1-1-2013

Multiple Conserved Enhancers of the Osteoblast Master Transcription Factor, Runx2, Integrate Diverse Signaling Pathways to Direct Expression to Developing Bone

Christopher Weber

University of Pennsylvania, topher.weber@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Genetics Commons, and the Molecular Biology Commons

Recommended Citation

Weber, Christopher, "Multiple Conserved Enhancers of the Osteoblast Master Transcription Factor, Runx2, Integrate Diverse Signaling Pathways to Direct Expression to Developing Bone" (2013). Publicly Accessible Penn Dissertations. 941. http://repository.upenn.edu/edissertations/941

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/941
For more information, please contact libraryrepository@pobox.upenn.edu.
Multiple Conserved Enhancers of the Osteoblast Master Transcription Factor, Runx2, Integrate Diverse Signaling Pathways to Direct Expression to Developing Bone

Abstract
The vertebrate skeleton forms via two distinct modes of ossification, membranous and endochondral. Osteoblasts are also heterogeneous in embryonic origin; bone formed by either mode can be derived from neural crest cells or mesoderm. In contrast, all bone develops via a common genetic pathway regulated by the transcription factor Runx2. Runx2 is required for bone formation, and haploinsufficiency in humans causes the skeletal syndrome cleidocranial dysplasia, demonstrating the importance of gene dosage. Despite the central role of Runx2 in directing bone formation, little is understood about how its expression is regulated in development. We took an unbiased approach to identify direct regulatory inputs into Runx2 transcription by identifying cis-regulatory elements associated with the human gene. We assayed conserved non-coding elements in a 1 Mb interval surrounding the gene for their ability to direct osteoblast expression in transgenic zebrafish. We identified three enhancers spaced out across the interval. Within each we identified conserved transcription factor binding sites required for their activity, and further showed distinct and specific regulation of each. The enhancer in the last intron of RUNX2 itself is positively regulated by the FGF signaling pathway, an enhancer in the last intron of the adjacent gene, SUPT3H, is regulated by canonical Wnt signaling, and a distant downstream enhancer requires a conserved Dlx binding site for its activity. While all of these pathways and factors have been previously implicated in bone formation, our results provide the first direct links to the common genetic pathway regulating osteogenesis, transcription of Runx2. These findings further illustrate the integration of multiple regulatory inputs at the level of transcription of a key developmental gene, and highlight the role of Runx2 as the gatekeeper for changes in skeletal morphology achieved through alterations in gene expression.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Shannon Fisher

Keywords
BMP, FGF, osteoblast, RUNX2, WNT, zebrafish

Subject Categories
Genetics | Molecular Biology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/941
MULTIPLE CONSERVED ENHANCERS OF THE OSTEOBLAST MASTER TRANSCRIPTION FACTOR, RUNX2, INTEGRATE DIVERSE SIGNALING PATHWAYS TO DIRECT EXPRESSION TO DEVELOPING BONE

Christopher William Weber

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

______________________________
Shannon Fisher, M.D, Ph.D.
Assistant Professor of Cell and Developmental Biology
Graduate Group Chairperson

______________________________
Daniel Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee

Douglas Epstein, Ph.D., Associate Professor of Genetics
Mary Mullins, Ph.D., Professor of Cell and Developmental Biology
Kurt Hankenson, M.S., D.V.M., Ph.D., Associate Professor of Equine Musculoskeletal Research
Eileen Shore, Ph.D., Cali and Weldon Research Professor in FOP
ACKNOWLEDGMENTS

Despite my peripatetic and inefficient path to actually sitting down and writing this thing, I’ve been surrounded constantly by an embarrassment of diversely sourced support and expertise that’s kept my big dumb head directed towards this goal. I’m going to dispense with the prose in favor of bullet points now; please keep in mind that the choice is motivated by sloth and not style.

To my mother: Susan: You believe in a lot of weird shit, but you also believe in me. That means a lot, even if you still think crystals were sent from outer space to positively regulate your mood and that it’s a good and useful exercise to pay a woman to mindmeld with your horses over the phone.

To my advisor, Shannon Fisher: Your support has been intellectual, material, patient, culinary and indispensible. Thank you. I don’t know what I did to deserve it.

To my brothers, Brian and Andy: You are the funniest people I know, and you’ve never made enough fun of me to satisfy my need for self-flagellation. I know there’s time.

To my thesis committee, Doug Epstein, Mary Mullins, Kurt Hankenson and Eileen Shore: Thank you for letting me fail in the privacy of your company, so that I could moderately succeed outside of it.

To other members of the Fisher lab, past and present, Thanks for showing me where we keep that. Also, I’m the one who left the scope on overnight and didn’t sign in. Gui, I’ve been mispronouncing your name for four years. Many apologies.
To my graduate school colleagues and contemporaries, in particular Jessica Taaffe, Austin Thiel, and Dustin Hancks, and: Let’s never do this again.

To the taxpayers of the Commonwealth of Pennsylvania and the United States of America: Thanks for subsidizing the last twelve years. I really wish I could promise you that it was money better spent than on a crumbling infrastructure, massive foreign debt, or the unfunded liabilities of Social Security and Medicare as the country undergoes a massive demographic shift. We’ll see. I gotta finish school first.

“Personally, I liked the university. They gave us money and facilities; we didn't have to produce anything! You've never been out of college! You don't know what it's like out there! I've worked in the private sector. They expect results.”

-Raymond Stantz, Ph.D

“Deformed, unfinish'd, sent before my time
Into this breathing world, scarce half made up”

--Richard III, Act 1, scene i
MULTIPLE CONSERVED ENHANCERS OF THE OSTEOBLAST MASTER TRANSCRIPTION FACTOR, RUNX2, INTEGRATE DIVERSE SIGNALING PATHWAYS TO DIRECT EXPRESSION TO DEVELOPING BONE

Christopher William Weber

Shannon Fisher

The vertebrate skeleton forms via two distinct modes of ossification, membranous and endochondral. Osteoblasts are also heterogeneous in embryonic origin; bone formed by either mode can be derived from neural crest cells or mesoderm. In contrast, all bone develops via a common genetic pathway regulated by the transcription factor Runx2. Runx2 is required for bone formation, and haploinsufficiency in humans causes the skeletal syndrome cleidocranial dysplasia, demonstrating the importance of gene dosage. Despite the central role of Runx2 in directing bone formation, little is understood about how its expression is regulated in development. We took an unbiased approach to identify direct regulatory inputs into Runx2 transcription by identifying cis–regulatory elements associated with the human gene. We assayed conserved non-coding elements in a 1 Mb interval surrounding the gene for their ability to direct osteoblast expression in transgenic zebrafish. We identified three enhancers spaced out across the interval. Within each we identified conserved transcription factor binding sites required for their activity, and further showed
distinct and specific regulation of each. The enhancer in the last intron of
RUNX2 itself is positively regulated by the FGF signaling pathway, an enhancer
in the last intron of the adjacent gene, SUPT3H, is regulated by canonical Wnt
signaling, and a distant downstream enhancer requires a conserved Dlx binding
site for its activity. While all of these pathways and factors have been previously
implicated in bone formation, our results provide the first direct links to the
common genetic pathway regulating osteogenesis, transcription of Runx2. These
findings further illustrate the integration of multiple regulatory inputs at the
level of transcription of a key developmental gene, and highlight the role of
Runx2 as the gatekeeper for changes in skeletal morphology achieved through
alterations in gene expression.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .............................................................................................................. ii

ABSTRACT ................................................................................................................................. iv

LIST OF TABLES ...................................................................................................................... viii

LIST OF FIGURES .................................................................................................................... ix

CHAPTER 1 ................................................................................................................................. 1

Characteristics of the vertebrate skeleton .............................................................................. 1
Embryonic origins of the vertebrate skeleton .......................................................................... 2
Genetic origins of the vertebrate skeleton – Runx2 ................................................................. 4
Genetic origins of the vertebrate skeleton – Sp7/osx ............................................................... 5
Molecular signaling and the vertebrate skeleton – The BMP pathway .................................. 6
Molecular signaling and the vertebrate skeleton – The Wnt pathway .................................... 10
Molecular signaling and the vertebrate skeleton – The FGF pathway .................................. 13
Genetic origins of the vertebrate skeleton – Signaling crosstalk ......................................... 15
Signaling crosstalk in the vertebrate skeleton - BMP and Wnt pathways .............................. 16
Signaling crosstalk in the vertebrate skeleton - BMP and FGF pathways .............................. 17
Signaling crosstalk in the vertebrate skeleton – Wnt and FGF pathways ............................. 18
Zebrafish as a model to study vertebrate skeletogenesis ......................................................... 19
Study aims ................................................................................................................................... 20

CHAPTER 2 ................................................................................................................................ 22

Introduction ............................................................................................................................... 23
Materials and Methods...............................................................................................................29
Results........................................................................................................................................31

CHAPTER 3....................................................................................................................................55
Introduction.................................................................................................................................56
Materials and Methods..............................................................................................................56
Results..........................................................................................................................................58

CHAPTER 4....................................................................................................................................83
Summary of human RUNX2 associated enhancers.................................................................84
Summary of human RUNX2 associated enhancer activity -- +210RUNX2 directs expression to osteoblasts separably through FGF signaling and Runx2 autoregulation.................................................................................................................................86
Summary of human RUNX2 associated enhancers -- +542RUNX2 directs expression to early osteoblasts..................................................................................................................................................87
Summary of human RUNX2 associated enhancers -460RUNX2 potentially links Wnt signaling, Runx2 regulation and variation in common skeletal phenotypes and diseases........................................................................................................................................88
Runx2 expression modulation as a source of evolutionary skeletal diversity.................................89
Future Directions.........................................................................................................................91
Conclusion....................................................................................................................................93
LIST OF TABLES

Table 2.1. Sequences at \textit{RUNX2} locus tested for enhancer activity...........................................50

Table 2.2. Sequences at \textit{the runx2a} locus tested for enhancer activity...........................................52

Table 2.3. Sequences at \textit{the runx2b} locus tested for enhancer activity.......................................... 54

Table 3.1. Sequences of +542\textit{RUNX2} tested for enhancer activity..................................................82

Table 4.1 – SNPs associated with human skeletal phenotypes in the human \textit{RUNX2} locus...........102
LIST OF FIGURES

Figure 2.1. – Broad distribution of osteoblast specific enhancers at the RUNX2 locus .................37

Figure 2.2 -- Expression patterns of three human RUNX2 associated enhancers .........................39

Figure 2.3 -- Differential expression initiation times in the cleithrum anlagen .........................41

Figure 2.4 – Distinct expression domains in the developing opercle ........................................42

Figure 2.5 – +154runx2a, a conserved ortholog of +210RUNX2, directs expression to the branchial arches and osteoblasts. ..................................................................................................................44

Figure 2.6 -- +542RUNX2 directs expression to the developing vertebral arches ......................45

Figure 2.7 – Analysis of chromatin environment at three RUNX2 associated enhancers in a normal human osteoblast cell line. ..................................................................................................................46

Figure 2.8 – Screen for skeletal enhancers associated with the zebrafish Runx2 orthologs, runx2a and runx2b ........................................................................................................................................48

Figure 3.1 -- Deep linear conservation between +210RUNX2 and other vertebrate orthologs ........................................................................................................................................64

Figure 3.2 – The RUNX2 binding site mediates +210RUNX2 directed bone expression ..........66

Figure 3.3 – The ETS binding sites mediate branchial arch expression ........................................68

Figure 3.4 – The POU binding site is not essential for +210RUNX2 activity during embryogenesis ........................................................................................................................................70

Figure 3.5 – +210RUNX2 is regulated by the FGF signaling pathway ........................................72
Figure 3.6 – Subcloning of +542RUNX2 localizes osteoblast activity to a 433 bp fragment.

Figure 3.7 – Identification and functional testing of conserved transcription factor binding sites in +542RUNX2.

Figure 3.8 – -460RUNX2 mediated expression is responsive to Wnt mediated signaling.

Figure 3.9 – -460RUNX2 regulatory competency requires two conserved TCF/LEF binding sites.

Figure 4.1 Model for integration of multiple signaling inputs at the Runx2 locus.

Figure 4.2 – SNPs associated with skeletal phenotypes and disorders cluster near the Wnt responsive enhancer -460RUNX2.

Figure 4.3 – Recent positive selection in the human lineage near +210RUNX2.
CHAPTER 1

INTRODUCTION
CHAPTER 1

Characteristics of the vertebrate skeleton

The presence of a mineralized endoskeleton is one of the common features of the vertebrate lineage\(^1\). In addition to its historically understood roles in support and as the sites of muscle attachment, the skeleton has more recently been understood to be the site of hematopoiesis\(^2\), endocrine regulation of glucose metabolism\(^3\), a reservoir for inorganic minerals\(^4\) and critical in male reproductive function\(^5\).

The vertebrate skeleton is chiefly composed of two tissue types: bone and cartilage.\(^a\) Cartilage is the more evolutionarily primary of the two\(^6\). While not possessing a ‘true’ skeleton, the chordate amphioxus expresses orthologs of cartilage marker genes in the nascent notochord\(^7\). Cartilage is composed of chondrocytes suspended in a rigid matrix rich in collagen fibrillar proteins and acidic polysaccharides\(^8\). The most abundant of these proteins are type II collagen and aggrecan, whose negative charge accounts for the osmotic swelling of the tissue\(^9\), resulting in the familiar rigid plasticity of the material. This property confers a biomechanical role in the fully realized skeleton, allowing articular surfaces of joints to tolerate compressive forces.

Conversely, bone is vascularized, has a higher metabolic activity and differs in its extracellular matrix (ECM) composition both in the content of secreted proteins, but also in the presence of inorganic calcium\(^10\). Unlike the

\(^a\) Two other tissue types found exclusively in teeth are dentin and enamel, though they will not be discussed further in this document.
cartilage ECM, 90% of the total dry protein weight is type I collagen. Collagen I forms an extensively crosslinked fiber, around which calcium crystals in the form of spindles of hydroxyapatite are deposited, resulting in the characteristic rigidity of bone tissue\(^{11}\). Other well-characterized components include alkaline phosphatase, osteopontin, and osteocalcin\(^{12}\).

Osteoblasts are the cells responsible for the deposition of this defining matrix. Correspondingly, they have a highly basophilic cytoplasm and extensive endoplasmic reticulum and Golgi apparatus\(^{13}\) to produce substantial amounts of secreted protein. Following matrix deposition, osteoblasts either become lining cells or remain embedded in bone, the latter defined as osteocytes\(^{14}\). These cells account for 95% of mature bone tissue. Another bone cell type, osteoclasts, arises from the monocytic/macrophage lineage postnatally\(^{15}\). These cells have a resorbative role in bone homeostasis and therefore regulate bone mass density.

