogenin.” *Developmental Biology* 395 (2): 255–67.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. “Capturing Chromosome Conformation.” *Science* 295 (5558): 1306–11.

Diogo, Rui. 2011. *Photographic and Descriptive Musculoskeletal Atlas of Gorilla : With Notes on the Attachments, Variations, Innervation, Synonymy, and Weight of the Muscles*. Enfield, N.H. : Boca Raton, FL: Science Publishers ; Distributed by CRC Press.

———. 2013. *Photographic and Descriptive Musculoskeletal Atlas of Chimpanzees : With Notes on the Attachments, Variations, Innervation, Function and Synonymy and Weight of the Muscles*. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Diogo, Rui, Josep M. Potau, Juan F. Pastor, Felix J. de Paz, Mercedes Barbosa, Eva M. Ferrero, Gaelle Bello, Mohammed Ashraf Aziz, Julia Arias-Martorell, and Bernard Wood. 2013. *Photographic and Descriptive Musculoskeletal Atlas of Orangutans*. 1st ed. CRC Press.

DiRenzo, Daniel M, Mirnal A Chaudhary, Xudong Shi, Sarah R Franco, Joshua Zent, Katie Wang, Lian-Wang Guo, and K Craig Kent. 2016. “A Crosstalk Between Tgf-B/Smad3 and Wnt/B-Catenin Pathways Promotes Vascular Smooth Muscle Cell Proliferation.” *Cellular Signalling* 28 (5). Elsevier: 498–505.

Dong, Z., and Y. Chen. 2013. “Transcriptomics: Advances and Approaches.” *Science in China Series C Life Sciences* 56 (10): 960–67.

Doral, MN, M Alam, M Bozkurt, Turhan - E Surgery, and Sports. 2010. “Functional

Anatomy of the Achilles Tendon." *Knee Surgery.*

Dostie, J., T. A. Richmond, R. A. Arnaout, and others. 2006. "Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions Between Genomic Elements." *Genome Research* 16 (10): 1299–1309.

Driscoll, Tristan P, Brian D Cosgrove, Su-Jin Heo, Zach E Shurden, and Robert L Mauck. 2015. "Cytoskeletal to Nuclear Strain Transfer Regulates Yap Signaling in Mesenchymal Stem Cells." *Biophysical Journal* 108 (12). Elsevier: 2783–93.

Dudley, A. T., M. A. Ros, and C. J. Tabin. 2002. "A Re-Examination of Proximodistal Patterning During Vertebrate Limb Development." *Nature* 418 (6897): 539–44.

Dunkman, A. A., M. R. Buckley, M. J. Mienaltowski, S. M. Adams, S. J. Thomas, A. Kumar, D. P. Beason, R. V. Iozzo, D. E. Birk, and Soslowsky LJ. 2014. "The Injury Response of Aged Tendons in the Absence of Biglycan and Decorin." *Matrix Biology* 35: 232–38.

Dunkman, Andrew, Mark Buckley, Michael Mienaltowski, Sheila Adams, Stephen Thomas, Lauren Satchell, Akash Kumar, et al. 2013. "Decorin Expression Is Important for Age-Related Changes in Tendon Structure and Mechanical Properties." *Matrix Biol* 32 (1): 3–13. doi:10.1016/j.matbio.2012.11.005.

Dupont, Sirio, Leonardo Morsut, Mariaceleste Aragona, Elena Enzo, Stefano Giulitti, Michelangelo Cordenonsi, Francesca Zanconato, et al. 2011. "Role of Yap/Taz in Mechanotransduction." *Nature* 474 (7350). Nature Publishing Group: 179.

Dyment, Nathaniel A, Yusuke Hagiwara, Brya G Matthews, Yingcui Li, Ivo Kalajzic, and David W Rowe. 2014. "Lineage Tracing of Resident Tendon Progenitor Cells During Growth and Natural Healing." *PloS One* 9 (4). Public Library of Science: e96113.

Dyment, Nathaniel A, Chia-Feng Liu, Namdar Kazemi, Lindsey E Aschbacher-Smith,

Keith Kenter, Andrew P Breidenbach, Jason T Shearn, Christopher Wylie, David W Rowe, and David L Butler. 2013. "The Paratenon Contributes to Scleraxis-Expressing Cells During Patellar Tendon Healing." *PloS One* 8 (3). Public Library of Science: e59944.

Egerbacher, Monika, Steven P Arnoczky, Oscar Caballero, Michael Lavagnino, and Keri L Gardner. 2008. "Loss of Homeostatic Tension Induces Apoptosis in Tendon Cells: An in Vitro Study." *Clinical Orthopaedics and Related Research* 466 (7). Springer: 1562.

Ezura, Yoichi, Shukti Chakravarti, Ake Oldberg, Inna Chervoneva, and David Birk. 2000. "Differential Expression of Lumican and Fibromodulin Regulate Collagen Fibrillogenesis in Developing Mouse Tendons." *J Cell Biology* 151 (4): 779–88. doi:10.1083/jcb.151.4.779.

Farber, C. R., S. A. Kelly, E. Baruch, and others. 2011. "Identification of Quantitative Trait Loci Influencing Skeletal Architecture in Mice: Emergence of Cdh11 as a Primary Candidate Gene Regulating Femoral Morphology." *Journal of Bone and Mineral Research* 26 (9): 2174–83.

Farooq, M., H. Nakai, A. Fujimoto, and others. 2013. "Characterization of a Novel Missense Mutation in the Prodomain of Gdf5, Which Underlies Brachydactyly Type c and Mild Grebe Type Chondrodysplasia in a Large Pakistani Family." *Human Genetics* 132 (11): 1253–64.

Favata, Michele, Pedro Beredjiklian, Miltiadis Zgonis, David Beason, Timothy Crombleholme, Abbas Jawad, and Louis Soslowsky. 2006. "Regenerative Properties of Fetal Sheep Tendon Are Not Adversely Affected by Transplantation into an Adult Environment." *J Orthopaed Res* 24 (11): 2124–32. doi:10.1002/jor.20271.

Florio, M., M. Albert, E. Taverna, and others. 2015. "Human-Specific Gene Arhgap11b

Promotes Basal Progenitor Amplification and Neocortex Expansion." *Science* 347 (6229): 1465–70.

Fossat, N., C. K. Ip, V. J. Jones, and others. 2015. "Context-Specific Function of the Lim Homeobox1 Transcription Factor in Head Formation of the Mouse Embryo." *Development* 142 (11): 2069–79.

Foudi, Adlen, Konrad Hochedlinger, Denille Van Buren, Jeffrey W Schindler, Rudolf Jaenisch, Vincent Carey, and Hanno Hock. 2009. "Analysis of Histone 2B-Gfp Retention Reveals Slowly Cycling Hematopoietic Stem Cells." *Nature Biotechnology* 27 (1). Nature Publishing Group: 84.

Francis-West, P. H., M. K. Richardson, E. Bell, and others. 1996. "The Effect of Overexpression of Bmps and Gdf-on the Development of Chick Limb Skeletal Elements." *Annals If the New York Academy of Science* 5 (785): 254–55.

Fukami, M., T. Tsuchiya, S. Takada, and others. 2012. "Complex Genomic Rearrangement in the Sox' Region in a Patient with Pierre Robin Sequence and Hypoplastic Left Scapula." *American Journal of Medical Genetics A* 9 (5): 1529–34.

Fullwood, M. J., Y. Han, C. L. Wei, X. Ruan, and Y. Ruan. 2010. "Chromatin Interaction Analysis Using Paired-End Tag Sequencing." *Current Protocol in Molecular Biology, Chapter* 21: 15–21.

Furey, T. S. 2012. "ChIP-Seq and Beyond: New and Improved Methodologies to Detect and Char- Acterize Protein–DNA Interactions." *Nature Reviews Genetics* 13 (12): 840–52.

Furniss, D., L. A. Lettice, I. B. Taylor, and others. 2008. "A Variant in the Sonic Hedgehog Regulatory Sequence Zrs) Is Associated with Triphalangeal Thumb and Deregulates Expression in the Developing Limb." *Human Molecular Genetics* 17 (16):

2417–23.

Garnier, Simon. 2018. *Viridis: Default Color Maps from 'Matplotlib'*. https://CRAN.R-project.org/package=viridis.

Gaspar, John M. 2018. "NGmerge: Merging Paired-End Reads via Novel Empirically-Derived Models of Sequencing Errors." *BMC Bioinformatics* 19 (1). BioMed Central: 536.

Gehrke, A. R., I. Schneider, E. de la Calle-Mustienes, and others. 2015. "Deep Conservation of Wrist and Digit Enhancers in Fish." *Proceedings of the National Academy of Sciences of the USA* 112 (3): 803–8.

Gehrke, Andrew R, Emily Neverett, Yi-Jyun Luo, Alexander Brandt, Lorenzo Ricci, Ryan E Hulett, Annika Gompers, et al. 2019. "Acoel Genome Reveals the Regulatory Landscape of Whole-Body Regeneration" 363 (6432): eaau6173. doi:10.1126/science.aau6173.

Gennequin, B., D. M. Otte, and A. Zimmer. 2013. "CRISPR/Cas-Induced Double-Strand Breaks Boost the Frequency of Gene Replacements for Humanizing the Mouse Cnr2 Gene." *Biochemical and Biophysical Research Communications* 441 (4): 815–19.

Gibson, S. V., and G. Muse. 2009. *A Primer in Genome Sciences*. New York, NY: Sinauer Associates.

Gilad, Y., S. A. Rifkin, and J. K. Pritchard. 2008. "Revealing the Architecture of Gene Regulation: The Promise of EQTL Studies." *Trends in Genetics* 24 (8): 408–15.

Gilbert, S. 2013. *Developmental Biology*. New York, NY: Sinauer Associates.

Giresi, P. G., J. Kim, R. M. McDaniell, V. R. Iyer, and J. D. Lieb. 2007. "FAIRE Formaldehyde- Assisted Isolation of Regulatory Elements) Isolates Active Regulatory

Elements from Human Chromatin." *Genome Research* 17 (6): 877–85.

Gittelman, Rachel M, Enna Hun, Ferhat Ay, Jennifer Madeoy, Len Pennacchio, William S Noble, R David Hawkins, and Joshua M Akey. 2015. "Comprehensive Identification and Analysis of Human Accelerated Regulatory Dna." *Genome Research* 25 (9). Cold Spring Harbor Lab: 1245–55.

Gomoll, Andreas, Jeffrey Katz, Jon Warner, and Peter Millett. 2004. "Rotator Cuff Disorders: Recognition and Management Among Patients with Shoulder Pain." *Arthritis Rheumatism* 50 (12): 3751–61. doi:10.1002/art.20668.