**Embryonic origins of the vertebrate skeleton**

Skeletogenesis describes the process by which mesenchymal stem cells (MSCs) differentiate into osteoblasts and chondrocytes in a defined program. MSCs are loose, multipotent cells with the capacity to differentiate into non-skeletal cell types such as adipocytes or myocytes\(^{16}\). Whether commitment to the skeletal lineage involves the existence of a bipotential skeletal precursor cell type, capable of adopting a bone or cartilage fate, is at issue in the literature\(^{17}\). Skeletal elements in the embryo forms via two distinct processes. Intramembrous ossification describes direct condensation of migrated MSCs and
subsequent transformation to bone. This process is employed in the creation of the flat bones of the skull as well as fracture repair. Elsewhere, particularly in the long bones, endochondral ossification results in the calcification and invasion of a cartilaginous scaffold by osteoprogenitor cells. A complementary heterogeneity is observed in the embryonic origin of MSCs, where neural crest, lateral plate, and somitic mesoderm all contribute to the developing skeleton.

**Genetic origins of the vertebrate skeleton – Runx2**

However, this diversity contrasts with the uniform genetic origin of skeletal tissues. Commitment to the osteoblast lineage requires the expression of the early marker gene and runt domain containing transcription factor Runx2. The runt domain is a site of protein-protein interaction, as well as binding to the core sequence 5’- PyGPyGGTPy-3’. Runx2−/− mice fail to generate any osteoblasts, and chondrocyte maturation and terminal differentiation are disturbed. Additionally, haploinsufficiency at the locus causes the skeletal disorder cleidocranial dysplasia, marked by delayed closure of the fontanelles of the skull, hypoplasticity of the clavicle, and other features (OMIM# 119600). RUNX2 binds to and upregulates other osteoblast marker genes, which are also upregulated following forced expression of Runx2 in non-skeletal tissues, including fibroblasts, C3H10T1/2 cells, primary myoblasts, and marrow stromal cells. For these reasons, Runx2 has been recognized as occupying an indispensable bottleneck position in the osteoblast fate switch and is often referred to as the master regulator of osteoblast development.
The dosage of Runx2 must be finely tuned in order to properly execute differentiation. Overexpression of Runx2 in osteoblasts arrests bone development in a mouse model, resulting in an osteopenic phenotype. Forced expression in chondrocytes produces precociously mature cells that produce osteoid tissue and bone marrow not present in orthologous structures in wild type animals. Despite Runx2’s unquestioned indispensability early in fate commitment, the notion that Runx2 might not have a role in mature osteoblasts has been proposed, as expression of a dominant negative form of the protein exclusively in mature osteoblasts does not affect transcription of the osteoblast marker osteocalcin, a gene that can be activated by forced expression of Runx2 in non osteoblastic cells.

Genetic origins of the vertebrate skeleton – sp7/osx

An answer to potential regulators of later osteoblast differentiation came with the identification of Sp7/Osx as a cDNA species specifically expressed in C2C12 cells undergoing osteoblastogenesis. Sp7 codes for a zinc finger-containing transcription factor from the Kruppel-like factor family. As with Runx2, inactivation of in mouse models yielded a skeleton devoid of osteoblasts; however, mineralization did occur in bones formed by endochondral ossification, though the features of those tissues were more akin to a mineralized form of cartilage. Interestingly, Runx2 expression levels were unaffected, indicating that Sp7 is not upstream of Runx2. Further work located Sp7 as a direct target of
Runx2. Sp7 is recruited to its own promoter in murine UMR106-01 osteosarcoma cells to the exclusion of other members of the sp transcription factor family in a manner that correlates with the expression of sp7.

Sp7 is thought to function exclusively in later osteoblast differentiation and distinct from the activities of Runx2 in cartilage. Among Sp7's target genes are the bone marker genes Col1a1, Bsp and Ocn. The regulation of Col1a1 by Sp7 is corroborated clinically by a report of a proband presenting with osteogenesis imperfecta and a frameshift mutation within the SP7 coding region. Finally, in contrast to Runx2, Sp7 function appears to be critical for postnatal growth and maintenance of bone.

Molecular signaling and the vertebrate skeleton – The BMP pathway

Bone morphogenetic proteins (BMPs) were initially identified on the basis of their ability to induce de novo bone and cartilage formation in vivo. Most BMPs are members of the transforming growth factor-β superfamily of proteins with important roles in both proper patterning and differentiation of the skeleton. Canonically, signaling starts upon BMP ligand binding to heteromeric cell surface receptors composed of BMPR-I and BMPR-II receptors. This activated complex phosphorylates cytoplasmic SMAD proteins via a serine-threonine kinase domain. SMADs 1,5 and 8 bind to a co-SMAD upon phosphorylation and enter the nucleus to directly affect gene transcription via

\[ \text{Allen, unpublished observation} \]
\[ \text{Notably, BMP-1 is a metalloproteinase.} \]
chromatin binding. As will be discussed in future sections, the Smad proteins offer a context for crosstalk with other signaling pathways.

Understanding the role of BMP signaling in skeletogenesis is complicated by the presence of multiple components with distinct, yet overlapping, activities, as well the necessity of BMP signaling in early embryo patterning formation. BMP-2, --4,--6,--7 are ligands with demonstrable osteogenic potential in vitro, yet genetic studies using conditional knockout alleles reveal more subtle and complementary roles. Both Bmp2 and Bmp4 activities are dispensable for the formation of the long bones, of the limbs, though deletion of the former results in an increase of fractures postnatally\(^{41,42}\). Similarly, loss of Bmpr2\(^{43}\) and Bmp7 have no demonstrable effect on bone formation or fracture repair in the limbs\(^{44}\). However, a double knockout of Bmp2 and Bmp4 results in a severe impairment of osteogenesis, indicating a redundancy in these roles\(^{45}\).

Runx2 upregulation has been observed in in vitro systems following BMP stimulation\(^{46,47}\), and consequently, Runx2 is thought to be the principle mediator of downstream BMP actions\(^{48}\). However, there are also thought to be BMP signals capable of driving osteoblastogenesis independently of Runx2. Although BMP-2 administration is not capable of driving full differentiation of osteoblasts and chondroblasts in Runx2-deficient mouse calvarial cell lines, upregulation of alkaline phosphatase, osteocalcin and sp7 is detectable\(^{22,49,50}\). BMP-2 treatment upregulates sp7 expression in C2C12 cells independently of Runx2\(^{51}\). Also, preosteoblastic cell lines require autocrine BMP signaling for proper
differentiation, although they already express Runx2\textsuperscript{52,53}. Finally, activated SMAD proteins interact physically and functionally with RUNX2, suggesting a synergistic relationship to complement the Runx2 dependent and independent BMP signalling axes. SMAD1 and RUNX2 transcription factors complex to drive gene expression on target gene promoters\textsuperscript{54}. An osteoblast specific deletion of Smad1 causes an osteopenic phenotype\textsuperscript{55}, and combined deletion of Smad1/5/8 results in severe chondrodysplasia\textsuperscript{56}.

Pretreatment with the ribosome inhibitor cyclohexamide prior to BMP-2 treatment blocks the induction of Runx2\textsuperscript{57} and sp7\textsuperscript{58}, indicating the need for the synthesis of an intermediate protein to complete the signaling axis. Among the direct targets of BMP signaling with known roles in skeletogenesis are homeodomain proteins. In particular, microarray experiments examining the transcriptional response to BMP-2 treatment in cultured C2C12 osteoprogenitor cells have identified members of the meshless(Msx), distalless(Dlx), and aristaless(Alx) transcription factor families as being immediately and transiently induced, prior to the commitment to osteogenesis evidenced by expression of Runx2\textsuperscript{59-61}. Mutations associated with the Msx1 and Msx2 loci demonstrate consequences in skeletal patterning and differentiation. Msx1\textsuperscript{−/−} mice exhibit craniofacial and tooth development abnormalities including a cleft palate phenotype\textsuperscript{62}, while Msx2\textsuperscript{−/−} mice possess delayed calvarial bone growth, defects in endochondrial ossification and chondrogenesis, as well as reduced expression of osteocalcin and Runx2\textsuperscript{63}. Simultaneous deletion of both Msx2 and Msx1
results in the complete absence of craniofacial bone\textsuperscript{63,64}. A reversal of this dosage effect is evidenced in a human MSX2 gain of function mutant with enhanced DNA binding, eliciting a premature fusion of the calvarial sutures and craniosynostosis\textsuperscript{65}. Significantly, microduplications upstream of MSX2 containing many conserved non-coding elements phenocopy CCD, suggesting a potentially rigid regulatory apparatus between BMP signaling and Runx2 in vivo\textsuperscript{66}.

In tetrapods, members of the Dlx gene family are grouped in binary clusters, facing each other via their 3' ends as a result of presumptive gene duplication events\textsuperscript{67}. Dlx1 and Dlx4 have important roles in tooth development\textsuperscript{68} and hematopoiesis\textsuperscript{69}, respectively, but they have not been identified as expressed in osteoblasts. Although Dlx3 inactivation results in embryonic lethality, it is expressed in osteoblastic lineage cells during endochondral ossification, and at its highest level in mature osteocalcin and Runx2 expressing osteoblasts\textsuperscript{70,71}. A 4bp frameshift deletion in the human DLX3 gene causes an autosomal dominant disease, tricho-dento-osseous syndrome (OMIM#600525), which is characterized by altered dermal bone formation in the skull as well as increased bone density\textsuperscript{72}. Consistent with this observation, interaction between DLX3 and RUNX2 reduces the capacity of RUNX2 to direct transcription at the osteocalcin promoter in a cell culture context\textsuperscript{73}.

Current opinion in the literature designates Dlx5 as a critical regulator of BMP mediated osteogenesis\textsuperscript{74}. Simultaneous knockout of the Dlx5\textsuperscript{5}/6 cluster
results in gross skeletal abnormalities, including absence of the calvaria, maxillary and mandibular bones, as well as a generalized ossification delay in the axial skeleton. These anomalies are also seen in Dlx5−/− mice; curiously, no data on a Dlx6−/− phenotype has been published. Dlx5 induction by BMPs has been observed in both the contexts of cell culture (MC3T3-E1 cells) and in vivo development (chick skull development), where Dlx5 expression is visible in proliferating suture mesenchyme not yet committed to an osteoblastic fate, suggesting a role in fate designation prior to Runx2 induction. Dlx5 induces Runx2 in immature calvaria mesenchyme culture, and Dlx5 and Runx2 have been shown to be recruited together at the stimulated promoters of induced osteoblast marker genes Alp and Ocn, so Dlx5 appears to possess a duality of roles during osteoblast differentiation, both as a direct regulator of Runx2 transcription, as well as a cooperative transcription factor at Runx2 target genes.

**Molecular signaling and the vertebrate skeleton – The Wnt pathway**

Wnts are secreted, lipid-modified glycoproteins that activate cell surface receptor-mediated signal transduction pathways to regulate a variety of cellular activities, including cell fate determination, proliferation, migration, polarity, and gene expression. In canonical, β-catenin-dependent WNT signaling, a Wnt ligand binds to a Frizzled receptor and their co-receptors low-density lipoprotein receptor-related protein 5 (LRP5) or LRP6 to stabilize cytosolic β-catenin via inhibition of a ubiquitinating complex. β-catenin then enters the nucleus and stimulates the transcription of WNT target genes by interacting
with lymphoid enhancer-binding factor 1 (LEF1), T cell factor 1 (TCF1), TCF3 or TCF4. Non-canonical Wnt pathways not utilizing β-catenin include the noncanonical planar cell polarity, which also does not employ LRP5 or LRP6, and noncanonical Wnt/calcium pathway, which requires modulation of intracellular calcium ion levels. These pathways are further separated by their choice of ligand; canonical Wnt signaling uses WNT1, WNT3a, WNT8 or WNT10b, while noncanonical signaling relies on WNT4, WNT5a or WNT11.

Recognition of the Wnt pathway’s involvement in bone biology began with a punctuation of discovery: in a single year, mutations causing severe alterations in bone density were identified in four groups of patients with bone mass disorders, pointing to the canonical branch of WNT signaling. Two of these mutations were detected in the LRP5 coreceptor necessary for Wnt signal propagation. Loss of function mutations in LRP5 cause the autosomal recessive disorder osteoporosis-pseudoglioma syndrome (OMIM# 259770) [81]. Affected individuals have very low bone mass and are prone to developing fractures and deformation, though they lack any identifiable defects in collagen synthesis, anabolic and catabolic hormones, calcium homeostasis, endochondral growth, or bone turnover. A knockout mouse model confirmed the genotype-phenotype relationship, and provided insight into the bone mass deficit. Lrp5−/− mice have low bone mass compared to their wild type littermates, though this feature was only detectable postnatally [82]. Intriguingly, no aberrations in Runx2 expression were detected in these mice.
The importance of the Wnt pathway in mediation of postnatal bone mass was further highlighted by the identification of gain of function mutations in LRP5 (G171V), causing an autosomal dominant high bone mass phenotype. Molecular investigations recognized the mutation as detrimental for the affinity of the protein for the extracellular Wnt signaling antagonists DKK1 and SOST. Sost itself is a locus for mutations affecting bone mass. Premature termination mutations in Sost cause sclerosteosis (OMIM #605740) whereas a 52 kb homozygous deletion downstream of the SOST gene is associated with van Buchem disease both of whom are characterized by bone overgrowth.

Genetic analysis in the decade following these initial discoveries detailed the importance of many additional canonical Wnt components (β-catenin, Gsk-3β, Axin2, and Dkk1; reviewed in ) in both osteoblast differentiation and postnatal bone mass density maintenance. Conditional deletion of β-catenin forces a chondrocytic fate on skeletal precursor cells, a fate suppressed in these progenitors in response to ectopic activation of Wnt signaling. However, finer dissection of this process reveals the stage of differentiation as a strong determinant of the response of a differentiating osteoblast to Wnt signaling. β-catenin stabilization in MSCs promotes proliferation at the expense of osteoblastic differentiation, while committed osteoblasts respond to the same stimulus by accelerating both growth and differentiation, at the expense of failure of terminal differentiation into mature osteoblasts. One possibility responsible for this state dependent response is the complex relationship
between Wnt signaling and the master regulatory transcription factors Runx2 and sp7. While the Runx2 P1 promoter possesses a Wnt responsive element that recruits β-catenin and TCF/LEF transcription factors, cells lacking β-catenin, are, like Lrp5−/− mice, still capable of expressing Runx2 in cells surrounding developing bone tissue. This suggests that other inputs in the Runx2 regulatory apparatus are sufficient to induce the primary osteoblast differentiation genetic program in the absence of Wnt.

Consistent with their effects on mature bone, the regulatory relationship between sp7 and the Wnt pathway appears to be reciprocal. While canonical Wnt signaling promotes both osteoblast differentiation and proliferation, sp7 promotes differentiation of maturing osteoblasts, while inhibiting their proliferative potential. It appears that this is accomplished at least partially by an sp7-mediated inhibition of Wnt activity. Sp7 appears to control the expression of the extracellular Wnt antagonist Dkk1 by direct binding to its promoter, and its expression is indeed abolished in sp7-null embryonic calvarial cells. Sp7 also inhibits β-catenin mediated transcription by direct interaction with the transcription factor TCF1. Therefore, it has been speculated that repeated downregulation of Wnt signaling is essential for balancing proliferative and cell fate priorities during osteoblastogenesis.

Molecular signaling and the vertebrate skeleton – The FGF pathway

Fibroblast growth factors (FGFs) are a family (22 members in both mouse and human) of secreted growth factors with roles in diverse biological functions.
processes that exert signaling activity by binding to tyrosine kinase fibroblast growth factor receptors (FGFRs), inducing intracellular pathways such as p38 MAPK, PLCγ, ERK1/2 or PI3K/AKT. Current thinking places the FGF signaling axis as a positive regulator of proliferation of progenitor cell populations and growth plate maturation during bone development. In cell culture, FGF signaling increases proliferation of immature osteoblasts while simultaneously blocking differentiation.

The identities of and roles of specific Fgf ligands at discrete stages of skeletal development are poorly understood. Fgf9 is expressed in early mesenchyme condensations prior to ossification, while Fgf2, Fgf5, Fgf6, and Fgf7 are expressed in the mesenchyme surrounding the ossification. All Fgfs have been identified in the coronal suture of E17.5 embryos, save for Fgf3 and Fgf4. While in vitro evidence has shown the capacity for Fgf ligands to stimulate osteoblast differentiation and or marker genes, animal models have failed to provide striking evidence of the necessity of a given ligand for a skeletal process, though it is clear that excessive ligand disrupts proper development. The construction of an Fgf2 knockout mouse provided an early illustration of this concept. Fgf2−/− mice are normal in appearance, but have lower bone mass density, concomitant with decrease thymidine incorporation in calvarial osteoblasts, suggesting an early proliferation defect behind the adult phenotype. Consistent with these observations, overexpression of Fgf2 in mice results in premature mineralization, achondroplasia and shortening of the long bone.
Similar observations were made in experiments using \textit{Fgf9} knockout and transgenic mice\textsuperscript{104, 105}.

As is not the case for the FGF ligands, there exists much human genetic evidence regarding the necessity of the FGFR genes in skeletal development. Of the four FGFRs, FGFRs1-3 are expressed in calvaria mesenchyme. Gain of function missense mutations in \textit{FGFR1}, \textit{FGFR2}, and \textit{FGFR3} cause a spectrum of fourteen disorders, most of whom share a craniosynostosis or chondrodysplasia feature\textsuperscript{106}. Despite the diversity in phenotypes resulting from those mutations, then, it makes sense to try to understand the common biology in these in these conditions. Craniosynostosis and chondrodysplasia differ fundamentally in the physiological process disrupted in their pathology. Craniosynostosis is a failure of the flat bones of the craniofacial skeleton to delay differentiation in progenitor cell populations, resulting in premature fusion of the sutures. However, chondrodysplasia is a defect in endochondral ossification, often resulting in shortening of the long bones of the limbs and the axial skeleton in general. So despite affecting two distinct pathways to mature bone, upregulation of FGF signaling in developing skeletal tissue results in a common cell biology defect: premature differentiation of progenitor cells.

\textbf{Genetic origins of the vertebrate skeleton – Signaling crosstalk}

While understanding the functions of individual signaling pathways in bone development is a necessary effort towards a complete theory of skeletogenesis, these deconstructions, in isolation, lead to an impoverished view
of a highly integrated process. Therefore, it is necessary to consider how these pathways interact, either to enhance or police each other's activities. Therefore, I will discuss known connections of each of the systems discussed above.

**Signaling crosstalk in the vertebrate skeleton - BMP and Wnt pathways**

Studies investigating interactions between Wnt and BMP in osteoblast differentiation have identified both synergistic and epistatic relationships between components of these pathways. At a fundamental level, BMP-2 driven osteogenesis is dependent on the presence of β-catenin. Many examples exist of synergistic activation of osteoblast marker genes by costimulation with BMP and Wnt ligands at early stages of osteoblastogenesis. Several possible explanations involving intracellular mediators of these signals have been proposed. In *Xenopus* embryos and cos-7 cells, Wnt signaling extends the duration of a ‘pulse’ of BMP signaling by regulating SMAD1 activity via GSK-3β dependent phosphorylation. Other researchers have described a mechanism involving the physical interaction of SMAD4 with TCF4 and the general co-activator protein p300.

However, at later stages in bone biology, they may have distinctly antagonistic roles. Where continued Wnt signaling is crucial for maintaining sufficient levels of bone mass density, BMP signaling at this stage actually acts in an catabolic manner. Deletion of *Noggin*, which codes for an extracellular inhibitor of BMP ligands, led to decreased bone mineral density (BMD) and bone formation in mice. The extracellular Wnt inhibitors *Dkk1* and *Sost* are
downstream of BMP signaling\textsuperscript{113}, and through their upregulation, Wnt signaling is attenuated. Genetic evidence for this interaction was observed in an osteoblast-specific \textit{Bmpr1a} knockout mouse, which had a high bone mass phenotype concomitant with upregulated Wnt signaling\textsuperscript{113}. BMP inhibition of Wnt occurs in uncommitted bone marrow cells via sequestration of the GSK-3β inhibitor/Wnt activator DSH by SMAD1\textsuperscript{114}.

Another level of BMP and Wnt integration to discuss is the combined cis regulation of both component and target genes. Chip-seq data from erythroid cells indicate that many active enhancers in these cells recruit both SMAD1 and TCF7L2\textsuperscript{115}. The promoters of \textit{Dlx5} and \textit{Msx2}, which are routinely and essentially upregulated in response to BMP signaling, respond synergistically to BMP and Wnt activation. Unsurprisingly, SMAD1, TCF4 and β-catenin are recruited to these promoters following dual stimulation of these pathways\textsuperscript{108}.

**Signaling crosstalk in the vertebrate skeleton - BMP and FGF pathways**

Unlike the complicated relationship between BMP and Wnt signaling, the association between BMP and FGF signaling has been described as largely cooperative. Similarly to its relationship with Wnt, many examples exist where BMP signaling is in part dependent on the presence of active FGF signaling to achieve full osteogenic effect. Mice null for \textit{Fgf2} have decreased \textit{Bmp2} expression\textsuperscript{116}, while FGF-2 and FGF-9 increase expression of \textit{Bmp2} in calvarial osteoblasts. Additionally, these ligands inhibit the expression of \textit{noggin}, an extracellular BMP inhibitor normally upregulated in response to BMP signaling\textsuperscript{117}. FGF
mediated suppression of noggin is also observed in vivo in the coronal dura mater during suture development. Noggin maintains the patency of flat bone sutures in the skull, so it is possible that some of the craniosynostosis phenotype arising from gain of function FGFR mutations is due, in part, to coordinately misregulated BMP signaling\textsuperscript{116,118}. FGF also upregulates BMP signaling beyond the context of increasing ligand-receptor association; $Fgf^{2-/-}$ osteoblasts have impaired colocalization of phosphorylated SMADs and RUNX2 in response to BMP-2 signaling, though the reason for this deficit is unclear\textsuperscript{116,119}. Finally, FGF-2 and BMP-2 have a synergistic effect on fracture healing: FGF-2 has a critical function at early stage while BMP-2 promotes mineralization at later stage\textsuperscript{120}.

**Signaling crosstalk in the vertebrate skeleton – Wnt and FGF pathways**

Wnt and FGF signaling have opposing effects during osteoblast differentiation\textsuperscript{121}. The convergence of Wnt and FGF signaling in skeletogenesis occurs primarily by the suppression of Wnt signaling by FGF signaling. Multiple mechanisms have been described underlying this process. At a fundamental level, the expression of components of the canonical Wnt pathway requires FGF signaling. mRNA expression of $Wnt10b$, $Lrp6$, and $\beta$-catenin are significantly downregulated in bone marrow stromal cells from $Fgf^{2-/-}$ mice\textsuperscript{122}. Exogenous application of Fgf2 ligand to these cells rescues both the osteogenesis defects while increasing $\beta$-catenin stabilization and nuclear localization. Comparative microarray analysis of osteoblasts derived from patients with gain of function
FGFR2 mutations identified the transcription factor SOX2 as dramatically (15 to 121 fold) upregulated compared to wildtype cells\textsuperscript{99}. Coimmunoprecipitation experiments demonstrated that SOX2 associates with β-catenin in osteoblasts and can repress activity of a reporter plasmid drive by TCF/LEF binding sites.

Wnt and FGF signaling interactions have also been studied genetically in the context of skull suture formation. Tellingly, deletion of the gene encoding the Wnt negative regulator, \textit{Axin2}, resulted in a phenotype similar to that observed in craniosynostosis in FGFR gain of function mutations\textsuperscript{123}. Upregulation of Wnt signaling in \textit{Axin2}-deficient mice was confirmed by increased nuclear accumulation of β-catenin. In concert, the proportion of FGFR positive cells at the suture was significantly reduced\textsuperscript{124}. Further altering the FGF/WNT balance by generating \textit{Axin2}\textsuperscript{−/−}, \textit{Fgfr1}\textsuperscript{−/−} mice produce sutures with ectopic cartilage formation\textsuperscript{125}. Cells at the front of these sutures had upregulated BMP signalling, as evidenced by increased SMAD phosphorylation. A complex mechanism in suture mesenchyme has been proposed, where Wnt signaling expands the population of skeletal precursors, while stimulating BMP signaling to counteract FGF signaling. In the presence of relatively high levels of FGF signaling, BMP signaling promotes osteoblastogenesis in the microenvironment, while reduced FGF signaling results in the effect of BMP to signaling to promote a chondrocytic fate.

\textit{Zebrafish as a model to study vertebrate skeletogenesis}
Zebrafish enjoy a burgeoning status as a more tractable alternative to the standard skeletal biology models of mouse and chick\textsuperscript{126}. Development of both the craniofacial\textsuperscript{127} and axial\textsuperscript{128} skeletons has been well characterized. Zebrafish produce the same skeletal cell types as higher vertebrates, albeit in simpler patterns\textsuperscript{129}. Additionally, gene expression events in skeletal elements are orthologous to those observed in higher vertebrates\textsuperscript{130}. Specifically for the purposes of this work, both zebrafish runx2 orthologs, \textit{runx2a} and \textit{runx2b} are expressed in nascent skeletal elements\textsuperscript{131}. Moreover, the appearance of the zebrafish skeleton is rapid; the first bony structure, the cleithrum\textsuperscript{d}, is visible within 72 HPF\textsuperscript{132}, though expression of bone marker genes in the anlagen begins at approximately 36 HPF\textsuperscript{133}. Potential bone specific deficiencies of the system, such as the lack of osteocytes or hematopoietic activity in the bone marrow, are not hindrances for exploring early development\textsuperscript{134}.

\textbf{Study aims}

Despite the identification of skeletogenesis specific roles of the signaling pathways discussed above and others\textsuperscript{e}, a coherent narrative of the genetics and cell biology underlying this process still eludes the field. Because of its singular

\textsuperscript{d} Most of the imaging in this document will focus on two bones as proxies for the expression in the rest of the skeleton. The cleithrum is a bone of mesodermal origin, and is roughly analogous to the shoulder girdle in mammals. Conversely, the opercle is derived from neural crest, and adopts a fan shaped morphology to lend structural support to the gill flap.

\textsuperscript{e} These include Notch, Indian Hedgehog, calcineurin, retinoic acid, p38 MAPK, among others.
role in both marking and initiating early stages of osteoblastogenesis, understanding the regulation of Runx2 itself is a potentially fruitful approach to understanding the biology of this process. Therefore, the rest of this document will be committed to describing the cis regulatory architecture responsible for regulating both the presence of the RUNX2 transcription factor itself in putative skeletal cells, but also the modulation of gene dosage that is so critical for proper execution of this process. Chapter 2 describes the results of a conservation based screen for conserved non-coding elements associated with the human RUNX2 locus capable of directing expression to bone. Chapter 3 relates a series of functional studies on individual elements, identifying upstream regulators with previously confirmed roles in skeletogenesis. Finally, Chapter 4 integrates the results from Chapters 2 & 3 for a summary, discussion of the implications of the work, and suggestions for future avenues of experimental inquiry informed by this effort.
CHAPTER 2

A SCREEN FOR RUNX2 ASSOCIATED ENHANCERS IDENTIFIES THREE CONSERVED NON-CODING ELEMENTS CAPABLE OF DIRECTING EXPRESSION TO OSTEOBLASTS IN VIVO.
CHAPTER 2

Introduction

The gene encoding the transcription factor RUNX2 was identified as the underlying cause of the human skeletal syndrome cleidocranial dysplasia (CCD)\textsuperscript{24,126}, resulting from haploinsufficiency. RUNX2 regulates expression of downstream genes important for osteoblast function, and its forced expression can upregulate those target genes\textsuperscript{25}. Mutation of the mouse gene demonstrated the requirement for \textit{Runx2} in bone formation throughout the skeleton, and its continued expression is also required for normal bone homeostasis\textsuperscript{22,127}.

The years since its identification have yielded a detailed understanding of the pathway downstream of \textit{Runx2} leading to differentiated osteoblasts, with identification of many genes whose transcription is directly regulated by \textit{Runx2}. Comparatively, almost nothing is known about the transcriptional regulation of \textit{Runx2} itself. This is a critical question, since initiation of \textit{Runx2} expression in development determines when and where bones will form, and its ongoing expression is important for proper maintenance of bone throughout life. Numerous signaling pathways have been implicated in its induction, but none has been shown to directly regulate \textit{Runx2} transcription \textit{in vivo}\textsuperscript{128}.

Direct regulation of a gene is accomplished by the binding of diffusible trans regulatory factors, either directly or to other trans factors, to cis-regulatory elements (CREs)\textsuperscript{129}. CREs are regions of genomic DNA with some role in activating, maintaining, or repressing transcription of an mRNA product.
Elements containing or immediately adjacent to the one of the transcriptional start sites of a gene are generally classified as promoters. They mark the site of recruitment of RNA polymerase and melting of the DNA strand and consequently possess an innate directionality. Conversely, enhancers can positively regulate gene transcription without regard to DNA strand orientation, and may be located at either a great distance from the transcriptional start site or, potentially on other chromosomes altogether\textsuperscript{130,131}, though this is not known to be a common phenomenon in vertebrate genomes. Additionally, exons of neighboring genes can also function as enhancers\textsuperscript{132,133}. Other forms of cis regulatory elements include locus control regions and silencers, capable of preventing gene activation and insulators, which form boundaries to prevent the spread of a repressive heterochromeric chromatin environment through the association of the CTCF protein.