Gordon, C. T., F. A. Rodda, and P. G. Farlie. 2009. "The Rcas Retroviral Expression System in the Study of Skeletal Development." *Developmental Dynamics* 238 (4): 797–811.

Gordon, C. T., T. Y. Tan, S. Benko, and others. 2009. "Long-Range Regulation at the Sox9 Locus in Development and Disease." *Journal of Medical Genetics* 46 (10): 649–56.

Green, Brian, Christiane Bouchier, Cécile Fairhead, Nancy Craig, and Brendan Cormack. 2012. "Insertion Site Preference of Mu, Tn5, and Tn7 Transposons." *Mobile Dna-Uk* 3 (1): 3. doi:10.1186/1759-8753-3-3.

Grigoryan, Tamara, Peter Wend, Alexandra Klaus, and Walter Birchmeier. 2008. "Deciphering the Function of Canonical Wnt Signals in Development and Disease: Conditional Loss-and Gain-of-Function Mutations of B-Catenin in Mice." *Genes and Development* 22 (17). Cold Spring Harbor Lab: 2308–41.

Grinstein, Mor, Heather L Dingwall, Rishita R Shah, Terence D Capellini, and Jenna L Galloway. 2018. "A Robust Method for Rna Extraction and Purification from a Single Adult Mouse Tendon." *PeerJ* 6. PeerJ Inc.: e4664.

Gross, D. S., and W. T. Garrard. 1988. "Nuclease Hypersensitive Sites in Chromatin."

Annual Review of Biochemistry 57: 159–97.

Grossman, S. R., K. G. Andersen, I. Shlyakhter, and others. 2013. "Identifying Recent Adaptations in Large-Scale Genomic Data." *Cell* 152 (4): 703–13.

Gu, Zuguang, Roland Eils, and Matthias Schlesner. 2016. "Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data." *Bioinformatics*.

Guenther, C., L. Pantalena-Filho, and D. M. Kingsley. 2008. "Shaping Skeletal Growth by Modular Regulatory Elements in the Bmp5 Gene." *PLoS Genetics* 4: 12. doi:e1000308.

Guo, Xing, and Xiao-Fan Wang. 2009. "Signaling Cross-Talk Between TGF-B/BMP and Other Pathways." *Cell Res* 19 (1): cr2008302. doi:10.1038/cr.2008.302.

Gupta, R. M., and K. Musunuru. 2014. "Expanding the Genetic Editing Tool Kit: ZFNs, Talens, and Crispr-Cas9." *Journal of Clinical Investigations* 124 (10): 4154–61.

Gurnett, C. A., A. M. Bowcock, F. R. Dietz, and others. 2007. "Two Novel Point Mutations in the Long- Range Shh Enhancer in Three Families with Triphalangeal Thumb and Preaxial Polydactyly." *American Journal of Medical Genetics A* 143: 27–32.

Harfe, B. D., P. J. Scherz, S. Nissim, and others. 2004. "Evidence for an Expansion-Based Temporal Shh Gradient in Specifying Vertebrate Digit Identities." *Cell* 118 (4): 517–28.

Hartl, D., and M. Ruvolo. 2011. *Genetics: Analysis of Genes and Genomes*. New York, NY: Jones & Bartlett Learning.

Hartmann, C., and C. J. Tabin. 2000. "Dual Roles of Wnt Signaling During Chondrogenesis in the Chicken Limb." *Development* 127 (14): 3141–59.

Hartmann, Christine, and Clifford Tabin. 2001. "Wnt-14 Plays a Pivotal Role in

Inducing Synovial Joint Formation in the Developing Appendicular Skeleton." *Cell* 104 (3): 341–51. doi:10.1016/S0092-8674(01)00222-7.

Hauptmann, G. 2015. *In Situ Hybridization Methods: Neuromethods*. New York, NY: Humana Press.

Havill, L. M., M. C. Mahaney, L. A. Cox, and others. 2005. "A Quantitative Trait Locus for Normal Variation in Forearm Bone Mineral Density in Pedigreed Baboons Maps to the Ortholog of Human Chromosome." *Journal of Clinical Endocrinology and Metabolism* 90 (6): 3638–45.

Havis, Emmanuelle, Marie-Ange Bonnin, Olivera-Martinez, Isabel, Nicolas Nazaret, Mathilde Ruggiu, Jennifer Weibel, Charles Durand, et al. 2014. "Transcriptomic Analysis of Mouse Limb Tendon Cells During Development." *Development* 141 (19): 3683–96. doi:10.1242/dev.108654.

Havis, Emmanuelle, Marie-Ange Bonnin, Joana Lima, Benjamin Charvet, Cécile Milet, and Delphine Duprez. 2016. "TGFB and FGF Promote Tendon Progenitor Fate and Act Downstream of Muscle Contraction to Regulate Tendon Differentiation During Chick Limb Development." *Development* 143 (20): 3839–51. doi:10.1242/dev.136242.

Hedbom, Erik, and D Heinegaard. 1989. "Interaction of a 59-KDa Connective Tissue Matrix Protein with Collagen I and Collagen Ii." *Journal of Biological Chemistry* 264 (12). ASBMB: 6898–6905.

Heinemeier, Katja Maria, Peter Schjerling, Tommy F Ohlenschlaeger, Christian Eismark, Jesper Olsen, and Michael Kjaer. 2018. "Carbon-14 Bomb Pulse Dating Shows That Tendinopathy Is Preceded by Years of Abnormally High Collagen Turnover." *The FASEB Journal* 32 (9). Federation of American Societies for Experimental Biology Bethesda, MD, USA: 4763–75.

Heinemeier, Katja, Peter Schjerling, Jan Heinemeier, Stig Magnusson, and Michael

Kjaer. 2013. “Lack of Tissue Renewal in Human Adult Achilles Tendon Is Revealed by Nuclear Bomb 14C.” *Faseb J* 27 (5): 2074–9. doi:10.1096/fj.12-225599.

Heinemeier, KM, and Kjaer - MJ Interact. 2011. “In Vivo Investigation of Tendon Responses to Mechanical Loading.” *J Musculoskelet Neuronal Interact*.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, and others. 2009. “Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression.” *Nature* 459 (7243): 108–12.

Hernando-Herraez, I., J. Prado-Martinez, P. Garg, and others. 2013. “Dynamics of Dna Methylation in Recent Human and Great Ape Evolution.” *PLoS Genetics* 9: 9. doi:e1003763.

Hess, Gregory P, William L Cappiello, Robert M Poole, and Stephen C Hunter. 1989. “Prevention and Treatment of Overuse Tendon Injuries.” *Sports Medicine* 8 (6). Springer: 371–84.

Hiller, M., B. T. Schaar, V. B. Indjeian, and others. 2012. “A Forward Genomics Approach Links Geno- Type to Phenotype Using Independent Phenotypic Losses Among Related Species.” *Cell Reports* 2 (4): 817–23.

Hoffmann, Andrea, Gadi Pelled, Gadi Turgeman, Peter Eberle, Yoram Zilberman, Hadassah Shinar, Keren Keinan-Adamsky, et al. 2006. “Neotendon Formation Induced by Manipulation of the Smad8 Signalling Pathway in Mesenchymal Stem Cells.” *The Journal of Clinical Investigation* 116 (4). Am Soc Clin Investig: 940–52.

Horvat-Gordon, M., C. A. Praul, R. Ramachandran, P. A. Bartell, and R. M. Leach Jr. 2010. “Use of Microarray Analysis to Study Gene Expression in the Avian Epiphyseal Growth Plate.” *Comparative Biochemistry and Physiology Part D Genomics and Proteomics* 5 (1): 12–23.

Howell, Kristen, Chun Chien, Rebecca Bell, Damien Laudier, Sara Tufa, Douglas

Keene, Andarawis-Puri, Nelly, and Alice Huang. 2017. "Novel Model of Tendon Regeneration Reveals Distinct Cell Mechanisms Underlying Regenerative and Fibrotic Tendon Healing." *Sci Rep-Uk* 7 (1): 45238. doi:10.1038/srep45238.

Huang, Alice H, Timothy J Riordan, Brian Pryce, Jennifer L Weibel, Spencer S Watson, Fanxin Long, Veronique Lefebvre, et al. 2015. "Musculoskeletal Integration at the Wrist Underlies the Modular Development of Limb Tendons." *Development* 142 (14). Oxford University Press for The Company of Biologists Limited: 2431–41.

Huang, J., J. Chen, J. Esparza, and others. 2015. "EQTL Mapping Identifies Insertion- and Deletion-Specific EQTLs in Multiple Tissues." *Nature Communications* 6: 6821.

Huang, R., B. Christ, and K. Patel. 2006. "Regulation of Scapula Development." *Anatomy and Embryology* 211: 65–71.

Hurle, JM, MA Ros, Y Ganan, D Macias, M Critchlow, and JR Hinchliffe. 1990. "Experimental Analysis of the Role of Ecm in the Patterning of the Distal Tendons of the Developing Limb Bud." *Cell Differentiation and Development* 30 (2). Elsevier: 97–108.

Ignotz, RA, and J Massagué. 1986. "Transforming Growth Factor-Beta Stimulates the Expression of Fibronectin and Collagen and Their Incorporation into the Extracellular Matrix." *J. Biol. Chem.* 261 (9): 4337–45.

Infante, C. R., S. Park, A. G. Mihala, D. M. Kingsley, and D. B. Menke. 2013. "Pitx1 Broadly Associates with Limb Enhancers and Is Enriched on Hindlimb Cis-Regulatory Elements." *Developmental Biology* 374 (1): 234–44.

Inoue, K., and Y. Imai. 2014. "Identification of Novel Transcription Factors in Osteoclast Differentiation Using Genome-Wide Analysis of Open Chromatin Determined by Dnase-Seq." *Journal of Bone and Mineral Research* 29 (8): 1823–32.

Ireland, M. L., and Ott SM. 2000. "The Effects of Pregnancy on the Musculoskeletal

System." *Clinical Orthopaedics and Related Research* 372: 169–79.

Ito, Yoshiaki, Naoya Toriuchi, Teruhito Yoshitaka, Hiroe Ueno-Kudoh, Tempei Sato, Shigetoshi Yokoyama, Keiichiro Nishida, et al. 2010. "The Mohawk Homeobox Gene Is a Critical Regulator of Tendon Differentiation." *Proceedings of the National Academy of Sciences* 107 (23). National Acad Sciences: 10538–42.