\textit{Runx2} is somewhat noteworthy in that it possesses two distinct promoters. The proximal P2 promoter regulates the type I isoform, while the distal P1 promoter (\textit{Runx2} P1) regulates the type II isoform. The two proteins share the same functional domains and are similarly capable of transactivating target genes\textsuperscript{134}. The P2 promoter is active at a basal level in a broad number of cells and tissues, including the thymus, cartilage, periosteum, and suture tissue of the calvarium\textsuperscript{135-137} whereas the P1 promoter is active in hypertrophic chondrocytes and mature osteoblasts\textsuperscript{25,138}. Although \textit{in vivo}, transcription from both promoters are necessary for fine-tuning \textit{Runx2} expression\textsuperscript{139}, they are
incapable of directing proper expression of a reporter transgene by themselves\textsuperscript{140,141}, indicating the existence and necessity of a more elaborate cis regulatory architecture. Also, while most characterized CCD mutations affect the \textit{RUNX2} coding sequence, some cases have been associated with translocations of distal regions\textsuperscript{142,143} or have no identified coding sequence\textsuperscript{66} mutations, suggesting the presence of critical regulatory sequences whose mutation or disruption severely impairs gene expression.

Therefore, I hypothesized that there must exist additional enhancer elements necessary to direct \textit{Runx2} expression\textsuperscript{144}. Methodologies for identifying functional CREs are an area of ongoing inquiry in the literature, each possessing relative strengths and inherent limitations. In a developmental context, the current ‘gold standard’ experiment for confirming regulatory potential of a DNA region is the deletion of that element in the germline or a relevant integrated BAC \textit{in vivo} and confirmation of a phenotypic of transgene expression change. The chief advantage of the approach is the opportunity to observe an element in its native regulatory environment in a variety of tissue types. Unfortunately, isolating individual elements using germline modifications in mice is costly and time consuming. Additionally, functional redundancy between elements may mask or buffer the consequences of element loss.

Other methods rely on the flood of bioinformatic data that has been made available as the result of numerous genome sequencing projects as well as massively parallel sequencing technologies\textsuperscript{145}. Current estimates place
approximately 5% of the human genome under negative purifying selection\textsuperscript{146,147}, where evolutionary forces conserve sequence against the caprice of mutation. However, only 1.5% of the genome appears to have exonic coding potential. The difference between these proportions—popularly referred to as the ‘dark matter’ of the genome—is due to the relative difficulty of annotating aspects of genes without protein coding potential. Conservation based methods assume conserved non-coding elements possess functional importance whose ablation or alteration would have fitness consequences. While conservation is a generally reliable and accepted criterion for identifying candidate CREs associated with a given gene, there has not yet been an agreement in the field regarding the algorithm or parameters that are best suited for identifying CREs amongst diverse biological contexts\textsuperscript{129}. Notably, deletion of many ‘ultraconserved’ elements in mice resulted in no observable phenotype\textsuperscript{148}. Additionally, while conservation might be a good approach to discern sites of input for relatively ancient signaling connections, it is less useful at identifying newly arisen CREs, which are likely to be of the greatest interest from an evolutionary perspective.

Marrying chromatin immunoprecipitation with genome wide interrogation techniques such as microarrays and next generation sequencing permits the mapping of locations of both known transcription factors as well as histone modifications associated with regulatory activity to the genome in a variety of cell types and environmental situations. In particular, enrichment of methylation of 4\textsuperscript{th} lysine of the N-terminus of the histone H3 has been repeatedly
shown to be a powerful technique for locating promoters (H3K4me3) and 
 enhancers (H3K4me1/2) in the absence of detectable sequence conservation\textsuperscript{149,150} . However, these approaches are generally poorly suited for study of enhancers 
 involved in development, because of cell number requirements of precious 
 embryonic materials as well as their necessarily static nature.

 With these experimental approaches and limitations in mind, I decided rely 
on moderate conservation amongst vertebrates to identify CREs associated with the human \textit{RUNX2} locus. A list of candidate CREs was generated by 
 interrogating the locus using PhastCons\textsuperscript{151} , an algorithm which relies on 
multiple vertebrate genomes to both identify conserved elements, as well as a quantitative measure of the evidence of that conservation, permitting a ranking 
and prioritization of element testing. Because location is a poor predictor of regulatory function, and enhancers can exist at great distances from their 
associated genes, I examined sequences in an interval of >1 Mb containing human \textit{RUNX2}.

 Finally, the qualities of the systems used to actually confirm regulatory 
activity merits discussion. Especially in developmental contexts, \textit{in vivo} systems 
are preferable to cell based assays because of the ability to simultaneously 
observe the full range of likely dynamic regulatory activity in all tissues. 
Traditionally, mice have been used for this purpose, but for the reasons 
discussed above, these are a difficult choice for studying embryonic expression, 
as it requires sacrificing a transgenic animal at each time point of interest.
Additionally, because many transgenes show position dependent effects resulting in ectopic activation in cells that do not actually reflect biological reality, it is necessary to examine a number of embryos to ensure that uniform and robust pattern of expression can be elucidated. In contrast, zebrafish produce hundreds of optically accessible embryos in single clutch. Finally, a highly efficient transgenesis methodology based on the Tol2 transposase permits the construction of potentially hundreds of mosaic transgenic fish in a single morning\textsuperscript{152}.

Specifically to skeletal biology, zebrafish are increasingly being employed as an alternative to the standard models of mouse and chick\textsuperscript{153}. In addition to the reasons of tractability listed above, they produce the same skeletal cell types as higher vertebrates, albeit in simpler patterns. The appearance of the zebrafish skeleton is rapid; the first bone, the cleithrum\textsuperscript{f}, is visible within 72 HPF\textsuperscript{154}, though expression of bone marker genes in the anlagen begins at approximately 36 HPF\textsuperscript{155}. Potential bone specific deficiencies of the system, such as the lack of osteocytes or hematopoietic activity in the bone marrow, are not hindrances for exploring early development\textsuperscript{156}.

In this chapter, I describe a screen for $RUNX2$-associated CREs in the human genome. Candidates were selected on the basis of moderate conservation

\textsuperscript{f} Most of the imaging in this document will focus on two bones as proxies for the expression in the rest of the skeleton. The cleithrum is a bone of mesodermal origin, and is roughly analogous to the shoulder girdle in mammals. Conversely, the opercle is derived from neural crest, and adopts a fan shaped morphology to lend structural support to the gill flap.
amongst vertebrates and tested for *in vivo* activity in a zebrafish system. Commonalities and differences in expression domains are also noted.

**Materials and Methods**

*Ethics statement.* All animal work was conducted according to national and international guidelines, and with the knowledge and approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania.

*Identification of conserved non-coding elements associated with RUNX2.* To identify sequences for *in vivo* analysis, the candidate locus (hg18; chr6: 44904448-45974166) was interrogated for conserved non-coding elements, as defined by the PhastCons Vertebrate Conserved Elements track during January, 2010. Elements overlapping known RefSeq exons were excluded from further analysis. The top 50 scoring elements (LOD >454) were amplified by PCR (Table 1) from human genomic DNA with LA Taq polymerase (Takara), and cloned into the Tol2 transposon containing vector pattP-Tol2-EGFP as previously described.

*Transgenesis and expression analyses.* Fish were cared for following standard protocols. Each construct for analysis was injected as previously described in at least two separate experiments, and mosaic expression of eGFP analyzed in a minimum of 150 embryos. Embryos were screened from 1-5 DPF using a Zeiss V12 Stereomicroscope, and imaged with AxioVision 4.5 software. For those constructs regulating a consistent expression pattern, embryos were raised to adulthood and their progeny examined for expression after germline...
transmission. All constructs were examined in at least three independent transgenic lines for consistent expression. The +210RUNX2:mCherry construct was made in a version of the same vector with mCherry coding sequence instead of eGFP. sp7:mCherry fish were made by injection of a modified BAC containing medaka sp7 regulatory sequences, a kind gift from Christoph Winkler.  

**Confocal imaging.** Embryos were anesthetized in Tricaine and mounted in MatTek glass bottom culture dishes in 1% low melt SeaPlaque agarose. Images were acquired on an Olympus IX 81 microscope equipped with a Yokogawa CSU 10 scan head combined with a Hamamatsu EMCCD camera (model C9100-13, Bridgewater, NJ). Hardware was controlled by Slidebook version 5.0 (Intelligent Imaging Innovations). Diode lasers for excitation (488 nm for eGFP and 561 nm for mCherry) were housed in a Spectral Applied Research launch (Richmond Hill, Ontario). Confocal image stacks were processed with ImageJ (http://rsbweb.nih.gov/ij).

**In situ hybridizations.** Whole mount in situ hybridizations were performed as described, with the following modifications: 0.05% CHAPS detergent was added to the pre-hybridization and hybridization solutions to prevent embryo clumping, and the concentration of NBT was reduced 10-fold in the staining solution to permit overnight development with low background. Stained embryos were dehydrated by successive methanol washes, cleared in methyl salicylate, and mounted in Permount medium (Fisher; SP15-100) between bridged coverslips. Microscopy was performed on an Olympus BX51 with Nomarski.
optics. Images were acquired using Spot Basic version 4.6 (Diagnostic Instruments). Further adjustments to white balance and contrast were performed with Adobe Photoshop.

Quantitative PCR. RNA was extracted from whole zebrafish embryos as previously described\(^{160}\). cDNA was synthesized using the cDNA high capacity transcription kit (Invitrogen). qPCR was performed in 20 \(\mu\)l reactions using ABI Sybr Green Master mix, and 250\(\mu\)M primer concentration. Samples were amplified using an ABI StepOnePlus real time PCR system.

Identification of cis regulatory elements associated with the zebrafish orthologs runx2a and runx2b. The program WPH finder was downloaded from http://rana.lbl.gov/downloads/wph.tar.gz. The three characterized human enhancers (-460RUNX2, +210RUNX2 and +542RUNX2) were used as substrates to build word profiles based on the occurrence of 8-mers. These were individually used scan the loci containing the zebrafish Runx2 orthologs runx2a (chr17: 5385672-5740215;Zv8) and runx2b (chr20: 44206838-44359224;Zv8) using 250 base pair windows offset by 100 base pairs. Repetitive sequences were removed with RepeatMasker prior to scanning. An arbitrary cut off of \(Z>5\) determined which candidate elements to progress to functional testing.

Results

A screen for RUNX2 associated enhancers identifies three conserved non-coding elements capable of directing expression to osteoblasts in vivo.
As potential cis–regulatory elements for *RUNX2*, we selected the 50 most conserved non–coding sequences within the ~1Mb interval bounded by the 3' UTR of the overlapping gene *SUPT3H* and the 3' UTR of the downstream gene *CLIC5* in the human genome (Figure 1). We grouped proximate sequences into larger amplicons, resulting in 36 constructs (Table 1). Each sequence was tested for *in vivo* enhancer activity, through its ability to direct tissue–specific eGFP expression in zebrafish. We initially screened by examining approximately 150 mosaic injected embryos for fluorescence from 1-6 days post fertilization (DPF); constructs positive in the initial screen were passed through the germline for detailed characterization.

In total, I identified three enhancers capable of directing reporter gene expression to osteoblasts: a distant upstream enhancer located in the last intron of *SUPT3* (-460RUNX2)*, a downstream enhancer in the intergenic space between *RUNX2* and *CLIC5* (+542RUNX2), and one in the last intron of *RUNX2* itself (+210RUNX2).

*Comparison of transgene activity by confocal microscopy and in situ hybridization reveals distinct, yet redundant expression patterns.*

While all three enhancers direct expression to osteoblasts, they do not have identical activities. Prior to formation of the first bones, +210RUNX2 is transiently active in the branchial arches, as evidenced by GFP expression in

* The nomenclature used here and subsequently in this document to identify specific enhancer elements is the distance from the transcriptional start site of a gene in kilobases, relative to the directionality of the open reading frame.
live embryos at 2DPF(Figure 2.2j). An earlier and less enduring expression in the branchial arches is directed by +542\textit{RUNX2} between 24-48 HPF (data not shown). +542\textit{RUNX2}:eGFP expression is first detectable in the cleithrum anlagen between 28-32 HPF, while this activity is relatively delayed in embryos carrying the other two transgenes (Figure 2.3a,d). Referring to the confocal data at 3DPF shows that, again, while all three transgenes drive expression in cleithrum in osteoblasts, the expression is much more pronounced in the +542\textit{RUNX2}:eGFP (Figure 2.2f) and -460\textit{RUNX2}:eGFP lines (Figure 2.2o).

Furthermore, +542\textit{RUNX2} mediated expression in the cleithrum at this stage is uniform along the dorsal ventral gradient, while in -460\textit{RUNX2}, expression is relatively punctuated at the dorsal and ventral extremes of the bone.

All three enhancers are active in cells of the opercle (Figure 2.4), a neural–crest derived bone\textsuperscript{161} that forms by membranous ossification lateral to the branchial arches. However, +210\textit{RUNX2} directs expression to the osteoprogenitors surrounding the edges of the bone (Figure 2.4j,m,p), while the activity of the other two enhancers is confined to cells within the bone itself. Similarly to the expression differences observed in the cleithrum, +542\textit{RUNX2} directs expression uniformly throughout the bone (Figure 2.4s,v,y), while +460\textit{RUNX2} expression is enhanced in the strut and fan structures of the opercle (Figure 2.4 a,d,g)

+154runx2a, the zebrafish ortholog of +210\textit{RUNX2}, is conserved at the levels of sequence and function.
While identifying conserved enhancers between distantly related species is not always a straightforward effort, alignment of vertebrate genomes using MultiZ in the UCSC genome browser revealed apparent linear alignment delement, +154runx2a, directs expression the branchial arches at 3DPF (Figure 2.5 a) as well as to bony elements such as the opercle and branchiostegal rays at 5DPF (Figure 2.5 b,e).

Only +542RUNX2 directs expression to the developing vertebrae.

Later in zebrafish development (10-21 DPF), the vertebrae form from migrating sclerotome cells that surround the spinal cord. Though at 5DPF, all three enhancers direct expression to all visible skeletal structures, the ability to direct expression to the vertebral arches at 14DPF is limited to +542RUNX2 (Figure 2.6c). Some expression by -460RUNX2:eGFP is visible in the centra of the vertebrae. (Figure 2.6a). +210RUNX2 failed to direct expression to any aspect of the vertebral column (Figure 2.6b).

Chromatin features of identified enhancers.

As discussed in the introduction, enhancer associated chromatin marks are an often-employed method to identify loci with regulatory activity in a cell or tissue type being studied. There exists an unpublished, publicly available ChIP-Seq data set from a normal human osteoblast cell line as part of the ENCODE project. Enrichment for regulatory element associated marks (H3K4me1, H3K27ac) in a normal human osteoblast cell line as part of the ENCODE project.

\(^1\) According the documentation accompanying the data set, the cell line is normal human osteoblasts (NHOst) from Lonza (#CC-2538). The ChIP-Seq data was produced by a collaboration of Bradley Bernstein and Greg Crawford.
H3K4me2, H3K27Ac) was examined as well as recruitment of the enhancer associated coactivator p300 (Figure 2.7). +542RUNX2 and +210RUNX2 were significantly enriched for p300, as well as H3K4me1 and H3K4me2, and were identified as peaks in the data set as determined by a hidden Markov model algorithm. Curiously, no enrichment for any of the studied histone modifications or proteins was observed in -460RUNX2.