Jaffe, A. E., R. Tao, A. L. Norris, M. Kealhofer, A. Nellore, J. H. Shin, D. Kim, Y. Jia, T. M. Hyde, and Kleinman JE. 2017. "QSVA Framework for Rna Quality Correction in Differential Expression Analysis." *Proceedings of the National Academy of Sciences:* 201617384.

James, C. G., L. A. Stanton, H. Agoston, and others. 2010. "Genome-Wide Analyses of Gene Expression During Mouse Endochondral Ossification." *PLoS ONE* 5: 1. doi:e8693.

Jang, J. S., V. A. Simon, R. M. Feddersen, F. Rakhshan, D. A. Schultz, M. A. Zschunke, W. L. Lingle, C. P. Kolbert, and J. Jen. 2011. "Quantitative MiRNA Expression Analysis Using Fluidigm Microfluidics Dynamic Arrays." *BMC Genomics* 12: 144.

Jasinska, A. J., M. K. Lin, S. Service, and others. 2012. "A Non-Human Primate System for Large-Scale Genetic Studies of Complex Traits." *Human Molecular Genetics* 21 (15): 3307–16.

Jelinsky, Scott A, Joanne Archambault, Li Li, and Howard Seeherman. 2010. "Tendon-Selective Genes Identified from Rat and Human Musculoskeletal Tissues." *Journal of Orthopaedic Research* 28 (3). Wiley Online Library: 289–97.

Jens, Bojsen-Møller, Philip Hansen, Per Aagaard, Ulla Svantesson, Michael Kjaer, and S Magnusson. 2004. "Differential Displacement of the Human Soleus and Medial Gastrocnemius Aponeuroses During Isometric Plantar Flexor Contractions in Vivo."

J Appl Physiol 97 (5): 1908–14. doi:10.1152/japplphysiol.00084.2004.

Jian, Hongyan, Xing Shen, Irwin Liu, Mikhail Semenov, XI He, and Xiao-Fan Wang. 2006. "Smad3-Dependent Nuclear Translocation of B-Catenin Is Required for Tgf-B1-Induced Proliferation of Bone Marrow-Derived Adult Human Mesenchymal Stem Cells." *Genes and Development* 20 (6). Cold Spring Harbor Lab: 666–74.

Jinek, M., K. Chylinski, I. Fonfara, and others. 2012. "A Programmable Dual-Rna-Guided Dna Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21.

Jones, F. C., M. G. Grabherr, Y. F. Chan, and others. 2012. "The Genomic Basis of Adaptive Evolution in Threespine Sticklebacks." *Nature* 484 (7392): 55–61.

Józsa, László, József B Bálint, and Susanne Demel. 1978. "Histochemical and Ultrastructural Study of Human Muscles After Spontaneous Rupture of the Tendon." *Acta Histochemica* 63 (1). Elsevier: 61–73.

Jun, Goo, Mary Kate Wing, Gonçalo R Abecasis, and Hyun Min Kang. 2015. "An Efficient and Scalable Analysis Framework for Variant Extraction and Refinement from Population-Scale Dna Sequence Data." *Genome Research* 25 (6). Cold Spring Harbor Lab: 918–25.

Kalson, Nicholas, Yinhui Lu, Susan Taylor, Tobias Starborg, David Holmes, and Karl Kadler. 2015. "A Structure-Based Extracellular Matrix Expansion Mechanism of Fibrous Tissue Growth." *Elife* 4: e05958. doi:10.7554/eLife.05958.

Kamberov, Y. G., S. Wang, J. Tan, and others. 2013. "Modeling Recent Human Evolution in Mice by Expression of a Selected Edar Variant." *Cell* 152 (4): 691–702.

Kan, Akinori, and Clifford J Tabin. 2013. "C-Jun Is Required for the Specification of

Joint Cell Fates." *Genes & Development* 27 (5). Cold Spring Harbor Lab: 514–24.

Kannus, P. 2000. "Structure of the Tendon Connective Tissue." *Scand J Med Sci Spor* 10 (6): 312–20. doi:10.1034/j.1600-0838.2000.010006312.x.

Kapacee, Zoher, Ching-Yan Chloé Yeung, Yinhui Lu, David Crabtree, David F Holmes, and Karl E Kadler. 2010. "Synthesis of Embryonic Tendon-Like Tissue by Human Marrow Stromal/Mesenchymal Stem Cells Requires a Three-Dimensional Environment and Transforming Growth Factor B3." *Matrix Biology* 29 (8). Elsevier: 668–77.

Kaufman, and PJ Rousseeuw. 1990. "Partitioning Around Medoids (Program PAM)." In, 65–128. John Wiley; Sons, Inc.

Kaux J.F., F. B., C. L. Goff, Crielaard J-m, and Croisier J.L. 2011. "Current Opinions on Tendinopathy." *Journal of Sports Science & Medicine* 10 (2): 238–53.

Kendziorski, C., R. Irizarry, Chen K-s, J. Haag, and M. Gould. 2005. "On the Utility of Pooling Biological Samples in Microarray Experiments." *Proceedings of the National Academy of Sciences of the United States of America* 102: 4252–7.

Kenney-Hunt, J. P., T. T. Vaughn, L. S. Pletscher, and others. 2006. "Quantitative Trait Loci for Body Size Components in Mice." *Mammalian Genome* 17 (6): 526–37.

Khaitovich, P., I. Hellmann, W. Enard, and others. 2005. "Parallel Patterns of Evolution in the Genomes and Transcriptomes of Humans and Chimpanzees." *Science* 309 (5742): 1850–4.

Khaitovich, P., B. Muetzel, X. She, and others. 2004. "Regional Patterns of Gene Expression in Human and Chimpanzee Brains." *Genome Research* 14 (8): 1462–73.

Khan, Z., M. J. Ford, D. A. Cusanovich, and others. 2013. "Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures." *Science*

342 (6162): 1100–1104.

Khayyeri, Hanifeh, Parmis Blomgran, Malin Hammerman, Mikael Turunen, Annika Löwgren, Guizar-Sicairos, Manuel, Per Aspenberg, and Hanna Isaksson. 2017. "Achilles Tendon Compositional and Structural Properties Are Altered After Unloading by Botox." *Sci Rep-Uk* 7 (1): 13067. doi:10.1038/s41598-017-13107-7.

Kim, Minseong, and Eek-hoon Jho. 2014a. "Cross-Talk Between Wnt/B-Catenin and Hippo Signaling Pathways: A Brief Review." *BMB Reports* 47 (10). Korean Society for Biochemistry; Molecular Biology: 540.

———. 2014b. "Cross-Talk Between Wnt/B-Catenin and Hippo Signaling Pathways: A Brief Review." *Bmb Rep* 47 (10): 540–45. doi:10.5483/BMBRep.2014.47.10.177.

King, M. C., and A. C. Wilson. 1975. "Evolution at Two Levels in Humans and Chimpanzees." *Science* 188 (4184): 107–16.

Kitchen, R. R., M. Kubista, and A. Tichopad. 2010. "Statistical Aspects of Quantitative Real-Time Pcr Experiment Design." *Methods* 50: 231–36.

Kjær, M., H. Langberg, K. Heinemeier, M. Bayer, M. Hansen, L. Holm, S. Doessing, M. Kongsgaard, M. Krogsgaard, and S. Magnusson. 2009. "From Mechanical Loading to Collagen Synthesis, Structural Changes and Function in Human Tendon." *Scand J Med Sci Spor* 19 (4): 500–510. doi:10.1111/j.1600-0838.2009.00986.x.

Kohler, Julia, Cvetan Popov, Barbara Klotz, Paolo Alberton, Wolf Christian Prall, Florian Haasters, Sigrid Muller-Deubert, et al. 2013. "Uncovering the Cellular and Molecular Changes in Tendon Stem/Progenitor Cells Attributed to Tendon Aging and Degeneration." *Aging Cell* 12 (6). Wiley Online Library: 988–99.

Korenkova, V., J. Scott, V. Novosadova, M. Jindrichova, Šidova M Langerova L. Švec D, and Sjoback R. 2015. "Pre-Amplification in the Context of High-Throughput

QPCR Gene Expression Experiment." *BMC Molecular Biology* 16: 5.

Kostka, Dennis, Alisha K Holloway, and Katherine S Pollard. 2018. "Developmental Loci Harbor Clusters of Accelerated Regions That Evolved Independently in Ape Lineages." *Molecular Biology and Evolution* 35 (8). Oxford University Press: 2034–45.

Koyama, E., Y. Shibukawa, M. Nagayama, and others. 2008. "A Distinct Cohort of Progenitor Cells Participates in Synovial Joint and Articular Cartilage Formation During Mouse Limb Skeletogenesis." *Developmental Biology* 316 (1): 62–73.

Kozhemyakina, E., A. B. Lassar, and E. Zelzer. 2015. "A Pathway to Bone: Signaling Molecules and Transcription Factors Involved in Chondrocyte Development and Maturation." *Development* 142 (5): 817–31.

Kraft, K., S. Geuer, A. J. Will, and others. 2015. "Cell Reports." *Deletions, Inversions, Duplications: Engineering of Structural Variants Using CRISPR/Cas in Mice*.

Kraus, P., S. V. Xing, and T. Lufkin. 2014. "Generating Mouse Lines for Lineage Tracing and Knock- Out Studies." In *Mouse Genetics: Methods and Protocols, Methods in Molecular Biology*, edited by S. R. Singh and V Coppola. New York, NY: Springer Science+Business Media.

Krebs, J. E., E. Gi, and S. T. Kilpatrick. 2014. *Lewin's Genes Xi*. Burlington, MA: Jones & Bartlett Learning.

Kroneis, T., E. Jonasson, D. Andersson, S. Dolatabadi, and A. Staahlberg. 2017. "Global Preamplification Simplifies Targeted MRNA Quantification." *Scientific Reports* 7: 45219.

Lanctôt, C., A. Moreau, M. Chamberland, M. L. Tremblay, and J. Drouin. 1999. "Hindlimb Patterning and Mandible Development Require the Pitx1 Gene." *Develop-*

ment 126 (9): 1805–10.

Langmead, Ben, and Steven L Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods* 9 (4). Nature Publishing Group: 357.

Lappalainen, T., M. Sammeth, M. R. Friedlander, and others. 2013. "Transcriptome and Genome Sequencing Uncovers Functional Variation in Humans." *Nature* 501 (7468): 506–11.

Leite, G. M., N. Magan, and A. Medina. 2012. "Comparison of Different Bead-Beating Rna Extraction Strategies: An Optimized Method for Filamentous Fungi." *J Microbiol Methods* 10 (1016): 413–18. doi:88:. /j.mimet.2012.01.011.