The three RUNX2-associated enhancers contain enough information content to efficiently identify other skeletal specific enhancers in a non-conservation based approach.

With the limitations of conservation based approaches to discovering CREs in mind, I wanted to explore whether other methodologies might be efficient at identifying enhancers associated with the zebrafish RUNX2 orthologs, runx2a and runx2b, for which there are few alignable genomes available to detect conservation. One such strategy involves utilizing the information content of known enhancers as a basis for predicted new ones from untested sequence data. WPH Finder is such an algorithm that has been successfully used to recognize enhancer elements not identifiable by linear conservation between Drosophila species. It trains itself by counting the occurrence of specific eight letter DNA ‘words’ (which likely correspond to transcription factor binding sites, or other features conferring regulatory activity upon an element), forming a profile of these word, and then testing windows of candidate sequences for overrepresentation relative to it.
Using this methodology, the loci of the zebrafish *Runx2* orthologs, *runx2a* and *runx2b* were scored for word profile similarity to each of the three known human enhancers (Tables 2.2, 2.3). Eighteen candidates were tested from the *runx2a* locus, and ten were examined from that of *runx2b*. In contrast to the relatively low rate of skeletal enhancer activity identified in conserved candidates from the human locus (3/38 = 7.8%), 38.9% (7/18) of amplicons tested from the *runx2a* locus and 50% (4/10) of those from *runx2b* showed some pattern of skeletal expression (Figure 2.8a,b). Notably, almost all of the positive elements (11/12) fall within the coding region of either one of the *Runx2* orthologs, or *supt3h*, whose syntenic relationship with *Runx2* is ancient. Representative views of expression patterns of these elements show expression in both subsets of osteoblasts as well as cartilaginous structures (Figure 2.8c-h).

\[^{i}\text{These numbers exclude +154runx2a, which had already been shown to regulate a skeletal expression pattern.}\]
Figure 2.1. – Broad distribution of osteoblast specific enhancers at the \textit{RUNX2} locus. The \textit{RUNX2} locus (chr6: 44,904,448-45,971,166, hg18) as represented in the UCSC Genome Browser is shown, with the conserved non–coding elements tested for enhancer assay indicated at top. Elements testing positive for osteoblast expression \textit{in vivo} are shown in green, while those with no activity in skeletal tissues are indicated in red. Tracks displaying all conserved elements as defined by PhastCons amongst vertebrates and mammals are displayed at the bottom of the figure to visualize relative conservation across the genomic region.
Figure 2.2 – Expression patterns of three human *RUNX2* associated enhancers. Lateral views of doubly transgenic zebrafish embryos for *sp7:mCh* and -460 *RUNX2:eGFP* (A-I), +210*RUNX2:eGFP* (J-R), and +542*RUNX2* (Q-AA) were imaged at 2DPF and 3DPF. In 2DPF embryos, coexpression of *sp7:mCh* and -460*RUNX2:eGFP* (A) and +542*RUNX2* :eGFP (S) in the developing cleithrum, but not in +210*RUNX2* (J). Conversely, branchial arch expression in +210*RUNX2* :eGFP is apparent at 2DPF. All three *RUNX2* transgenes express in the opercle anlage at 3DPF with *sp7:mCh* (D,M,V). Ventral views imaged at 5DPF demonstrate concomitant *RUNX2* transgene expression in later ossifications (G,P,Y). *ba*, branchial arches; *bs*, brachiostegal ray; *cl*, cleithrum; *de*, dentary *mx*, maxilla; *op*, opercle; *Scale bar = 50 mm*
Figure 2.3 – Differential expression initiation times in the cleithrum anlagen. Embryos from each transgenic line were fixed at 4 hour intervals and analyzed by in situ hybridization with an eGFP probe to determine onset of expressing in the cleithrum. Dorsal views of representative embryos of +542RUNX2:eGFP (a,d), +210RUNX2:eGFP (b,e), and +542RUNX2:eGFP (c,f).
Figure 2.4 – Distinct expression domains in the developing opercle.

Lateral views of the opercle in doubly transgenic zebrafish embryos for sp7:mCh and (a-i), -460 RUNX2:eGFP (j-r) +210RUNX2:eGFP, and (q-aa) +542RUNX2 were imaged at 3DPF, 4DPF and 5DPF; Scale bar = 10 mm
Figure 2.5 – +154runx2a, a conserved ortholog of +210RUNX2, directs expression to the branchial arches and osteoblasts. Lateral (a-d) and ventral (e-g) of doubly transgenic zebrafish for +154runx2a:eGFP and +210RUNX2:mCh at 3DPF(a) and 5DPF (b-g). ba, branchial arches; bsr2, brachioseptal ray 2; bsr3, brachioseptal ray 3; op, opercle;
Figure 2.6 – +542RUNX2 directs expression to the developing vertebral arches. Lateral views of the developing vertebrate column of doubly transgenic zebrafish for the indicated RUNX2:eGFP transgene and sp7:mCh at 14 DPF. -460RUNX2:eGFP expression (a) is directed to the anterior edge of the presumptive vertebrate. +210RUNX2:eGFP (b) does not express in any portion of the anatomy and +542RUNX2:eGFP (c) directs expression to the vertebral arches. (d-f) demonstrate sp7:mCh mediated expression; (g-i) shows a merge of these two pattern,
Figure 2.7 – Analysis of chromatin environment at three RUNX2 associated enhancers in a normal human osteoblast cell line. The human skeletal enhancers -460 RUNX2 (a), +210 RUNX2 (b), +542 RUNX2 (c) are shown as represented in the UCSC genome browser along with tracks indicating both the enrichment of and presence of peak of the following histone modification/proteins in normal human osteoblast cell lines as defined by ChIP-Seq.
Figure 2.8 – Screen for skeletal enhancers associated with the zebrafish Runx2 orthologs, runx2a and runx2b. The runx2a(a; chr17: 5385672-5740215;Zv8) and runx2b(b; chr20: 44206838-44359224;Zv8) loci(chr6:44,904,448-45,971,166, hg18) as represented in the UCSC Genome Browser is shown, with the elements tested for enhancer assay indicated at top. Elements testing positive for skeletal expression in vivo are shown in green, while those with no activity in skeletal tissues are indicated in red. Tracks displaying all conserved elements as defined by 6 way MultiZ alignment displayed at the bottom of the figure to visualize relative conservation across the genomic region. Representative images of transgenic fish carrying these elements illustrate the diversity of skeletal expression observed. -52runx2a:eGFP (c) expresses in a subset of ossifying structures at 5DPF; +54runx2b:eGFP (d) expression is exclusive to osteoblasts; +28runx2b:eGFP (e) expression is found in the cartilaginous elements of the neurocranium; -32runx2a:eGFP (f) expressed in the branchial arches as well as the dentary and maxilla forming the mouth; and +38runx2b:eGFP (h) expresses in the pharyngeal skeleton and dermal bones. ba, branchial arches; bh, basihyal; bs, brachiostegal ray; cb5, ceratobranchial 5; ch, ceratohyal; chb, ceratohyal bone; cl, cleithrum; de, dentary; e, ethmoid; mk, Meckel’s cartilage; mx, maxilla; op, opercle; pq, palatoquadrate
<table>
<thead>
<tr>
<th>Element</th>
<th>Location</th>
<th>Size</th>
<th>LOD</th>
<th>F primer</th>
<th>R primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>-488</td>
<td>44808206-8579</td>
<td>373</td>
<td>1105</td>
<td>agccggcacgagagacttttta</td>
<td>aagagcagcagagaaaagggagcag</td>
</tr>
<tr>
<td>-472</td>
<td>44824048-4376</td>
<td>328</td>
<td>780</td>
<td>ttggcgtacattcatttcattg</td>
<td>tatcattagcgtccggttcttg</td>
</tr>
<tr>
<td>-469</td>
<td>44827066-7407</td>
<td>341</td>
<td>677</td>
<td>gcacatgagagagcactga</td>
<td>cagttcacgggacagacg</td>
</tr>
<tr>
<td>-460</td>
<td>44835694-6992</td>
<td>398</td>
<td>692</td>
<td>gcgaaaccaacaccctcctca</td>
<td>aggacagaaactgggcccttcctca</td>
</tr>
<tr>
<td>-410</td>
<td>44886294-6648</td>
<td>354</td>
<td>1190</td>
<td>agccacgtgggttcctccaa</td>
<td>gcccccctgaaacaaagaaacct</td>
</tr>
<tr>
<td>-336</td>
<td>4495617-9931</td>
<td>314</td>
<td>454</td>
<td>gctctccacagatgaagcact</td>
<td>tcgltgcacgttggggtagl</td>
</tr>
<tr>
<td>-326</td>
<td>44970259-0571</td>
<td>312</td>
<td>741</td>
<td>gcccgtgcctgtgacctccca</td>
<td>tgtgcgagtcgctccctca</td>
</tr>
<tr>
<td>-307</td>
<td>44988634-9028</td>
<td>394</td>
<td>1126</td>
<td>cttggcgaataaaacacaga</td>
<td>gcccattatcttcctacaacag</td>
</tr>
<tr>
<td>-252</td>
<td>45043906-4200</td>
<td>294</td>
<td>625</td>
<td>gcagttgctaggagccagaaag</td>
<td>ggctgaattgtaaagggcgtca</td>
</tr>
<tr>
<td>-230</td>
<td>45065752-8659</td>
<td>907</td>
<td>1983</td>
<td>acagaaagcagagagcagga</td>
<td>aacctgcccaaaatgtgcag</td>
</tr>
<tr>
<td>-202</td>
<td>45093924-4357</td>
<td>433</td>
<td>855</td>
<td>ggggctcagcagaaacccgt</td>
<td>gcacctgtcttcccctcctttt</td>
</tr>
<tr>
<td>-132</td>
<td>45163927-4407</td>
<td>480</td>
<td>2070</td>
<td>ttcgtatccgccgaggtta</td>
<td>ttcagttgtggcttcttacg</td>
</tr>
<tr>
<td>-127</td>
<td>45169299-9615</td>
<td>316</td>
<td>600</td>
<td>tgcacgtgagtccttcagga</td>
<td>tgcctgttttgagagctgggal</td>
</tr>
<tr>
<td>-6</td>
<td>45290403-1852</td>
<td>1449</td>
<td>973</td>
<td>670</td>
<td>tcgtactagctcaagaaacaga</td>
</tr>
<tr>
<td>-1</td>
<td>45294619-8531</td>
<td>1912</td>
<td>896</td>
<td>888; 3526</td>
<td>cgcgtagtggctgttttgta</td>
</tr>
<tr>
<td>+4</td>
<td>45299716-0238</td>
<td>522</td>
<td>697</td>
<td>668</td>
<td>tgtggaaaagagcttgggtt</td>
</tr>
<tr>
<td>+9</td>
<td>45305220-5485</td>
<td>265</td>
<td>758</td>
<td>aagccaccccggacagttgga</td>
<td>cccgctgacagacattctggtga</td>
</tr>
<tr>
<td>+40</td>
<td>45335673-5997</td>
<td>324</td>
<td>892</td>
<td>aaccccgctcctcaagaaagt</td>
<td>ttcgtgaggaagggctgtcct</td>
</tr>
<tr>
<td>+90</td>
<td>45386404-6653</td>
<td>249</td>
<td>520</td>
<td>gggagcggggggcggaggata</td>
<td>cctctccctctccactgct</td>
</tr>
<tr>
<td>+62</td>
<td>45388499-0699</td>
<td>2200</td>
<td>674</td>
<td>gagcagctgatcggagga</td>
<td>ctctccctcactccagagacag</td>
</tr>
<tr>
<td>+104</td>
<td>45395557-1838</td>
<td>2281</td>
<td>1546</td>
<td>tcgcctgattgacttcataa</td>
<td>tcaagtctgattgagctgcct</td>
</tr>
<tr>
<td>+114</td>
<td>45406604-0034</td>
<td>430</td>
<td>1238</td>
<td>ttcctccctctcgcagtagac</td>
<td>ttttcctccctctcgacaaacaa</td>
</tr>
<tr>
<td>+118</td>
<td>45413657-3853</td>
<td>196</td>
<td>688</td>
<td>tcgcaagtggcacaacagggata</td>
<td>gcagagtggcccaacaaacaaalag</td>
</tr>
<tr>
<td>+145</td>
<td>45441259-1580</td>
<td>314</td>
<td>734</td>
<td>cccacgtgcaattgttgct</td>
<td>gcacaaatctggcaggcgca</td>
</tr>
<tr>
<td>+184</td>
<td>45479980-1419</td>
<td>1439</td>
<td>726</td>
<td>gacgctggtgcccccctcct</td>
<td>gcagtcatgcctgtttgatoc</td>
</tr>
<tr>
<td>+205</td>
<td>45501790-1314</td>
<td>235</td>
<td>516</td>
<td>agatgctgccccctcctgct</td>
<td>ggtatctgcttcggcactg</td>
</tr>
<tr>
<td>+210</td>
<td>45506418-6731</td>
<td>313</td>
<td>1302</td>
<td>aatagtgcctccagtttgg</td>
<td>ggcagcggcgttagattgtgga</td>
</tr>
<tr>
<td>+224</td>
<td>45520471-2527</td>
<td>251</td>
<td>511</td>
<td>agccagagctggattggact</td>
<td>ttgctctatcctctcccaag</td>
</tr>
<tr>
<td>+228</td>
<td>45524043-4304</td>
<td>261</td>
<td>608</td>
<td>gcatcttgcctccacctg</td>
<td>gcagctgattgctcactgcta</td>
</tr>
<tr>
<td>+235</td>
<td>45531043-1392</td>
<td>349</td>
<td>766</td>
<td>ccccagttggtggccttct</td>
<td>ttccaaagctgctggcagta</td>
</tr>
<tr>
<td>+262</td>
<td>45556534-8961</td>
<td>327</td>
<td>1283</td>
<td>tgacacactggctcaggtl</td>
<td>gggagtgagctgacgatgt</td>
</tr>
<tr>
<td>+298</td>
<td>45583889-4243</td>
<td>354</td>
<td>648</td>
<td>ttcctgcccaatcagctccag</td>
<td>atgcctgagcggcctgggt</td>
</tr>
<tr>
<td>+429</td>
<td>45725477-5774</td>
<td>297</td>
<td>1021</td>
<td>aatcaacagatggcagaggg</td>
<td>cttgaaccccaagacagacag</td>
</tr>
<tr>
<td>+488</td>
<td>45784423-4811</td>
<td>388</td>
<td>911</td>
<td>cccctctcttctccatctcc</td>
<td>ccccccgtatcttctctctct</td>
</tr>
<tr>
<td>+526</td>
<td>45822026-2328</td>
<td>302</td>
<td>629</td>
<td>gggcagattgtgcctgaggg</td>
<td>acagctgaccctggtcctcacc</td>
</tr>
<tr>
<td>+542</td>
<td>45838035-8397</td>
<td>362</td>
<td>773</td>
<td>acagcgcattttactccca</td>
<td>cagcagagagacagagagaggggaaa</td>
</tr>
<tr>
<td>+552</td>
<td>45848037-8491</td>
<td>454</td>
<td>635</td>
<td>643</td>
<td>caacacccagggagaaactcgg</td>
</tr>
</tbody>
</table>
Table 2.1. Sequences at *RUNX2* locus tested for enhancer activity. Each element is named for the distance in kilobases from the transcription start site of *RUNX2*; the location of each conserved region is given as coordinates in hg19, chr6. Where multiple conserved sequences were grouped into one amplicon, the LOD scores for each are listed separately. The primer sequences are those used to amplify the elements from genomic DNA for testing. The three elements in bold are those with activity in osteoblasts.
<table>
<thead>
<tr>
<th>Name</th>
<th>Coordinates(Zv8)</th>
<th>F primer</th>
<th>R primer</th>
<th>Product size (bp)</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>+321runx2a</td>
<td>chr17:5,278,105-5,278,353</td>
<td>ACAGATGGCGGGAAGGAGCAGT</td>
<td>GCACGTCCGAAATGCAACA</td>
<td>1061</td>
<td>-460</td>
</tr>
<tr>
<td>+261runx2a</td>
<td>chr17:5,338,455-5,338,703</td>
<td>TGTGAAGGTGGTTGAGCAACA</td>
<td>CGGTCTCTGGTTGACTAGT</td>
<td>870</td>
<td>+542</td>
</tr>
<tr>
<td>+258runx2a</td>
<td>chr17:5,341,255-5,341,503</td>
<td>CGAGCGTCTGAGGACAA</td>
<td>AGGAAACACTGAAGGACAAAATGC</td>
<td>852</td>
<td>-460</td>
</tr>
<tr>
<td>+229runx2a</td>
<td>chr17:5,371,055-5,371,303</td>
<td>TGTGAAGCTGCTTTGACACA</td>
<td>TTGCAGTGAGATGCGTTG</td>
<td>892</td>
<td>+542</td>
</tr>
<tr>
<td>+206runx2a</td>
<td>chr17:5,382,750-5,382,106</td>
<td>AGCACCACACCAACTGGA</td>
<td>GGACGACTGGAAGGACTTGG</td>
<td>938</td>
<td>210</td>
</tr>
<tr>
<td>+179runx2a</td>
<td>chr17:5,420,750-5,420,105</td>
<td>AGCACCACCATGACAA</td>
<td>ATCAGATTCATGCGTTG</td>
<td>991</td>
<td>-460</td>
</tr>
<tr>
<td>+79runx2a</td>
<td>chr17:5,529,955-5,530,203</td>
<td>AGCAGTTAGCCTTGTGAGAAG</td>
<td>ACATATTPGGCCGTGCAGT</td>
<td>1017</td>
<td>-460</td>
</tr>
<tr>
<td>+50runx2a</td>
<td>chr17:5,579,405-5,579,653</td>
<td>TCTGCTGGCCCACAAAGA</td>
<td>TCTTGGAGGAAGACTGGCAAAAGC</td>
<td>882</td>
<td>-460</td>
</tr>
<tr>
<td>+14runx2a</td>
<td>chr17:5,585,655-5,585,903</td>
<td>TTCAATGGACTTTGATTCAGCTT</td>
<td>CAAGCAGTGACTGAAAGATTGAAAT</td>
<td>852</td>
<td>+542</td>
</tr>
<tr>
<td>+1runxa</td>
<td>chr17:5,599,105-5,599,353</td>
<td>CGCGATGATGACAAACAA</td>
<td>TCACCAAGACCTTGAGAAG</td>
<td>885</td>
<td>-460</td>
</tr>
<tr>
<td>+1runxa</td>
<td>chr17:5,600,005-5,600,253</td>
<td>TCTAAAGCGGAGGAGCAAAAAA</td>
<td>TGACCGGGAAAACAGGGAGA</td>
<td>865</td>
<td>-460</td>
</tr>
<tr>
<td>-1runx2a</td>
<td>chr17:5,600,505-5,600,753</td>
<td>TCTAAAGCGGAGGAGCAAAAAA</td>
<td>TGACCGGGAAAACAGGGAGA</td>
<td>884</td>
<td>+542</td>
</tr>
<tr>
<td>-2runx2a</td>
<td>chr17:5,632,105-5,632,353</td>
<td>GGCAACCACAAATTGAAACC</td>
<td>CCTAAACCAGGGGACTAA</td>
<td>1455</td>
<td>+542</td>
</tr>
<tr>
<td>-32runx2a</td>
<td>chr17:5,651,555-5,651,803</td>
<td>GGGCAACCACAAATTGAAACC</td>
<td>CCTAAACCAGGGGACTAA</td>
<td>949</td>
<td>-460</td>
</tr>
<tr>
<td>-52runx2a</td>
<td>chr17:5,740,055-5,740,303</td>
<td>GGGCAACCACAAATTGAAACC</td>
<td>CCTAAACCAGGGGACTAA</td>
<td>2456</td>
<td>+542</td>
</tr>
</tbody>
</table>

52
<table>
<thead>
<tr>
<th></th>
<th>chr</th>
<th>start</th>
<th>end</th>
<th>sequence</th>
<th>primer sequence</th>
<th>length</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>chr17:5,741,305-5,741,553</td>
<td>A</td>
<td></td>
<td>CGAACAGACAGATGAAATAAAAGACA</td>
<td>GCAACCCATCTCTGGGAAC</td>
<td>1282</td>
<td>+542</td>
</tr>
<tr>
<td>-141runx2a</td>
<td>chr17:5,757,755-5,758,003</td>
<td>TGGGCTGGAGACCAAGAAAA</td>
<td>TGCATTGCACATAGGGGACA</td>
<td>1192</td>
<td>+210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-150runx2a</td>
<td>chr17:5,753,255-5,753,503</td>
<td>TCCCGAAGATCTGGCAAATG</td>
<td>GGCTGGATGCATCATTTTT</td>
<td>863</td>
<td>+460</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2. Sequences at the runx2a locus tested for enhancer activity.

Each element is named for the distance in kilobases from the transcription start site of runx2a. The primer sequences are those used to amplify the elements from genomic DNA for testing. The seven elements in bold are those with skeletal activity. Similarity indicates the human RUNX2 associated enhancer to which WPHFinder identified word profile similarity.
<table>
<thead>
<tr>
<th>Name</th>
<th>Coordinates(Zv8)</th>
<th>F primer</th>
<th>R primer</th>
<th>Product size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-45runx2b</td>
<td>chr20:44,131,047-</td>
<td>TGGCGCTTTTCCAGCAATT</td>
<td>TGGGCACGCTATGATGAC</td>
<td>921</td>
</tr>
<tr>
<td>-36runx2b</td>
<td>chr20:44,121,758-</td>
<td>GTGAAGGCTCCGTCACCT</td>
<td>GGCAAAGACATACGAGCAGAAA</td>
<td>993</td>
</tr>
<tr>
<td>+11runx2b</td>
<td>chr20:44,075,029-</td>
<td>GGCGGGCATGTGAGTTCTA</td>
<td>CGGAGAGAGGAGGTACAGTGGA</td>
<td>907</td>
</tr>
<tr>
<td>+28runx2b</td>
<td>chr20:44,058,608-</td>
<td>CAAAACATGAACGGCGAGA</td>
<td>CGGCAGCCAGAAGAGAGAAGAG</td>
<td>918</td>
</tr>
<tr>
<td>+52runx2b</td>
<td>chr20:44,032,209-</td>
<td>TGTGCTCACCTTTAAGTGGGA</td>
<td>GGGAGAGGCGCCCTGAGCATA</td>
<td>884</td>
</tr>
<tr>
<td>+54runx2b</td>
<td>chr20:44,019,065-</td>
<td>TGCTTTTATATATATCTCCTTCA</td>
<td>GGAACACACTCAGTACACACTCA</td>
<td>1000</td>
</tr>
<tr>
<td>+61runx2b</td>
<td>chr20:44,002,595-</td>
<td>TGCTTTTATTATCTCCTTTGCA</td>
<td>CCTACGTGCGGAAAGGCTA</td>
<td>895</td>
</tr>
</tbody>
</table>

Table 2.3. Sequences at the runx2b locus tested for enhancer activity.

Each element is named for the distance in kilobases from the transcription start site of runx2b. The primer sequences are those used to amplify the elements from genomic DNA for testing. The five elements in bold are those with skeletal activity.
CHAPTER 3

FUNCTIONAL CHARACTERIZATION OF \textit{RUNX2} ASSOCIATED ENHANCERS.
CHAPTER 3

Introduction

Having identified three RUNX2 associated enhancers from the human genome capable of directing expression to osteoblasts in Chapter 2, I attempted to characterize upstream regulators for each CRE. Fortunately, the transgenic fish lines I created to describe expression patterns also serve as a useful platform to study the underlying biology of their reporter constructs.

Materials and Methods

Site directed mutagenesis. Predicted, conserved transcription factor binding sites in +210RUNX2 were identified via linear alignment to its zebrafish ortholog, +154runx2a, using Clustal W (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Binding sites were individually mutated via PCR. Briefly, each site was converted into a unique restriction site via two parallel PCR reactions with primer pairs that introduced the restriction site and attB sequence flanking each amplicon\(^j\). A subsequent digestion and ligation step produced a mutagenized enhancer competent for Gateway recombination. For -460RUNX2mutTCF and +542RUNX2mutDLX, the mutated enhancer sequences were synthesized (GeneWiz), and cloned directly into pattP-Tol2-egfp as described in Chapter 2.

Drug treatments. To screen for responsiveness to candidate signaling pathways, embryos transgenic for each enhancer construct were treated from 48

\(^j\) Mutagenized +210RUNX2 constructs were constructed by Gina Mahatma.
to 72 HPF with inhibitors of FGF signaling (SU5402; 10µM) BMP signaling (dorsomorphin; 30µM), retinoic acid (DEAB; 50µM), notch (DAPT; 20µM), hedgehog (cyclopamine; 50µM) and calcineurin (FK506; 3µM) signaling and screened for changes in GFP expression. Additionally, +210RUNX2:egfp embryos were treated from 28-32 HPF, or +210RUNX2:mCh embryos from 100-104 HPF, with SU5402 at 10 µM, and -460RUNX2:mCh transgenic embryos were treated from 52-56 HPF with GSK3β inhibitor XV at 5µM, before being harvested for in situ hybridization or Q-PCR analysis. For all treatments, drugs were dissolved in DMSO to make stock solutions, which were diluted into embryo medium; additional DMSO was added to equalize concentration for all treatments.

Heat shock treatments. Embryos doubly transgenic for -460RUNX2:mCh and hsp70:dkk1 were immersed in pre–warmed embryo medium at 37°C for 30 minutes at 52 HPF. Following heat shock, embryos were transferred to fresh embryo medium at 28.5°C and incubated for 3.5 hours before harvesting for analysis. For +542RUNX2:egfp, embryos doubly transgenic with either hsp70:bmp2b or hsp70:chd were similarly heat shocked at 48 HPF, and harvested at 56 HPF.

DNA sequence alignments. Orthologs of human sequences were identified by BLAT. Sequences were downloaded from the UCSC genome browser, curated into FASTA files and aligned using Clustal X (http://www.clustal.org).

Confocal imaging, in situ hybridizations, zebrafish transgenesis and quantitative RT-PCR all performed as described in Chapter 2.
Results – Characterization of +210RUNX2

Identification of conserved predicted transcription factor binding sites in +210RUNX2

The search for sequence features potentially critical for the ability of an enhancer to direct skeletal expression was facilitated greatly in the example of +210RUNX2. Unlike -460RUNX2 and +542RUNX2, +210RUNX2 is deeply conserved, with orthologous sequences alignable from mammals to teleosts (Fig. 3.1). The site of conserved sequences themselves was similarly preserved; all are located in the final intron of either Runx2 or a putatively orthologous gene in more poorly annotated genomes. There are several conserved predicted transcription factor binding sites, including two adjacent inverted binding sites for Ets-related factors (containing a characteristic 5’-GGA(A/T)-3’ core), a binding site for proteins containing a POU DNA binding domain, and one for RUNX2 itself.

Functional testing of conserved predicted transcription factor binding sites in +210RUNX2

In order to test what, if any, function these deeply conserved sequences had with regard to the function of the enhancer, these potentially critical residues were individually and specifically ablated in new transgenic constructs (Figure 3.1) Single insert transgenic lines were constructed as in Chapter 2, and these were crossed onto fish carrying the wild type +210RUNX2 sequence driving the expression of mCherry (+210RUNX2:mCh). Although only representative
microscopic photography is demonstrated below, at least three lines of each transgene have been constructed to confirm the changes in enhancer activity.

The $^{+210}$RUNX2 binding site mediates bone expression.

Removal of the RUNX2 binding site produced an enhancer that failed to direct expression to bone at any time during the first five days post fertilization (Figure 3.2 a-c). However, this altered enhancer was still capable of directing early expression to the branchial arches (Figure 3.2d-f).

The $^{+210}$RUNX2 Ets binding sites mediate branchial arch expression.

Without the conserved Ets binding sites, $^{+210}$RUNX2 is still competent to direct expression to bony tissues, though this does appear to be less robust compared to the activity of the wild type enhancer (Figure 3.3 a-c). Possible position integration effects on the autonomy of the transgene currently confound confirming this quantitatively. More strikingly, however, this altered enhancer failed to direct expression to the branchial arches at 3DPF (Figure 3.3 d-f).

The $^{+210}$RUNX2 POU binding site has no confirmable effect on transgene activity.

Eliminating the POU binding site in $^{+210}$RUNX2 did not compromise its ability to direct expression to the domains of either the branchial arches or osteoblasts (Figure 3.4a,d). eGFP expression driven by $^{+210}$RUNX2mutPOU did appear to be more intense than that typically driven by the wildtype enhancer, suggesting that the conserved sequence might actually have a role in attenuating expression.
+210RUNX2 directed expression is a direct target of FGF signaling.

Because of the presence and functionality of two inverted Ets binding sites in +210RUNX2, I sought to understand what signaling pathways might be mediating gene regulation through those conserved sites. Ets transcription factors are often found to be downstream of the FGF signaling pathway\textsuperscript{168-170}, which, in turn, has a well-appreciated role in skeletogenesis and Runx2 regulation\textsuperscript{k}. The small molecule inhibitor SU-5402 is a potent and selective vascular endothelial growth factor receptor (VEGFR) and fibroblast growth factor receptor (FGFR) inhibitor\textsuperscript{171}. An initial treatment of 10uM from 24-48 HPF reduced +210RUNX2:eGFP mediated fluorescence in the branchial arches (Fig3.5a,b) To confirm this downregulation was a direct result of modulating FGF signaling, I treated +210RUNX2:eGFP transgenic embryos with the Fgf inhibitor SU5402 from 28-32 HPF. In situ hybridization showed specific reduction of egfp expression in the branchial arches (Fig. 3.5c-f), and Q-PCR confirmed a quantitative reduction in transcript levels (Fig. 3.5g). While the ETS binding sites are not absolutely required for the later activity of +210RUNX2 in differentiated osteoblasts, activity of the enhancer was quantitatively decreased by pharmacological inhibition of FGF signaling from 100-104 HPF (Fig. 3.5h), demonstrating continued regulation of enhancer activity by the FGF pathway during osteoblast differentiation.