Lejard, Véronique, Frédéric Blais, Marie-Justine Guerquin, Aline Bonnet, Marie-Ange Bonnin, Emmanuelle Havis, Maryline Malbouyres, et al. 2011. "EGR1 and EGR2 Involvement in Vertebrate Tendon Differentiation." *J Biol Chem* 286 (7): 5855–67. doi:10.1074/jbc.M110.153106.

Lemay, P., M. C. Guyot, E. Tremblay, and others. 2015. "Loss-of-Function de Novo Mutations Play an Important Role in Severe Human Neural Tube Defects." *Journal of Medical Genetics* 52 (7): 493–97.

Lettice, L. A., S. J. Heaney, L. A. Purdie, and others. 2003. "A Long-Range Shh Enhancer Regulates Expression in the Developing Limb and Fin and Is Associated with Preaxial Polydactyly." *Human Molecular Genetics* 12 (14): 1725–35.

Lettice, L. A., A. E. Hill, P. S. Devenney, and R. E. Hill. 2008. "Point Mutations in a Distant Sonic Hedgehog Cis-Regulator Generate a Variable Regulatory Output Responsible for Preaxial Polydactyly." *Human Molecular Genetics* 17 (7): 978–85.

Li, Dongming, Ana Marie S Palanca, So Youn Won, Lei Gao, Ying Feng, Ajay A Vashisht, Li Liu, et al. 2017. "The Mbd7 Complex Promotes Expression of Methylated Transgenes Without Significantly Altering Their Methylation Status." *Elife* 6. eLife

Sciences Publications Limited: e19893.

Li, Hao, Jonathan C Radford, Michael J Ragusa, Katherine L Shea, Scott R McKercher, Jeffrey D Zaremba, Walid Soussou, et al. 2008. "Transcription Factor Mef2c Influences Neural Stem/Progenitor Cell Differentiation and Maturation in Vivo." *Proceedings of the National Academy of Sciences* 105 (27). National Acad Sciences: 9397–9402.

Li, Y. R., and B. J. Keating. 2014. "Trans-Ethnic Genome-Wide Association Studies: Advantages and Challenges of Mapping in Diverse Populations." *Genome Medicine* 6 (10): 91.

Liang, P., Y. Xu, X. Zhang, and others. 2015. "CRISPR/Cas9-Mediated Gene Editing in Human Tripronuclear Zygotes." *Protein and Cell* 6 (5): 363–72.

Lieberman-Aiden, E., N. L. van Berkum, L. Williams, and others. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science* 326 (5950): 289–93.

Lima, Joana Esteves de, Marie-Ange Bonnin, Adeline Bourgeois, Alice Parisi, Fabien Le Grand, and Delphine Duprez. 2014. "Specific Pattern of Cell Cycle During Limb Fetal Myogenesis." *Developmental Biology* 392 (2). Elsevier: 308–23.

Lin, Qing, Jianrong Lu, Hiromi Yanagisawa, Robert Webb, Gary E Lyons, James A Richardson, and Eric N Olson. 1998. "Requirement of the Mads-Box Transcription Factor Mef2c for Vascular Development." *Development* 125 (22). Oxford University Press for The Company of Biologists Limited: 4565–74.

Lindsay, WK, and JR Birch. 1964. "The Fibroblast in Flexor Tendon Healing." *Plastic and Reconstructive Surgery* 34 (3). LWW: 223–32.

Liu, Han, Jingyue Xu, Chia-Feng Liu, Yu Lan, Christopher Wylie, and Rulang Jiang. 2015. "Whole Transcriptome Expression Profiling of Mouse Limb Tendon Development

by Using RNA-seq" 33 (6): 840–48. doi:10.1002/jor.22886.

Liu, Huanhuan, Shouan Zhu, Can Zhang, Ping Lu, Jiajie Hu, Zi Yin, Yue Ma, Xiao Chen, and OuYang, Hongwei. 2014. "Crucial Transcription Factors in Tendon Development and Differentiation: Their Potential for Tendon Regeneration." *Cell Tissue Res* 356 (2): 287–98. doi:10.1007/s00441-014-1834-8.

Liu, Yanxin, Andrea G Schwartz, Victor Birman, Stavros Thomopoulos, and Guy M Genin. 2014. "Stress Amplification During Development of the Tendon-to-Bone Attachment." *Biomechanics and Modeling in Mechanobiology* 13 (5). Springer: 973–83.

Livak, Kenneth J, and Thomas D Schmittgen. 2001. "Analysis of Relative Gene Expression Data Using Real-Time Quantitative Pcr and the DdCT Method." *Methods* 25 (4). Elsevier: 402–8.

Lizio, M., J. Harshbarger, H. Shimoji, and others. 2015. "Gateways to the Fantom5 Promoter Level Mammalian Expression Atlas." *Genome Biology* 16: 22.

Logan, M., and C. J. Tabin. 1999. "Role of Pitx1 Upstream of Tbx4 in Specification of Hindlimb Identity." *Science* 283 (5408): 1736–9.

Loiselle, Alayna, Kiminori Yukata, Michael Geary, Sirish Kondabolu, Shanshan Shi, Jennifer Jonason, Hani Awad, and O'Keefe, Regis. 2015. "Development of Antisense Oligonucleotide (ASO) Technology Against Tgf-B Signaling to Prevent Scarring During Flexor Tendon Repair." *J Orthopaed Res* 33 (6): 859–66. doi:10.1002/jor.22890.

Lorda-Diez, Carlos I, Juan A Montero, Carmen Martinez-Cue, Juan A Garcia-Porrero, and Juan M Hurle. 2009. "Transforming Growth Factors B Coordinate Cartilage and Tendon Differentiation in the Developing Limb Mesenchyme." *Journal of Biological Chemistry* 284 (43). ASBMB: 29988–96.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for Rna-Seq Data with Deseq2." *Genome Biology* 15

(12): 550. doi:10.1186/s13059-014-0550-8.

Low, Boon Chuan, Catherine Qiurong Pan, GV Shivashankar, Alexander Bershadsky, Marius Sudol, and Michael Sheetz. 2014. "YAP/Taz as Mechanosensors and Mechanotransducers in Regulating Organ Size and Tumor Growth." *FEBS Letters* 588 (16). Wiley Online Library: 2663–70.

Lui, J. C., A. C. Andrade, P. Forcinito, and others. 2010. "Spatial and Temporal Regulation of Gene Expression in the Mammalian Growth Plate." *Bone* 46 (5): 1380–90.

Lupiáñez, D. G., K. Kraft, V. Heinrich, and others. 2015. "Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions." *Cell* 161 (5): 1012–25.

M-j, Guerquin, B. Charvet, G. Nourissat, E. Havis, O. Ronsin, Bonnin M-a, M. Ruggiu, I. Olivera-Martinez, N. Robert, and Y. Lu. 2013. "Transcription Factor Egr1 Directs Tendon Differentiation and Promotes Tendon Repair." *The Journal of Clinical Investigation* 123: 3564.

MacCabe, A. B., M. T. Gasseling, and J. W. 1973. "Spatiotemporal Distribution of Mechanisms That Control Outgrowth and Anteroposterior Polarization of the Limb Bud in the Chick Embryo." *Mechanisms of Ageing and Development* 2 (1): 1–12.

Macedo, N. J., and Ferreira TL. 2014. "Maximizing Total Rna Yield from Trizol Reagent Protocol: A Feasibility Study." *ASEE Zone I Conference*.

Macosko, E. Z., A. Basu, R. Satija, and others. 2015. "Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell* 161 (5): 1202–14.

Madrigal, P., and P. Krajewski. 2012. "Current Bioinformatic Approaches to Identify Dnase I Hyper- Sensitive Sites and Genomic Footprints from Dnase-Seq Data."

Frontiers in Genetics 3: 230.

Maeda, Toru, Tomoya Sakabe, Ataru Sunaga, Keiko Sakai, Alexander L. Rivera, Douglas R. Keene, Takako Sasaki, et al. 2011. "Conversion of Mechanical Force into TGF-B-Mediated Biochemical Signals." *Curr Biol* 21 (11): 933–41. doi:10.1016/j.cub.2011.04.007.

Magavi, Sanjay S, and Jeffrey D Macklis. 2008. "Identification of Newborn Cells by Brdu Labeling and Immunocytochemistry in Vivo." In *Neural Stem Cells*, 335–43. Springer.

Magnusson, SP, Katja M Heinemeier, and Michael Kjaer. 2016. "Collagen Homeostasis and Metabolism." *Adv. Exp. Med. Biol.* 920: 11–25. doi:10.1007/978-3-319-33943-6_2.

Majewski, J., and T. Pastinen. 2011. "The Study of EQTL Variations by Rna-Seq: From Snps to Phenotypes." *Trends in Genetics* 27 (2): 72–79.

Mann, B, M Gelos, A Siedow, ML Hanski, A Gratchev, M Ilyas, WF Bodmer, et al. 1999. "Target Genes of B-Catenin–T Cell-Factor/Lymphoid-Enhancer-Factor Signaling in Human Colorectal Carcinomas." *Proceedings of the National Academy of Sciences* 96 (4). National Acad Sciences: 1603–8.

Marinić, M., T. Aktas, S. Ruf, and F. Spitz. 2013. "An Integrated Holo-Enhancer Unit Defines Tis- Sue and Gene Specificity of the Fgf8 Regulatory Landscape." *Developmental Cell* 24 (5): 530–42.

Massague, Joan. 2012. "TGF-B Signalling in Context." *Nature Reviews Molecular Cell Biology* 13 (10). Nature Publishing Group: 616.

Mauviel, Alain. 2005. "Transforming Growth Factor-Beta: A Key Mediator of Fibrosis"

117: 69–80. doi:10.1385/1-59259-940-0:069.

McLean, C. Y., P. L. Reno, A. A. Pollen, and others. 2011. "Human-Specific Loss of Regulatory Dna and the Evolution of Human-Specific Traits." *Nature* 471 (7337): 216–19.

McNeill, Helen, and James R Woodgett. 2010. "When Pathways Collide: Collaboration and Connivance Among Signalling Proteins in Development." *Nature Reviews Molecular Cell Biology* 11 (6). Nature Publishing Group: 404.

Mendias, Christopher L, Konstantin I Bakhurin, and John A Faulkner. 2008. "Tendons of Myostatin-Deficient Mice Are Small, Brittle, and Hypocellular." *Proceedings of the National Academy of Sciences* 105 (1). National Acad Sciences: 388–93.

Mendias, Christopher L, Jonathan P Gumucio, Konstantin I Bakhurin, Evan B Lynch, and Susan V Brooks. 2012. "Physiological Loading of Tendons Induces Scleraxis Expression in Epitenon Fibroblasts" 30 (4): 606–12. doi:10.1002/jor.21550.

Mendias, Christopher, Jonathan Gumucio, Max Davis, Caleb Bromley, Carol Davis, and Susan Brooks. 2012. "Transforming Growth Factor-beta Induces Skeletal Muscle Atrophy and Fibrosis Through the Induction of Atrogin-1 and Scleraxis." *Muscle Nerve* 45 (1): 55–59. doi:10.1002/mus.22232.

Menke, D. B. 2013. "Engineering Subtle Targeted Mutations into the Mouse Genome." *Genesis* 51 (9): 605–18.

Merino, R., D. Macias, Y. Ganan, and others. 1999. "Expression and Function of Gdf-5 During Digit Skeletogenesis in the Embryonic Chick Leg Bud." *Developmental Biology* 206 (1): 33–45.

Messier, B, and CP Leblond. 1960. "Cell Proliferation and Migration as Revealed by Radioautography After Injection of Thymidine-H3 into Male Rats and Mice."

American Journal of Anatomy 106 (3). Wiley Online Library: 247–85.

Meyer, C. A., and X. S. Liu. 2014. "Identifying and Mitigating Bias in Next-Generation Sequencing Methods for Chromatin Biology." *Nature Reviews Genetics* 15 (11): 709–21.

Miyajima, Atsushi, Minoru Tanaka, and Tohru Itoh. 2014. "Stem/Progenitor Cells in Liver Development, Homeostasis, Regeneration, and Reprogramming." *Cell Stem Cell* 14 (5). Elsevier: 561–74.

Montesano, R, and Orci - L the of Sciences. 1988. "Transforming Growth Factor Beta Stimulates Collagen-Matrix Contraction by Fibroblasts: Implications for Wound Healing." *... of the National Academy of Sciences*. doi:10.1073/pnas.85.13.4894.

Montgomery, R D. 1989. "Healing of Muscle, Ligaments, and Tendons." *Semin Vet Med Surg (Small Anim)* 4 (4): 304–11.

Mori, Y., U. I. Chung, S. Tanaka, and T. Saito. 2014. "Determination of Differential Gene Expres- Sion Profiles in Superficial and Deeper Zones of Mature Rat Articular Cartilage Using Rna Sequencing of Laser Microdissected Tissue Specimens." *Biomedical Research* 35 (4): 263–70.

Mortlock, D. P., C. Guenther, and D. M. Kingsley. 2003. "A General Approach for Identifying Distant Regulatory Elements Applied to the Gdf6 Gene." *Genome Research* 13 (9): 2069–81.

Mouselimis, Lampros. 2019. *ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering*. https://CRAN.R-project.org/package=ClusterR.

Murchison, Nicholas, Brian Price, David Conner, Douglas Keene, Eric Olson, Clifford Tabin, and Ronen Schweitzer. 2007. "Regulation of Tendon Differentiation by Scleraxis Distinguishes Force-Transmitting Tendons from Muscle-Anchoring Tendons."

Development 134 (14): 2697–2708. doi:10.1242/dev.001933.

Nagai, H., and M. Aoki. 2002. "Inhibition of Growth Plate Angiogenesis and Endochondral Ossifi- Cation with Diminished Expression of Mmp-13 in Hypertrophic Chondrocytes in Fgf-2-Treated Rats." *Journal of Bone and Mineral Metabolism* 20 (3): 142–47.

Nam, K., K. Munch, A. Hobolth, and others. 2015. "Extreme Selective Sweeps Independently Targeted the X Chromosomes of the Great Apes." *Proceedings of the National Academy of Sciences of the USA* 112 (20): 6413–8.

Nielsen, R., N. Clausen, P. Schjerling, J. O. Larsen, T. Martinussen, E. List, J. Kopchick, M. Kjær, and K. Heinemeier. 2014. "Chronic Alterations in Growth Hormone/Insulin-Like Growth Factor-I Signaling Lead to Changes in Mouse Tendon Structure." *Matrix Biology* 34: 96–104.

Nikolskiy, I., D. F. Conrad, S. Chun, and others. 2015. "Using Whole-Genome Sequences of the Lg/J and Sm/J Inbred Mouse Strains to Prioritize Quantitative Trait Genes and Nucleotides." *BMC Genomics* 16: 415.

Nilsson, O., M. H. Guo, N. Dunbar, and others. 2014. "Short Stature, Accelerated Bone Maturation, and Early Growth Cessation Due to Heterozygous Aggrecan Mutations." *Journal of Clinical Endocrinology and Metabolism* 99 (8): E1510–1518.

Niu, Y., B. Shen, Y. Cui, and others. 2014. "Generation of Gene-Modified Cynomolgus Monkey via Cas9/Rna-Mediated Gene Targeting in One-Cell Embryos." *Cell* 156 (4): 836–43.

Nourissat, G., F. Berenbaum, and D. Duprez. 2015. "Tendon Injury: From Biology to Tendon Repair." *Nature Reviews Rheumatology* 11: 223–33.

Ogle, Derek H., Powell Wheeler, and Alexis Dinno. 2019. *FSA: Fisheries Stock*

Analysis. https://github.com/droglenc/FSA.

Oh, C. D., Y. Lu, S. Liang, and others. 2014. "SOX9 Regulates Multiple Genes in Chondrocytes, Including Genes Encoding Ecm Proteins, Ecm Modification Enzymes, Receptors, and Transporters." *PLoS ONE* 9: 9. doi:e107577.

Ohba, S., X. He, H. Hojo, and A. P. McMahon. 2015. "Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte." *Cell Report* 12 (2): 229–43.

Orlando, V. 2000. "Mapping Chromosomal Proteins in Vivo by Formaldehyde-Crosslinked-Chromatin Immunoprecipitation." *Trends in Biochemical Science* 25 (3): 99–104.

Ozsolak, F., and P. M. Milos. 2011. "RNA Sequencing: Advances, Challenges and Opportunities." *Nature Reviews Genetics* 12 (2): 87–98.

Pacifici, M., E. Koyama, Y. Shibukawa, and others. 2006. "Cellular and Molecular Mechanisms of Synovial Joint and Articular Cartilage Formation." *Annals of the New York Academy of Science* 1068: 74–86.

Pan, Yu Chih, Xiao Wen Wang, Han Feng Teng, Yi Ju Wu, Hsuan Chia Chang, and Shen Liang Chen. 2015. "Wnt3a Signal Pathways Activate MyoD Expression by Targeting Cis-Elements Inside and Outside Its Distal Enhancer." *Bioscience Rep* 35 (2): e00180. doi:10.1042/BSR20140177.

Pardo-Diaz, C., C. Salazar, and C. D. Jiggins. 2015. "Towards the Identification of the Loci of Adaptive Evolution." *Methods in Ecology and Evolution* 6 (4): 445–64.

Parry, D. A., C. V. Logan, A. P. Stegmann, and others. 2013. "SAMS, a Syndrome of Short Stature, Auditory-Canal Atresia, Mandibular Hypoplasia, and Skeletal Abnormalities Is a Unique Neurocristopathy Caused by Mutations in Goosecoid." *American*

Journal of Human Genetics 93 (6): 1135–42.

Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. 2017. "Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression." *Nature Methods* 14 (4). Nature Publishing Group: 417.

Pazin, D. E., L. W. Gamer, K. A. Cox, and V. Rosen. 2012. "Molecular Profiling of Synovial Joints: Use of Microarray Analysis to Identify Factors That Direct the Development of the Knee and Elbow." *Developmental Dynamics* 241 (11): 1816–26.

Peffers, Mandy Jayne, Yongxiang Fang, Kathleen Cheung, Tim Koh Jia Wei, Peter David Clegg, and Helen Lucy Birch. 2015. "Transcriptome Analysis of Ageing in Uninjured Human Achilles Tendon." *Arthritis Research and Therapy* 17 (1). BioMed Central: 33.

Perry, G. H., P. Melsted, J. C. Marioni, and others. 2012. "Comparative Rna Sequencing Reveals Substantial Genetic Variation in Endangered Primates." *Genome Research* 22 (4): 602–10.

Piersma, Bram, Ruud A Bank, and Miriam Boersema. 2015. "Signaling in Fibrosis: TGF-B, Wnt, and Yap/Taz Converge." *Frontiers in Medicine* 2. Frontiers: 59.

Pokholok, D. K., C. T. Harbison, S. Levine, and others. 2005. "Genome-Wide Map of Nucleosome Acetylation and Methylation in Yeast." *Cell* 122 (4): 517–27.

Pollard, Katherine S, Sofie R Salama, Bryan King, Andrew D Kern, Tim Dreszer, Sol Katzman, Adam Siepel, et al. 2006. "Forces Shaping the Fastest Evolving Regions in the Human Genome." *PLoS Genetics* 2 (10). Public Library of Science: e168.

Pon, JR, and Marra - MA Oncotarget. 2016. "MEF2 Transcription Factors: Developmental Regulators and Emerging Cancer Genes." *Oncotarget*.

Pon, Julia R, and Marco A Marra. 2016. "MEF2 Transcription Factors: Developmental

Regulators and Emerging Cancer Genes." *Oncotarget* 7 (3). Impact Journals, LLC: 2297.

Porcu, E., S. Sanna, C. Fuchsberger, and L. G. Fritsche. 2013. "Genotype Imputation in Genome-Wide Association Studies." *Current Protocol in Human Genetics* 1: 25.

Porrello, Enzo, Ahmed Mahmoud, Emma Simpson, Joseph Hill, James Richardson, Eric Olson, and Hesham Sadek. 2011. "Transient Regenerative Potential of the Neonatal Mouse Heart." *Science* 331 (6020): 1078–80. doi:10.1126/science.1200708.

Potthoff, Matthew J, and Eric N Olson. 2007. "MEF2: A Central Regulator of Diverse Developmental Programs." *Development* 134 (23). The Company of Biologists Ltd: 4131–40.

Potthoff, Matthew, and Eric Olson. 2007. "MEF2: A Central Regulator of Diverse Developmental Programs." *Development* 134 (23): 4131–40. doi:10.1242/dev.008367.

Prabhakar, S., J. P. Noonan, S. Paabo, and E. M. Rubin. 2006. "Accelerated Evolution of Conserved Noncoding Sequences in Humans." *Science* 314 (5800): 786.

Prabhakar, S., A. Visel, J. A. Akiyama, and others. 2008. "Human-Specific Gain of Function in a Developmental Enhancer." *Science* 321 (5894): 1346–50.

Prado-Martinez, J., P. H. Sudmant, J. M. Kidd, and others. 2013. "Great Ape Genetic Diversity and Population History." *Nature* 499 (7459): 471–75.