**Results – Characterization of +542RUNX2**

\textsuperscript{k} Reviewed in Chapter 1
The -460RUNX2 and +542RUNX2 enhancers are less deeply conserved, complicating prediction of transcription factor binding sites. To provide evidence for specific regulatory inputs, I pharmacologically altered activity of candidate signaling pathways in transgenic embryos. Changes in BMP, FGF, retinoic acid, notch, hedgehog, calcineurin, and canonical Wnt signaling had no effect on +542RUNX2 activity (data not shown).

A conserved subelement of +542RUNX2 is sufficient to direct osteoblast expression.

To better localize the essential components of +542RUNX2, I created transgenic lines containing the most conserved cores of the element (Figure 3.6a; Table 3.1) driving eGFP expression. The more conserved of the two (MC1; Phastcons LOD = 773) directed expression to osteoblasts in a similar manner to the entire element (Figure 3.6b,c). However, bone expression was notably less robust than that driven by +542RUNX2. +542RUNX2MC2:eGFP (Phastcons = 334) expresses in the basihyal and ceratohyal cartilages (Figure 3.6d,e), components of the pharyngeal skeleton, but this does not comprise part of the expression pattern normally dictated by the intact wild type element.

A conserved DLX binding site is necessary for +542RUNX2 activity.

Alignment of the +542RUNX2 enhancer with the orthologous sequence from chicken revealed several conserved predicted binding sites (Figure 3.7a). Initially, I hypothesized that SATB2 might be directly regulating +542RUNX2 because of the generalized delay in bone formation observed in Satb2−/− mice.  

61
However, ablation of this sequence failed to curtail osteoblast expression directed by the enhancer (Figure 3.7b,c). Subsequently, I mutated the core of the Dlx binding site and observed no eGFP expression in 500 embryos injected with the resulting construct, compared to the readily observable mosaic expression regulated by the wild-type sequence (Figure 3.7d,e). Transgenic embryos from +542RUNX2mutDLX:eGFP founders showed no eGFP expression, despite evidence of germline transmission of the transgene (Figure 3.7h) confirming requirement of the Dlx binding site for enhancer activity.

Results – Characterization of -460RUNX2

A drug screen identifies GSK-3β as an inhibitor of -460RUNX2 mediated expression.

Treatment -460RUNX2:egfp fish with a small molecule inhibitor of GSK3β from 48-72 HPF resulted in a broad upregulation of eGFP expression (Figure 3.8a,b). To confirm that this effect is a direct effect of modulating the Wnt pathway, a narrower window of treatment from 52-56 HPF (which is relevant to the initiation of -460RUNX2 mediated expression in the cleithrum; Figure 2.3) demonstrated a similar pattern of upregulation (Figure 3.8c,d). GSK3β is an inhibitory component of the canonical Wnt signaling pathway, but can also function in other signaling pathways. To confirm Wnt regulation of -460RUNX2, we generated embryos doubly transgenic for -460RUNX2:mCh and the Wnt inhibitory protein Dkk-1 under control of the hsp70 promoter (hs:Dkk1GFP)165. A brief heat shock substantially reduced expression of mCherry (Figure 3.8e,f).
**-460RUNX2 requires two conserved TCF/LEF binding sites to direct expression.**

Canonically, the endpoint of Wnt mediated signaling is the recruitment of members of the TCF/LEF family of transcription factors to cognate DNA binding sites\textsuperscript{173}. There are two predicted TCF/Lef1 binding sites in the -460RUNX2 sequence, conserved among mammals (Figure 3.9a,b), so I created a transgene in which these sites had been mutated (-460RUNX2mutTCF:eGFP). We observed no eGFP expression in >500 injected embryos (Figure 3.9d). -460RUNX2mutTCFLEF:eGFP founders showed no eGFP expression, despite evidence of germline transmission of the transgene (Figure 3.9e,f) confirming requirement of the TCF/LEF binding sites for enhancer activity.
Figure 3.1 – Deep linear conservation between +210RUNX2

and other vertebrate orthologs. Alignment to orthologous sequences from other vertebrates reveals conservation of predicted transcription factor binding sites, including RUNX2 itself, a binding site for transcription factor containing a POU domain and a pair of inverted sites for the Ets family of transcription factors.
Figure 3.2 – The Runx2 binding site mediates +210RUNX2 directed bone expression. (a-c) Lateral views of a doubly transgenic +210RUNX2mutRUNX2:eGFP; +210RUNX2:mCh zebrafish show no expression in any bony tissue at 5DPF driven by +210RUNX2mutRUNX2, while this activity is intact in +210RUNX2:mCh. (d-f) Ventral views of a doubly transgenic +210RUNX2mutRUNX2:eGFP; +210RUNX2:mCh zebrafish show both transgenes expressing in the branchial arches at 3 DPF. ba, branchial arches; cb5, ceratobranchial 5; cl, cleithrum; op, opercle
Figure 3.3 – The ETS binding sites mediate branchial arch expression.

(a-c) Lateral views of a doubly transgenic +210RUNX2mutETS:eGFP; +210RUNX2:mCh zebrafish show attenuated expression in bony tissue at 5DPF driven by +210RUNX2mutETS, while this activity is intact in +210RUNX2:mCh. (d-f) Dorsal views of a doubly transgenic +210RUNX2mutETS:eGFP; +210RUNX2:mCh zebrafish show failure of +210RUNX2mutETS, to express in branchial arches at 3DPF, while this activity is intact in +210RUNX2:mCh. ba, branchial arches; cb5, ceratobranchial 5; cl, cleithrum; op, opercle
Figure 3.4 – The POU binding site is not essential for +210RUNX2 activity during embryogenesis. (a-c) Lateral views of a doubly transgenic +210RUNX2mutPOU:eGFP; +210RUNX2:mCh zebrafish demonstrate coexpression in bony tissue at 5DPF by both transgenes. (d-f) Dorsal views of a doubly transgenic +210RUNX2mutPOU:eGFP; +210RUNX2:mCh zebrafish show coexpression in branchial arches at 3DPF by both transgenes. ba, branchial arches; cb5, ceratobranchial 5; cl, cleithrum; op, opercle
Figure 3.5 – +210RUNX2 is regulated by the FGF signaling pathway. a,b) Treatment of +210RUNX2:egfp transgenics with the FGF inhibitor SU5042 from 24-48 HPF resulted in loss of transgene expression in the branchial arches. c-f) As show shown by in situ hybridization for egfp, a narrow window of treatment from 28-32 HPF this loss of expression is specific and direct. g,h) Q-PCR confirmed a quantitative decrease in reporter gene expression following treatment for the same interval, and similarly following a later treatment from 100-104 HPF. Views in a and b are lateral, and in c-f, dorsal, with anterior to the left.
Figure 3.6 – Subcloning of +542RUNX2 localizes osteoblast activity to a 433 bp fragment. a) Genome browser view of the original +542RUNX2 element with tracks indicating the location of the amplicons for subcloning (black) and Phastcon elements upon which those amplicons were designed (brown). b,d) Lateral views of 5DPF transgenic embryos carrying +542RUNX2MC1:eGFP (b) or (d) +542RUNX2MC2:eGFP. c,e) Ventral views of 5DPF transgenic embryos carrying +542RUNX2MC1:eGFP (c) or (e) +542RUNX2MC2:eGFP. ba, branchial arches; bh, basihyal; bs, brachiostegal ray; ch, ceratohyal; cl, cleithrum; de, dentary; op, opercle;
<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>ATGCATTTGCCACACACTGATGTTCACAAAATATATCAGGATATACGGAATGATGATGATGATG 308</td>
</tr>
<tr>
<td>Chicken</td>
<td>ATGC--CCACAGTTATGCTACCCACAAATATGCTAAAGGATATACGGAATGATGATGATGATG 195</td>
</tr>
<tr>
<td>Human</td>
<td>ATGCATTTGCCACACACTGATGTTCACAAAATATATCAGGATATACGGAATGATGATGATGATG 308</td>
</tr>
<tr>
<td>Chicken</td>
<td>ATGC--CCACAGTTATGCTACCCACAAATATGCTAAAGGATATACGGAATGATGATGATGATG 195</td>
</tr>
<tr>
<td>Human</td>
<td>ATGCATTTGCCACACACTGATGTTCACAAAATATATCAGGATATACGGAATGATGATGATGATG 308</td>
</tr>
<tr>
<td>Chicken</td>
<td>ATGC--CCACAGTTATGCTACCCACAAATATGCTAAAGGATATACGGAATGATGATGATGATG 195</td>
</tr>
<tr>
<td>Human</td>
<td>ATGCATTTGCCACACACTGATGTTCACAAAATATATCAGGATATACGGAATGATGATGATGATG 308</td>
</tr>
<tr>
<td>Chicken</td>
<td>ATGC--CCACAGTTATGCTACCCACAAATATGCTAAAGGATATACGGAATGATGATGATGATG 195</td>
</tr>
<tr>
<td>Human</td>
<td>ATGCATTTGCCACACACTGATGTTCACAAAATATATCAGGATATACGGAATGATGATGATGATG 308</td>
</tr>
<tr>
<td>Chicken</td>
<td>ATGC--CCACAGTTATGCTACCCACAAATATGCTAAAGGATATACGGAATGATGATGATGATG 195</td>
</tr>
<tr>
<td>Human</td>
<td>ATGCATTTGCCACACACTGATGTTCACAAAATATATCAGGATATACGGAATGATGATGATGATG 308</td>
</tr>
<tr>
<td>Chicken</td>
<td>ATGC--CCACAGTTATGCTACCCACAAATATGCTAAAGGATATACGGAATGATGATGATGATG 195</td>
</tr>
</tbody>
</table>

**Legend:**
- **SIX1/SIX6**
- **SATB/BRN/AC1/BDPT**
- **CEPBA/IRF**
- **MEF2**

**Images:**
- **(a)** Sequence alignments for SIX1/SIX6 and SATB/BRN/AC1/BDPT genes in different species.
- **(b)** Imaging of gene expression in early stages of development.
- **(c)** Imaging of gene expression in mid-development stages.
- **(d)** Sequence alignments for mutSABT2 and mutDLX genes in different species.
- **(e)** Imaging of gene expression after injection with different conditions.
- **(f)** Imaging of gene expression in fully developed stages.
- **(g)** Sequence alignments for mutSABT2 and mutDLX genes in different species.
- **(h)** Graphical representation of gene expression levels.
Figure 3.7 – Identification and functional testing of conserved transcription factor binding sites in +542RUNX2

A) Alignment of +542RUNX2 and +312RUNX2(gg) along with transcription factor binding sites identified in both by Transfac (blue), Genomatix (orange), or Uniprobe (red) b,c) Representative views of eGFP expression pattern in +542RUNX2mutSATB2:eGFP at 5DPF. d) Alignment of a predicted SATB2 binding site to other vertebrate orthologs. Sequence of mutagenized construct is shown below. e) Following injection of +542RUNX2:egfp, mosaic expression in bones, including the cleithrum of two different embryos (arrows) is readily apparent. f) In contrast, >500 embryos injected with +542RUNX2mutDLX:eGFP showed no mosaic expression. g) Alignment of a predicted Dlx binding site to other vertebrate orthologs. Sequence of mutagenized construct is shown below. h) Presence of +542RUNX2mutDLX:eGFP in non-expressing progeny of injected founder was confirmed by PCR and sequencing of transgene. cb5; ceratobranchial 5; cl, cleithrum; de, dentary; mx, maxilla; op, opercle.
Figure 3.8 – *-460RUNX2 mediated expression is responsive to Wnt mediated signaling.* Compared to control (a), activity of -460RUNX2 is increased by treatment from 48-72 HPF with GSK3b inhibitor XV. (b) A narrower window of treatment (52-56HPF) shows rapid upregulation (d) relative to control (c). -460RUNX2 mediated expression (e) is abolished by ectopic expression of the Wnt inhibitor *dkk1* by heat shock (f). cl, cleithrum; op, opercle
Figure 3.9 – -460RUNX2 regulatory competency requires two conserved TCF/LEF binding sites. a, b) Alignment of vertebrate genomes to conserved TCF/LEF binding sites (a) at chr6:44,835,639-44,835,645(hg19), and b) chr6:44,836,005-44,836,012(hg19) in -460RUNX2. Residues that have been changed in the mutagenized transgene are indicated below the alignment. c) Following injection of -460RUNX2:eGFP, mosaic expression in bone is readily apparent at 5 DPF, seen in the cleithra of multiple injected embryos (arrows). d) In contrast, following injection of -460RUNX2mutTCFLEF:eGFP, no expression was seen in >500 embryos. PCR and sequencing of progeny from injected fish confirmed the presence of the transgene with ablated TCF/LEF binding sites (f), compared to sequencing of the transgenics with wildtype sequence (e).
<table>
<thead>
<tr>
<th>Name</th>
<th>F primer</th>
<th>R primer</th>
<th>Coordinates (hg18)</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+542RUNX2</td>
<td>AGACAACACGGGCTCATCGT</td>
<td>CCCCAAGGGTCTCTGGATTT</td>
<td>chr6:45,945,982-</td>
<td></td>
</tr>
<tr>
<td>MC 1</td>
<td></td>
<td></td>
<td>45,946,414</td>
<td>433</td>
</tr>
<tr>
<td>+542RUNX2</td>
<td>CTGGATGGCCAGAGAGGG</td>
<td>TGGCTTCGATATGCCTCTAGTGA</td>
<td>chr6:45,946,358-</td>
<td></td>
</tr>
<tr>
<td>MC 2</td>
<td></td>
<td></td>
<td>45,946,610</td>
<td>253</td>
</tr>
</tbody>
</table>

Table 3.1. Sequences of +542RUNX2 tested for enhancer activity. Both elements are wholly located within +542RUNX2. The primer sequences are those used to amplify the elements from genomic DNA for testing.
CHAPTER 4

DISCUSSION, SUMMARY AND FUTURE DIRECTIONS
CHAPTER 4

Summary of human RUNX2 associated enhancers

Runx2 is the common denominator in osteoblast development throughout the skeleton, and its levels of expression are critical. As a crucial step in understanding the regulation of the gene, and subsequently the skeleton itself, I have identified distant osteoblast specific enhancers associated with RUNX2 and characterized signaling pathways acting on them. Despite their common feature of directing osteoblast expression, they are strikingly diverse. They are widely spaced across the locus (Figure 2.1), have no obvious sequence similarity to each other, and are conserved across species to varying degrees. While they all direct expression to osteoblasts, they do so with differing spatiotemporal dynamics.