Preuschoft, H. 1992. "Biomechanical Factors That Influence Overall Body Shape of Large Apes and Humans." *Topics in Primatology* 3. University of Tokyo Press: 259–89.

Prihantono, Prihantono, Mochammad Hatta, Christian Binekada, Daniel Sampepajung, Haryasena Haryasena, Berti Nelwan, Andi Asadul Islam, and Andi Nilawati Usman. 2017. "Ki-67 Expression by Immunohistochemistry and Quantitative Real-

Time Polymerase Chain Reaction as Predictor of Clinical Response to Neoadjuvant Chemotherapy in Locally Advanced Breast Cancer." *Journal of Oncology* 2017. Hindawi.

Provot, S., H. Kempf, L. C. Murtaugh, and others. 2006. "Nkx/Bapx1 Acts as a Negative Regulator of Chondrocyte Maturation." *Development* 133 (4): 651–62.

Prüfer, K., K. Munch, I. Hellmann, and others. 2012. "The Bonobo Genome Compared with the Chimpanzee and Human Genomes." *Nature* 486 (7404): 527–31.

Pryce, Brian, Spencer Watson, Nicholas Murchison, Julia Staverosky, Nicole Dünker, and Ronen Schweitzer. 2009. "Recruitment and Maintenance of Tendon Progenitors by TGFB Signaling Are Essential for Tendon Formation." *Development* 136 (8): 1351–61. doi:10.1242/dev.027342.

Quinlan, Aaron R, and Ira M Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6). Oxford University Press: 841–42.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rada, Jody A, Pamela K Cornuet, and John R Hassell. 1993. "Regulation of Corneal Collagen Fibrillogenesis in Vitro by Corneal Proteoglycan (Lumican and Decorin) Core Proteins." *Experimental Eye Research* 56 (6). Elsevier: 635–48.

Rada-Iglesias, A., R. Bajpai, T. Swigut, and others. 2011. "A Unique Chromatin Signature Uncovers Early Developmental Enhancers in Humans." *Nature* 470 (7333): 279–83.

Rajkumar, A. P., P. Qvist, R. Lazarus, F. Lescai, J. Ju, M. Nyegaard, O. Mors, A. D. Borglum, Q. Li, and Christensen JH. 2015. "Experimental Validation of Methods

for Differential Gene Expression Analysis and Sample Pooling in Rna-Seq." *BMC Genomics* 16: 548.

Rauwerdink, IG. 1991. "Muscle Fibre and Tendon Lengths in the Distal Limb Segments of Primates." *Z Morphol Anthropol* 78 (3): 331–40.

Record, Comer - RD. 1956. "An Experimental Study of the 'Laws' of Muscle and Tendon Growth." *The Anatomical Record*. doi:10.1002/ar.1091250402.

Reilly, S. K., J. Yin, A. E. Ayoub, and others. 2015. "Evolutionary Genomics: Evolutionary Changes in Promoter and Enhancer Activity During Human Corticogenesis." *Science* 347 (6226): 1155–9.

Reno, C., L. Marchuk, P. Sciore, C. B. Frank, and Hart DA. 1997. "Rapid Isolation of Total Rna from Small Samples of Hypocellular, Dense Connective Tissues." *Biotechniques* 22: 1082–6.

reviews, Kjær - M. 2004. "Role of Extracellular Matrix in Adaptation of Tendon and Skeletal Muscle to Mechanical Loading." *Physiological Reviews*.

Riddle, R. D., R. L. Johnson, E. Laufer, and C. Tabin. 1993. "Sonic Hedgehog Mediates the Polarizing Activity of the Zpa." *Cell* 75 (7): 1401–16.

Riemondy, Kent A., Ryan M. Sheridan, Austin Gillen, Yinni Yu, Christopher G. Bennett, and Jay R. Hesselberth. 2017. "Valr: Reproducible Genome Interval Arithmetic in R." *F1000Research*. doi:10.12688/f1000research.11997.1.

Rigozzi, S., R. Müller, and J. Snedeker. 2010. "Collagen Fibril Morphology and Mechanical Properties of the Achilles Tendon in Two Inbred Mouse Strains." *J Anat* 216 (6): 724–31. doi:10.1111/j.1469-7580.2010.01225.x.

Rio, Donald C, Manuel Ares, Gregory J Hannon, and Timothy W Nilsen. 2010. "Purification of Rna Using Trizol (Tri Reagent)." *Cold Spring Harbor Protocols* 2010

(6). Cold Spring Harbor Laboratory Press: pdb–prot5439.

Rivera, C. M., and B. Ren. 2013. "Mapping Human Epigenomes." *Cell* 155 (1): 39–55.

Rock, J. R., M. C. Lopez, H. V. Baker, and B. D. Harfe. 2007. "Identification of Genes Expressed in the Mouse Limb Using a Novel Zpa Microarray Approach." *Gene Expression Patterns* 8 (1): 19–26.

Rodriguez-Esteban, C., T. Tsukui, S. Yonei, and others. 1999. "The T-Box Genes Tbx4 and Tbx5 Regu- Late Limb Outgrowth and Identity." *Nature* 398 (6730): 814–18.

Romero, Gallego, Pavlovic I., B. J., I. Hernando-Herraez, and others. 2015. "A Panel of Induced Pluripotent Stem Cells from Chimpanzees: A Resource for Comparative Functional Genomics." *Elife* 4.

Romero, I. G., A. A. Pai, J. Tung, and Y. Gilad. 2014. "RNA-Seq: Impact of Rna Degradation on Transcript Quantification." *BMC Biology* 12 (42).

Ross-Innes, Caryn S., Rory Stark, Andrew E. Teschendorff, Kelly A. Holmes, H. Raza Ali, Mark J. Dunning, Gordon D. Brown, et al. 2012. "Differential Oestrogen Receptor Binding Is Associated with Clinical Outcome in Breast Cancer." *Nature* 481: –4. http://www.nature.com/nature/journal/v481/n7381/full/nature10730.html.

Rosso, Claudio, Philipp Schuetz, Caroline Polzer, Lukas Weisskopf, Ulrich Studler, and Victor Valderrabano. 2012. "Physiological Achilles Tendon Length and Its Relation to Tibia Length." *Clinical Journal of Sport Medicine* 22 (6). LWW: 483–87.

Roux, J., M. Rosikiewicz, and M. Robinson-Rechavi. 2015. "What to Compare and How: Comparative Transcriptomics for Evo–Devo." *Journal of Experimental Zoology Series B Molecular Development and Evolution* 324 (4): 372–82.

RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston,

MA: RStudio, Inc. http://www.rstudio.com/.

Ruchti, Charles, Denis Haller, Markus Nuber, and Hans Cottier. 1983. "Regional Differences in Renewal Rates of Fibroblasts in Young Adult Female Mice." *Cell and Tissue Research* 232 (3). Springer: 625–36.

Rudnicki, Michael, Patrick Schnegelsberg, Ronald Stead, Thomas Braun, Hans-Henning Arnold, and Rudolf Jaenisch. 1993. "MyoD or Myf-5 Is Required for the Formation of Skeletal Muscle." *Cell* 75 (7): 1351–9. doi:10.1016/0092-8674(93)90621-V.

Runesson, Eva, Paul Ackermann, Helena Brisby, Jón Karlsson, and Bengt I Eriksson. 2013. "Detection of Slow-Cycling and Stem/Progenitor Cells in Different Regions of Rat Achilles Tendon: Response to Treadmill Exercise." *Knee Surgery, Sports Traumatology, Arthroscopy* 21 (7). Springer: 1694–1703.

Sabo, P. J., R. Humbert, M. Hawrylycz, and others. 2004. "Genome-Wide Identification of Dnasei Hypersensitive Sites Using Active Chromatin Sequence Libraries." *Proceedings of the National Academy of Science of the USA* 101 (13): 4537–42.

Sagai, T., M. Hosoya, Y. Mizushina, M. Tamura, and T. Shiroishi. 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 (4): 797–803.

Sakabe, Tomoya, Keiko Sakai, Toru Maeda, Ataru Sunaga, Nao Furuta, Ronen Schweitzer, Takako Sasaki, and Takao Sakai. 2018. "Transcription Factor Scleraxis Vitally Contributes to Progenitor Lineage Direction in Wound Healing of Adult Tendon in Mice." *Journal of Biological Chemistry* 293 (16). ASBMB: 5766–80.

Salic, Adrian, and Timothy J Mitchison. 2008. "A Chemical Method for Fast and Sensitive Detection of Dna Synthesis in Vivo." *Proceedings of the National Academy*

of Sciences 105 (7). National Acad Sciences: 2415–20.

Scally, A., J. Y. Dutheil, L. W. Hillier, and others. 2012. "Insights into Hominid Evolution from the Gorilla Genome Sequence." *Nature* 483 (7388): 169–75.

Scherz, P. J., E. McGlinn, S. Nissim, and C. J. Tabin. 2007. "Extended Exposure to Sonic Hedgehog Is Required for Patterning the Posterior Digits of the Vertebrate Limb." *Developmental Biology* 308 (2): 343–54.

Schiele, Nathan, Joseph Marturano, and Catherine Kuo. 2013. "Mechanical Factors in Embryonic Tendon Development: Potential Cues for Stem Cell Tenogenesis." *Curr Opin Biotech* 24 (5): 834–40. doi:10.1016/j.copbio.2013.07.003.

Schleifman, Erica B, Rupal Desai, Jill M Spoerke, Yuanyuan Xiao, Cheryl Wong, Ilma Abbas, Carol O'Brien, et al. 2014. "Targeted Biomarker Profiling of Matched Primary and Metastatic Estrogen Receptor Positive Breast Cancers." *PLoS One* 9 (2). Public Library of Science: e88401.

Schmittgen, Thomas, and Kenneth Livak. 2008. "Analyzing Real-Time PCR Data by the Comparative CT Method." *Nat Protoc* 3 (6): 1101–8. doi:10.1038/nprot.2008.73.

Schreiner, C. M., S. M. Bell, and W. J. 2009. "Microarray Analysis of Murine Limb Bud Ectoderm and Mesoderm After Exposure to Cadmium or Acetazolamide." *Birth Defects Research Series A Clinical and Molecular Teratology* 85 (7): 588–98.

Schweitzer, R, JH Chyung, LC Murtaugh, and Brent - AE .... 2001. "Analysis of the Tendon Cell Fate Using Scleraxis, a Specific Marker for Tendons and Ligaments." ....