The cleithrum is the first bone to ossify in the zebrafish skeleton and does so intramembraneously. Using it as a proxy for the relative timing of expression onset yields a sequence of +542RUNX2 → +210RUNX2 → -460RUNX2 (Figure 2.3). Whether this is consistent across all bony structures is unclear; it appears that the ability to drive expression to the vertebral arches is exclusively a property of +542RUNX2 (Figure 2.6).

In addition to expression at the resolution of individual bones, subpopulations of osteoblasts express the three transgenes differentially. As is evidenced by study of opercle development in these transgenic lines (Figure 2.4) these enhancers direct expression to different cells within that structure. +542RUNX2:eGFP expression is uniform throughout the opercle and cleithrum
(Figure 2.3). Combining these observations with its unique expression in the vertebral arches and early expression in the cleithrum, it is intriguing to speculate that +542RUNX2-mediated expression is the most ‘fundamental’ of the three characterized enhancers and tied primarily to osteoblast identity itself. The expression pattern directed by +210RUNX2 with respect to the developing opercle is similar to that of the fli1:eGFP transgenic line, which labels all neural-crest derived mesenchyme and may indicate cells that have recently become RUNX2+. Finally, -460RUNX2 expression is relatively strongest in the strut and the leading edge in the fan structure of the bone. These cells are also sp7:mCh positive, indicating their likely active deposition of bone ECM components. The opercle fan structure expands via a banding pattern, so -460RUNX2 positive cells may be those that have committed to remaining in a specific iteration of that process.

Finally, these three CREs do not fit the definition of redundant ‘shadow’ enhancers that reinforce a response to a single inductive event and ensure transcriptional robustness to environmental variability. Rather, they appear to integrate inputs from different signaling pathways to induce or maintain Runx2 expression (Figure 4.1). This observation parallels and complements the diversity of signaling inputs capable of accelerating Runx2 expression and osteogenic differentiation in the literature.
Summary of human RUNX2 associated enhancer activity -- +210RUNX2 directs expression to osteoblasts separably through FGF signaling and Runx2 autoregulation.

That +210RUNX2 shares orthology both in sequence and function with an element similarly placed in the last intron of the zebrafish Runx2 ortholog runx2a (+154runx2a;Figure 2.5) suggests its role in regulating Runx2 activity is ancient, and consequential. It also possesses a modularity competent to respond to FGF signaling to direct expression to the branchial arches (essentially) or osteoblasts (qualitatively), while also possessing a conserved binding site for RUNX2 required to direct bone expression. The involvement for FGF signaling during osteoblast differentiation generally has been discussed in Chapter 1, so it is not surprising that +210RUNX2 directs expression to the calvarial sutures that are so sensitive to that signaling axis\(^1\).

+210RUNX2 also presumably functions as a site for positive autoregulation of Runx2 activity. In diverse biological systems, the existence of a positive feedback loop is an essential step in the creation of switches with an all-or-none ‘digital’ output characteristic\(^{175}\). And where better to place a switch incapable of nuance than at a gatekeeper gene whose expression above threshold is sufficient to completely alter cell fate? Whether the +210Runx2 response to RUNX2 is a required step in the commitment of MSCs or chondrocytes to an osteoblastic fate is unclear. There may be other unannotated positive and negative feedback loops

---

\(^1\) Kague, E. Unpublished observation
involving recruitment of RUNX2 to target CREs, to alter cell fate kinetics.

Analysis of the rat and mouse Runx2 P1 promoters indicated that RUNX2 binding to the 5' UTR coding region of Runx2 was capable of suppressing transcription \textit{in vitro}. However, a reasonable hypothesis is that +210RUNX2 functions to 'lock in' a cell to an osteoblastic fate commitment, due to its inability to direct bone-specific expression without a conserved binding site.

\textit{Summary of human RUNX2 associated enhancers} -- +542RUNX2 directs expression to early osteoblasts.

Dissection of the activity of +542RUNX2 was focused on the two most biologically likely conserved direct upstream regulators. SATB2 is a nuclear matrix attachment protein that also functions as a transcription factor. Satb2\textsuperscript{-/-} mice have generalized osteoblast differentiation delays as well as craniofacial patterning defects. Deletion of the predicted SATB2 binding site from +542RUNX2 did not affect the ability of the enhancer to direct expression to osteoblasts. Although, SATB2 binds the promoters of and upregulates bone marker genes, it is also hypothesized to act as a negative regulator of Hoxa2 expression during osteoblast differentiation. Whether it might be executing a similar role with regards to regulation of +542Runx2 is unclear with respect to current experimental evidence.

Members of the \textit{Dlx} family are dynamically expressed during osteoblast maturation, suggesting roles in different aspects of this process. During skull formation in chick, \textit{Dlx5} is expressed in osteoblast progenitors, specifically in
response to BMP but not FGF signaling, and its expression activates Runx2 and osteogenic differentiation in uncommitted embryonic calvarial mesenchyme. Zebrafish dlx5a is expressed in the cleithrum at least as early as the long pec stage (~42 HPF), consistent both with early expression of runx2b and early activity of +542RUNX2. Coexpression of bmp2a and bmp2b early in the cleithrum is also consistent with a BMP->DLX->RUNX2 signaling axis in these cells. However, +542RUNX2 did not demonstrate response to perturbation by induction of bmp2b or chd via heat shock (data not shown), confounding the impulse to arrive at such a conclusion. While the ablated binding site was identified by multiple algorithms as one similar to others capable of recruiting DLX proteins, transcription factor binding site profiles are famously degenerate, and so it is reasonable that other homeodomain containing proteins could be signaling through +542RUNX2.

Summary of human RUNX2 associated enhancers -460RUNX2 potentially links Wnt signaling, Runx2 regulation and variation in common skeletal phenotypes and diseases.

Although experimental and clinical data indicate that gross aberrations in Runx2 expression cause skeletal disorders, smaller individual variations in Runx2 dosage might be responsible for differences in variation of non-pathologic skeletal phenotypes or susceptibility to disease. A cluster of SNPs associated with skeletal conditions (bone mass density (BMD) and osteoarthritis (OA), and height in three different populations) by genome wide association studies cluster
around the Wnt responsive enhancer -460RUNX2 (Figure 4.3; Table 4.1) No other SNPs associated with skeletal phenotypes are located in the remainder of the RUNX2 locus. Wnt signaling has been well implicated in affecting BMD, although the precise mechanism of that effect is not clear in the literature. Some evidence suggests the effect of Wnt signaling on bone mass is indirect, mediated by serotonin secretion by neuroendocrine cells of the gut\textsuperscript{179}, although this has been disputed\textsuperscript{180}. The presented data strongly support a direct role for Wnt signaling in osteoblasts, acting via transcriptional regulation of *Runx2*.

The location of -460RUNX2 suggests that variations in the enhancer itself alter the risk of low BMD and OA and influence height through changes in RUNX2 expression. Interestingly, in addition to its positive association with BMD, the canonical Wnt pathway has been implicated in increased osteoarthritis risk\textsuperscript{181,182}, as has increased *RUNX2* expression\textsuperscript{183,184}. Therefore, sequence variants in the population may affect either the basal activity of -460RUNX2 or its responsiveness to Wnt signaling, accounting for the genetic associations with both of these skeletal phenotypes. An intriguing possibility is that two alleles at a single location could lead either to increased enhancer activity and increased arthritis risk, or decreased enhancer activity and increased risk of osteoporosis.

*Runx2 expression modulation as a source of evolutionary skeletal diversity.*

Runx2 protein activity is positively correlated to facial length in carnivores, especially domesticated dogs\textsuperscript{185}. This relationship is not generally true among
placental mammals, suggesting that other changes, such as in gene expression levels, are more likely to correlate with intra- or inter-specific variation. The sensitivity of normal skeletal development to precise levels of Runx2 has led to the suggestion that alterations in Runx2 activity provide a mechanism for skeletal evolution, acting as a ‘tuning knob’ to either accelerate or delay osteoblast differentiation during development. Following assembly of the Neanderthal genome sequence, the RUNX2 locus was identified as one of the regions with the strongest evidence of positive selection in the evolution of modern humans. Specifically, the 3’ end of RUNX2, encompassing +210RUNX2, shows a deficit of derived alleles in modern humans (Figure 4.3a). No fixed differences in the RUNX2 coding sequence are present between Neanderthal and modern humans, suggesting that the positive selection has acted on changes in regulatory sequences. Comparing the human +210RUNX2 sequence to other primates identifies three derived, human specific SNPs that could potentially link this element to the evolution of the human skeleton (Figure 4.3b). Interestingly, many of the differences between the skeletons of Neanderthal and modern humans—clavicular morphology, frontal bossing of the skull—are similar to the differences observed in cleidocranial dysplasia, which is caused by a Runx2 gene dosage defect.
Future directions

Future investigation of the functional consequences of specific sequence alterations in the RUNX2 enhancers will shed light on the role of its regulation in development, evolution, and disease. An obvious question resulting from the screen in Chapter 2 is the thoroughness of it. While it is likely impossible to ever know the cis regulatory architecture of single gene in a complex eukaryotic genome has been exhaustively annotated, some potential experiments in other model systems present themselves to address this question. However, this will require divesting ourselves of the zebrafish model. While it has been shown to be an effective system for evaluating the regulatory potential of discrete elements in the human genome, the existence of two runx2 genes as well as the current size limitations of BAC-mediated transgenesis make the fish a poor choice to study the intact human Runx2 locus. To try to get a broader locus-wide view of Runx2 enhancer dynamics during development, we must turn to a system with a more similar Runx2 structure, namely the mouse.

Although not discussed in this document, it is not difficult to obtain a population of cells uniformly positive for the Runx2 transgene from early embryos. This process necessitates enzymatic digestion followed by flow-cytometry sorting to derive an enriched population of transgene expressing osteoblasts at a relatively discrete stage of development. Sorting based on multiple colors/transgenes can further refine this staging. Creating one or more mouse transgenic lines using either the +542RUNX2 or -460RUNX2 elements
would, assuming expression patterns are similar in the mouse, permit exploration for bone specific enhancers in early osteoblasts. This could be done by looking for enrichment for histone modifications associated with regulatory sequences, or for regions directly associated with RUNX2 itself. Circular chromosome conformation capture (4C) allows to us to ask questions specifically about the dynamics of the Runx2 locus itself. Using one or both Runx2 promoters as ‘bait’, comparing the physical interaction of distal elements with the Runx2 TSS would presumably yield a list of currently unknown cis regulators, as well as informing how they and the currently known Runx2 enhancers function dynamically through osteoblast development.

Additionally, as the number of and knowledge regarding individual RUNX2 cis regulators grows, the genetic tools they offer might be applicable to studying bone biology in other contexts than embryonic differentiation. RUNX2 transgenic fish have already been used in a to study bone regeneration post-amputation as well as suture development in the skull vault. Certainly, how bones heal post-fracture is a robust area of research, and the ability to visualize Runx2 expression during in vivo assays would presumably augment them. Finally, the transgenic lines could be incorporate in a high throughput screen against a pharmaceutical library, enhancing drug discovery for Runx2 expression mediators that may aid in therapies for common skeletal disorders such as OA and osteoporosis.

---

\(^m\) E. Kague, unpublished
Conclusion

Taken together, the data presented in this document provide evidence for direct regulation of Runx2 transcription by biologically important signaling pathways and transcription factors through three independent enhancers. This complex regulatory landscape has allowed the fine–tuning of expression of this critical developmental gene through alterations in enhancer activities. Furthermore, I hypothesize that these alterations have been selected for in evolution, and help account for differences in skeletal morphology among species. This data also supports the model that variations in a distal enhancer of RUNX2 account for genetic associations in the region with height, BMD, and increased OA risk.
Figure 4.1 – Model for integration of multiple signaling inputs at the Runx2 locus. Three identified enhancers at the Runx2 locus are regulated by different upstream factors, and each is capable of interacting independently with the transcriptional start site (dotted lines) to activate gene transcription. Once transcription is activated via one or more external signals, it is stabilized by Runx2 auto-regulation through the intronic enhancer. Downstream, expression of Runx2 in mesenchymal precursor cells of diverse embryologic origins leads to activation of genes necessary for development of osteoblasts.
Figure 4.2 – SNPs associated with skeletal phenotypes and disorders cluster near the Wnt responsive enhancer -460RUNX2. A genome browser view of the human RUNX2 locus interrogated for regulatory activity shows the location of all SNPs associated with a human phenotype by genome wide association studies. SNPs associated with skeletal phenotypes are listed in Table 4.1; rs1932040 is associated with attention deficit hyperactivity disorder.
Figure 4.3 – Recent positive selection in the human lineage near $+210RUNX2$. (a) Genome browser view of the human $RUNX2$ locus. Signals of positive selection based on scoring of individual SNPs. A negative score indicates more derived alleles in modern humans relative to Neanderthals and is evidence of positive selection. SNP scan data obtained from $^{187}$. (b) Three derived SNPs (indicated by red rectangles) in $+210RUNX2$ are candidate alleles for a recent selective sweep in the human lineage. All three are derived with respect to the ancestral primate state and have not been observed to be polymorphic in human populations.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Phenotype</th>
<th>MAF</th>
<th>Dist from -460RUNX2 (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs556621</td>
<td>Atherosclerotic Stroke</td>
<td>0.3</td>
<td>-241</td>
</tr>
<tr>
<td>rs11755164</td>
<td>Bone mass density</td>
<td>0.4</td>
<td>-196</td>
</tr>
<tr>
<td>rs10948172</td>
<td>Osteoarthritis</td>
<td>0.29</td>
<td>58</td>
</tr>
<tr>
<td>rs9472414</td>
<td>Adult Height (European)</td>
<td>0.22</td>
<td>111</td>
</tr>
<tr>
<td>rs10948197</td>
<td>Adult Height (Korean)</td>
<td>0.34</td>
<td>132</td>
</tr>
<tr>
<td>rs9395066</td>
<td>Adult Height (DECODE)</td>
<td>0.48</td>
<td>260</td>
</tr>
</tbody>
</table>

**Table 4.1 – SNPs associated with human skeletal phenotypes in the human RUNX2 locus.** Minor allele frequency is given in the studied population. Distance from the Wnt responsive enhancer -460RUNX2 is given in kilobases. A negative distance indicates distance 5’ to the element, while a positive one denotes 3’ separation.
References


73. Hassan, M. Q. et al. Dlx3 Transcriptional Regulation of Osteoblast Differentiation: Temporal Recruitment of Msx2, Dlx3, and Dlx5 Homeodomain


165. Lan, C. C., Tang, R., Un San Leong, I. & Love, D. R. Quantitative real-time RT-PCR (qRT-PCR) of zebrafish transcripts: optimization of RNA extraction,