Senyo, Samuel E, Richard T Lee, and Bernhard Kuhn. 2014. "Cardiac Regeneration Based on Mechanisms of Cardiomyocyte Proliferation and Differentiation." *Stem Cell Research* 13 (3). Elsevier: 532–41.

Shao, Y., Y. Guan, L. Wang, and others. 2014. "CRISPR/Cas-Mediated Genome

Editing in the Rat via Direct Injection of One-Cell Embryos." *Nature Protocols* 9 (10): 2493–2512.

Shapiro, M. D., M. E. Marks, C. L. Peichel, and others. 2004. "Genetic and Developmental Basis of Evolutionary Pelvic Reduction in Threespine Sticklebacks." *Nature* 428 (6984): 717–23.

Sherwood, R. J., D. L. Duren, L. M. Havill, and others. 2008. "A Genomewide Linkage Scan for Quantitative Trait Loci Influencing the Craniofacial Complex in Baboons Papio Hamadryas Spp." *Genetics* 180 (1): 619–28.

Shih, J. H., A. M. Michalowska, K. Dobbin, Y. Ye, T. H. Qiu, and Green JE. 2004. "Effects of Pooling MRNA in Microarray Class Comparisons." *Bioinformatics* 20: 3318–25.

Shou, S., V. Scott, C. Reed, R. Hitzemann, and H. S. Stadler. 2005. "Transcriptome Analysis of the Murine Forelimb and Hindlimb Autopod." *Developmental Dynamics* 234 (1): 74–89.

Shukunami, Chisa, Yusuke Oshima, and Yuji Hiraki. 2001. "Molecular Cloning of Tenomodulin, a Novel Chondromodulin-I Related Gene." *Biochemical and Biophysical Research Communications* 280 (5). Elsevier: 1323–7.

Shukunami, Chisa, Aki Takimoto, Miwa Oro, and Yuji Hiraki. 2006. "Scleraxis Positively Regulates the Expression of Tenomodulin, a Differentiation Marker of Tenocytes." *Developmental Biology* 298 (1). Elsevier: 234–47.

Silver, Frederick, Joseph Freeman, and Gurinder Seehra. 2003. "Collagen Self-Assembly and the Development of Tendon Mechanical Properties." *J Biomech* 36 (10): 1529–53. doi:10.1016/S0021-9290(03)00135-0.

Simons, Benjamin D, and Hans Clevers. 2011. "Stem Cell Self-Renewal in Intestinal

Crypt." *Experimental Cell Research* 317 (19). Elsevier: 2719–24.

Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2015. "Differential Analyses for Rna-Seq: Transcript-Level Estimates Improve Gene-Level Inferences." *F1000Research* 4 (1521). doi:10.12688/f1000research.7563.1.

Soriano, P. 1999. "Generalized LacZ Expression with the Rosa26 Cre Reporter Strain." *Nature Genetics* 21 (1): 70–71.

Stephens, T. D., and T. R. McNulty. 1981. "Evidence for a Metameric Pattern in the Development of the Chick Humerus." *Journal of Embryology and Experimental Morphology* 61: 191–205.

Stergachis, A. B., S. Neph, R. Sandstrom, and others. 2014. "Conservation of Trans-Acting Circuitry During Mammalian Regulatory Evolution." *Nature* 515 (7527): 365–70.

Subramanian, Arul, and Thomas Schilling. 2015. "Tendon Development and Musculoskeletal Assembly: Emerging Roles for the Extracellular Matrix." *Development* 142 (24): 4191–4204. doi:10.1242/dev.114777.

Sudmant, P. H., J. Huddleston, C. R. Catacchio, and others. 2013. "Evolution and Diversity of Copy Number Variation in the Great Ape Lineage." *Genome Research* 23 (9): 1373–82.

Sugg, Kristoffer, James Markworth, Nathaniel Disser, Andrew Rizzi, Jeffrey Talarek, Dylan Sarver, Susan Brooks, and Christopher Mendias. 2018. "Postnatal Tendon Growth and Remodeling Require Platelet-Derived Growth Factor Receptor Signaling." *Am J Physiol-Cell Ph* 314 (4): C389–C403. doi:10.1152/ajpcell.00258.2017.

Sultan, M., I. Piccini, D. Balzereit, R. Herwig, N. G. Saran, H. Lehrach, R. H. Reeves, and Yaspo M-L. 2007. "Gene Expression Variation in down's Syndrome Mice Allows

Prioritization of Candidate Genes." *Genome Biology* 8: R91.

Summerbell, D. 1979. "The Zone of Polarizing Activity: Evidence for a Role in Normal Chick Limb Morphogenesis." *Journal of Embryology and Experimental Morphology* 50: 217–33.

Summerbell, D., P. R. Ashby, O. Coutelle, and others. 2000. "The Expression of Myf5 in the Developing Mouse Embryo Is Controlled by Discrete and Dispersed Enhancers Specific for Particular Populations of Skeletal Muscle Precursors." *Development* 127 (17): 3745–57.

Sun, M., F. Ma, X. Zeng, and others. 2008. "Triphalangeal Thumb-Polysyndactyly Syndrome and Syndactyly Type Iv Are Caused by Genomic Duplications Involving the Long Range, Limb-Specific Shh Enhancer." *Journal of Medical Genetics* 45 (9): 589–95.

Swindler, Daris Ray, and Charles D. Wood. 1973. *An Atlas of Primate Gross Anatomy: Baboon, Chimpanzee, and Man.* Seattle: University of Washington Press.

Szymaniak, Aleksander D, John E Mahoney, Wellington V Cardoso, and Xaralabos Varelas. 2015. "Crumbs3-Mediated Polarity Directs Airway Epithelial Cell Fate Through the Hippo Pathway Effector Yap." *Developmental Cell* 34 (3). Elsevier: 283–96.

Tai, P. W., H. Wu, J. A. Gordon, and others. 2014. "Epigenetic Landscape During Osteoblastogenesis Defines a Differentiation-Dependent Runx2 Promoter Region." *Gene* 550 (1): 1–9.

Takeuchi, J. K., K. Koshiba-Takeuchi, K. Matsumoto, and others. 1999. "Tbx5 and Tbx4 Genes Determine the Wing/Leg Identity of Limb Buds." *Nature* 398 (6730): 810–14.

Tan, Qi, Pauline Po Yee Lui, and Yuk Wa Lee. 2013. "In Vivo Identity of Tendon Stem

Cells and the Roles of Stem Cells in Tendon Healing." *Stem Cells and Development* 22 (23). Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA: 3128–40.

Tapscott, Stephen. 2005. "The Circuitry of a Master Switch: Myod and the Regulation of Skeletal Muscle Gene Transcription." *Development* 132 (12): 2685–95. doi:10.1242/dev.01874.

Tetreault, M., E. Bareke, J. Nadaf, N. Alirezaie, and J. Majewski. 2015. "Whole-Exome Sequencing as a Diagnostic Tool: Current Challenges and Future Opportunities." *Expert Review of Molecular Diagnostics* 15 (6): 749–60.

Tickle, C., G. Shellswell, A. Crawley, and L. Wolpert. 1976. "Positional Signalling by Mouse Limb Polarising Region in the Chick Wing Bud." *Nature* 259 (5542): 396–97.

Timmons, James, Kristian Wennmalm, Ola Larsson, Tomas Walden, Timo Lassmann, Natasa Petrovic, D. Hamilton, et al. 2007. "Myogenic Gene Expression Signature Establishes That Brown and White Adipocytes Originate from Distinct Cell Lineages." *Proc National Acad Sci* 104 (11): 4401–6. doi:10.1073/pnas.0610615104.

Trella, K. J., J. Li, E. Stylianou, V. M. Wang, J. M. Frank, J. Galante, J. D. Sandy, A. Plaas, and R. Wysocki. 2017. "Genome-wide Analysis Identifies Differential Promoter Methylation of Leprel2, Foxf1, Mmp25, Igfbp6, and Peg12 in Murine Tendinopathy." *Journal of Orthopaedic Research* 35: 947–55.

Tuli, Richard, Suraj Tuli, Sumon Nandi, Xiaoxue Huang, Paul Manner, William Hozack, Keith Danielson, David Hall, and Rocky Tuan. 2003. "Transforming Growth Factor-B-Mediated Chondrogenesis of Human Mesenchymal Progenitor Cells Involves N-Cadherin and Mitogen-Activated Protein Kinase and Wnt Signaling Cross-Talk." *J Biol Chem* 278 (42): 41227–36. doi:10.1074/jbc.M305312200.

Tung, J. W., K. Heydari, R. Tirouvanziam, and others. 2007. "Modern Flow

Cytometry: A Practical Approach." *Clinical Laboratory Medicine* 27 (3): 453–68.

Urbanowicz, Maria, and Prejzner-Morawska, Alicja. 1972. "The Triceps Surae Muscle in Chimpanzees" XXXI (4): 433–40.

Urnov, F. D., E. J. Rebar, M. C. Holmes, H. S. Zhang, and P. D. Gregory. 2010. "Genome Editing with Engineered Zinc Finger Nucleases." *Nature Reviews Genetics* 11 (9): 636–46.

Ustanina, Svetlana, Jaime Carvajal, Peter Rigby, and Thomas Braun. 2007. "The Myogenic Factor Myf5 Supports Efficient Skeletal Muscle Regeneration by Enabling Transient Myoblast Amplification." *Stem Cells* 25 (8): 2006–16. doi:10.1634/stemcells.2006-0736.

Valasek, P., D. J. Evans, F. Maina, M. Grim, and K. Patel. 2005. "A Dual Fate of the Hindlimb Muscle Mass: Cloacal/Perineal Musculature Develops from Leg Muscle Cells." *Development* 132 (3): 447–58.

Vargesson, N., J. D. Clarke, K. Vincent, and others. 1997. "Cell Fate in the Chick Limb Bud and Relationship to Gene Expression." *Development* 124 (10): 1909–18.

Vaux, D. L., F. Fidler, and G. Cumming. 2012. "Replicates and Repeats—what Is the Difference and Is It Significant?" *EMBO Reports* 13: 291–96.

Visel, A., M. J. Blow, Z. Li, and others. 2009. "ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers." *Nature* 457 (7231): 854–58.

Vitti, J. J., S. R. Grossman, and P. C. Sabeti. 2013. "Detecting Natural Selection in Genomic Data." *Annual Review of Genetics* 47: 97–120.

Vogel, Kathryn G, M Paulsson, and D Heinegård. 1984. "Specific Inhibition of Type I and Type Ii Collagen Fibrillogenesis by the Small Proteoglycan of Tendon."

Biochemical Journal 223 (3). Portland Press Limited: 587–97.

Voleti, P. B., M. R. Buckley, and L. J. n. Soslowsky. 2012. "Tendon Healing: Repair and Regeneratio." *Annu Rev Biomed Eng* 14: 47–71.

Voronova, Anastassia, Erin Coyne, Ashraf Al Madhoun, Joel V Fair, Neven Bosiljcic, Catherine St-Louis, Grace Li, et al. 2013. "Hedgehog Signaling Regulates Myod Expression and Activity." *Journal of Biological Chemistry* 288 (6). ASBMB: 4389–4404.

Vortkamp, A., K. Lee, B. Lanske, and others. 1996. "Regulation of Rate of Cartilage Differentiation by Indian Hedgehog and Pth-Related Protein." *Science* 273 (5275): 613–22.

Wan, H., C. Feng, F. Teng, and others. 2015. "One-Step Generation of P53 Gene Biallelic Mutant Cynomolgus Monkey via the Crispr/Cas System." *Cell Research* 25 (2): 258–61.

Wang, H., H. Yang, C. S. Shivalila, and others. 2013. "One-Step Generation of Mice Carrying Mutations in Multiple Genes by Crispr/Cas-Mediated Genome Engineering." *Cell* 153 (4): 910–18.

Wang, Y., F. Middleton, J. A. Horton, and others. 2004. "Microarray Analysis of Proliferative and Hypertrophic Growth Plate Zones Identifies Differentiation Markers and Signal Pathways." *Bone* 35 (6): 1273–93.

Wang, Y., X. Zhang, H. Huang, Y. Xia, Y. Yao, Mak Af-t, Yung Ps-h, Chan K-m, L. Wang, and C. Zhang. 2017. "Osteocalcin Expressing Cells from Tendon Sheaths in Mice Contribute to Tendon Repair by Activating Hedgehog Signaling." *ELife* 6.

Wang, Yi, Xu Zhang, Huihui Huang, Yin Xia, YiFei Yao, Arthur Fuk-Tat Mak, Patrick Shu-Hang Yung, et al. 2017. "Osteocalcin Expressing Cells from Tendon Sheaths in Mice Contribute to Tendon Repair by Activating Hedgehog Signaling." *ELife* 6. eLife

Sciences Publications Limited: e30474.

Wareing, Katy Ann. 2016. “Adaptation of the Non-Human Great Ape Lower Limb in Response to Locomotor Behaviour.”

Watkins-Chow, D. E., and Pavan WJ. 2008. “Genomic Copy Number and Expression Variation Within the C57bl/6J Inbred Mouse Strain.” *Genome Research* 18: 60–66.

Weir, B. S., L. R. Cardon, A. D. Anderson, D. M. Nielsen, and W. G. Hill. 2005. “Measures of Human Population Structure Show Heterogeneity Among Genomic Regions.” *Genome Research* 15 (11): 1468–76.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2017. *Tidyverse: Easily Install and Load the ’Tidyverse’*. https://CRAN.R-project.org/package=tidyverse.

Widdig, A., M. J. Kessler, F. B. Bercovitch, and others. 2016. “Genetic Studies on the Cayo Santi- Ago Rhesus Macaques: A Review of 40 Years of Research.” *American Journal of Primatology* 78 (1): 44–62.

Wood, A. R., T. Esko, J. Yang, and others. 2014. “Defining the Role of Common Variation in the Genomic and Biological Architecture of Adult Human Height.” *Nature Genetics* 46 (11): 1173–86.

Wu, C., P. M. Bingham, K. J. Livak, R. Holmgren, and S. C. Elgin. 1979. “The Chromatin Structure of Specific Genes: Evidence for Higher Order Domains of Defined Dna Sequence.” *Cell* 16 (4): 797–806.

Wu, S., and F. De Luca. 2006. “Inhibition of the Proteasomal Function in Chondrocytes down-Regulates Growth Plate Chondrogenesis and Longitudinal Bone Growth.”

Endocrinology 147 (8): 3761–8.

Xin, Mei, Yuri Kim, Lillian Sutherland, Masao Murakami, Xiaoxia Qi, McAnally, John, Enzo Porrello, et al. 2013. "Hippo Pathway Effector Yap Promotes Cardiac Regeneration." *Proc National Acad Sci* 110 (34): 13839–44. doi:10.1073/pnas.1313192110.

Xue, Y., J. Prado-Martinez, P. H. Sudmant, and others. 2015. "Mountain Gorilla Genomes Reveal the Impact of Long-Term Population Decline and Inbreeding." *Science* 348 (6231): 242–45.

Yamamoto, Masakazu, Nicholas Legendre, Arpita Biswas, Alexander Lawton, Shoko Yamamoto, Shahragim Tajbakhsh, Gabrielle Kardon, and David Goldhamer. 2018. "Loss of MyoD and Myf5 in Skeletal Muscle Stem Cells Results in Altered Myogenic Programming and Failed Regeneration." *Stem Cell Rep* 10 (3): 956–69. doi:10.1016/j.stemcr.2018.01.027.

Yang, Y., G. Drossopoulou, P. T. Chuang, and others. 1997. "Relationship Between Dose, Distance and Time in Sonic Hedgehog-Mediated Regulation of Anteroposterior Polarity in the Chick Limb." *Development* 124 (21): 4393–4404.

Ye, Jian, George Coulouris, Irena Zaretskaya, Ioana Cutcutache, Steve Rozen, and Thomas L Madden. 2012. "Primer-Blast: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction." *BMC Bioinformatics* 13 (1). BioMed Central: 134.

Yin, Zi, Jia Guo, Tian-yi Wu, Xiao Chen, Liang-liang Xu, Si-en Lin, Yun-xin Sun, Kai-Ming Chan, Hongwei Ouyang, and Gang Li. 2016. "Stepwise Differentiation of Mesenchymal Stem Cells Augments Tendon-Like Tissue Formation and Defect Repair in Vivo." *Stem Cells Translational Medicine* 5 (8). Wiley Online Library: 1106–16.

Yoon, J, and J Halper. 2005. "Tendon Proteoglycans: Biochemistry and Function." *J Musculoskelet Neuronal Interactions* 5 (1): 22–34.

Young, N. M. 2006. "Function, Ontogeny and Canalization of Shape Variance in the

Primate Scapula." *Journal of Anatomy* 209 (5): 623–36.

Yu, Guangchuang, Li-Gen Wang, and Qing-Yu He. 2015. "ChIPseeker: An R/Bioconductor Package for Chip Peak Annotation, Comparison and Visualization." *Bioinformatics* 31 (14): 2382–3. doi:10.1093/bioinformatics/btv145.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." *OMICS: A Journal of Integrative Biology* 16 (5): 284–87. doi:10.1089/omi.2011.0118.

Yue, F., Y. Cheng, A. Breschi, and others. 2014. "A Comparative Encyclopedia of Dna Elements in the Mouse Genome." *Nature* 515 (7527): 355–64.

Zammit, Peter, Jaime Carvajal, Jon Golding, Jennifer Morgan, Dennis Summerbell, Joseph Zolnerciks, Terence Partridge, Peter Rigby, and Jonathan Beauchamp. 2004. "Myf5 Expression in Satellite Cells and Spindles in Adult Muscle Is Controlled by Separate Genetic Elements." *Dev Biol* 273 (2): 454–65. doi:10.1016/j.ydbio.2004.05.038.

Zhang, F., Y. Wen, and X. Guo. 2014. "CRISPR/Cas9 for Genome Editing: Progress, Implications and Challenges." *Human Molecular Genetics* 23 (R1): R40–46.

Zhang, G, BB Young, Y Ezura, M Favata, LJ Soslowsky, S Chakravarti, and DE Birk. 2005. "Development of Tendon Structure and Function: Regulation of Collagen Fibrillogenesis." *J Musculoskelet Neuronal Interact* 5 (1): 5–21.

Zhang, Jianying, and James Wang. 2013. "The Effects of Mechanical Loading on Tendons - an in Vivo and in Vitro Model Study." *Plos One* 8 (8): e71740. doi:10.1371/journal.pone.0071740.

Zhang, M., M. R. Pritchard, F. A. Middleton, J. A. Horton, and T. A. Damron. 2008. "Microarray Analysis of Perichondral and Reserve Growth Plate Zones Identifies

Differential Gene Expressions and Signal Pathways." *Bone* 43 (3): 511–20.

Zhang, Ming, Meina Wang, Xiaohong Tan, Tian-Fang Li, Ying E Zhang, and Di Chen. 2010. "Smad3 Prevents B-Catenin Degradation and Facilitates B-Catenin Nuclear Translocation in Chondrocytes." *Journal of Biological Chemistry* 285 (12). ASBMB: 8703–10.

Zhang, Yong, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of Chip-Seq (Macs)." *Genome Biology* 9 (9). BioMed Central: R137.

Zhao, Bin, Li Li, and Kun-Liang Guan. 2010. "Hippo Signaling at a Glance." *J Cell Sci* 123 (23): 4001–6. doi:10.1242/jcs.069070.

Zhao, Bin, Li Li, Qunying Lei, and Kun-Liang Guan. 2010. "The Hippo–YAP Pathway in Organ Size Control and Tumorigenesis: An Updated Version." *Genes and Development* 24 (9). Cold Spring Harbor Lab: 862–74.

Zhao, Z., G. Tavoosidana, M. Sjolinder, and others. 2006. "Circular Chromosome Conformation Capture (4C) Uncovers Extensive Networks of Epigenetically Regulated Intra- and Interchromosomal Interactions." *Nature Genetics* 38 (11): 1341–7.

Zhou, H. Y., Y. Katsman, N. K. Dhaliwal, and others. 2014. "A Sox2 Distal Enhancer Cluster Regulates Embryonic Stem Cell Differentiation Potential." *Genes and Development* 28 (24): 2699–2711.

Zhou, Shuanhu. 2011. "TGF-B Regulates B-Catenin Signaling and Osteoblast Differentiation in Human Mesenchymal Stem Cells." *Journal of Cellular Biochemistry* 112 (6). Wiley Online Library: 1651–60.

Zhu, Xuming, Huang Zhu, Lingling Zhang, Sixia Huang, Jingjing Cao, Gang Ma, Guoying Feng, Lin He, Yingzi Yang, and Xizhi Guo. 2012. "Wls-Mediated Wnts Differentially Regulate Distal Limb Patterning and Tissue Morphogenesis." *Dev Biol*

365 (2): 328–38. doi:10.1016/j.ydbio.2012.02.019.

Zou, H., R. Wieser, J. Massague, and L. Niswander. 1997. “Distinct Roles of Type I Bone Morphogenetic Protein Receptors in the Formation and Differentiation of Cartilage.” *Genes and Development* 11 (17): 2191–2203.