DYNAMICS OF LONG-RANGE GENE REGULATION AT THE *BMP2* LOCUS

By

Eva Marie Broeckelmann

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Interdisciplinary Studies: Training for a Career in Human Genetics

May, 2013

Nashville, Tennessee

Approved:

Dr. Douglas P. Mortlock

Dr. Michelle Southard-Smith

Dr. Xiangli Yang

To the art of dance,

without which I could have never made it through these past two years of fear

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to all Vanderbilt faculty, students and staff who have taught me and facilitated my time at this wonderful University so far. Amongst all the outstanding faculty I have had the pleasure to meet, learn from, and work with over the years, I specifically would like to thank Dr. Jim Patton for not only a wonderful experience during the IGP year, but also especially for his continued, selfless, and most importantly unbiased support even five years later. Moreover, I owe very special thanks to Dr. Tony Weil, quite possibly the single most inspiring teacher I ever had the privilege to learn from, both as a student and teaching assistant – an invaluable experience I will cherish always -, as he taught me not only gene regulation, but moreover serves as perfect living example for the invaluable virtue of extraordinary attention to detail, and I honestly could not be more grateful for his continued support even in the face of serious adversity. Finally, I cannot thank Dr. Xiangli Yang enough for not only allowing for a great experience rotating in her lab during the IGP year, but also especially for her unwavering support as committee member and true mentor, as she sacrificed numerous hours of her valuable time to talk to me and always honored her role as teacher. I seriously could not be any more grateful for all her support and advice throughout the years, and most importantly, for truly understanding, valuing, and above all accepting me as a human being for who I am -- therefore I thank her from the bottom of my heart.

Of course I also would like to acknowledge and thank all previous members of the Mortlock lab: Kelly and Ron Chandler and Nyk Reed for being such great teachers, role

models, mentors, and inspiration all at once; Dawn Clendenning as most wonderful friend, support, and invaluable help in the lab on a daily basis, and – perhaps most importantly – the dear Yue Hou who more than lived up to her nickname as “lab-mom” in every possible way, as she was not only instrumental in teaching me a large majority of lab skills and techniques, but moreover, was absolutely essential in creating the most wonderfully supportive and inspiring work environment I could have ever hoped for during her time in the lab.

Last but not least, I would also like to thank both the Graduate School and the MPB department very much for being the only entities who graciously contributed to my travel expenses to several scientific conferences and thus greatly facilitated these invaluable opportunities to present my work to a wider audience.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES vii

LIST OF FIGURES viii

Chapter

I. INTRODUCTION 1

Molecular physiology and function of BMP2 1
Bmp2 significance in bone 3
Bmp2, FGF2, and Runx2 interactions during osteogenesis 6
Long-range regulation of gene expression in evolution and development 7
Bmp2 gene regulation 16
Chromosome Conformation Capture 20

II. COMPLEXITY OF DISTANT *BMP2* ENHANCERS IN BONE 25

Introduction 25
Methods 26
ECR1hspLacZ transgenic mouse line 26
ECR1 deletion from 3' lacZ-BAC 26
Purification of BAC DNA for pronuclear injection 27
Generation of transient transgenic mice 28
Yolk sac DNA preps 28
X-gal staining 28
PCR genotyping 29
Fetal long bone osteoblast isolation and culture 29
Results 30
ECR1 enhancer activity alone is insufficient to recapitulate all osteoblast-specific transgene expression of the *Bmp2* 3'lacZ-BAC 30
ECR1 is essential for all 3'BAC specific transgene expression in bone 33
ECR1 transgene expression cannot be induced by FGF2 *in vitro* 37
Discussion 39

III. INTEGRATION OF EVOLUTIONARY CONSERVATION, RT-PCR, 3C, DNaseHS, AND ChIP ON CHIP DATA HIGHLIGHTS NOVEL CANDIDATE ENHANCERS OF *BMP2* IN OSTEOBLASTS 42

Introduction 42
Methods 44
Neonatal calvarial osteoblast isolation 44
Screen of osteoblast cell lines for appropriate *Bmp2* expression 45
Chromosome Conformation Capture 46
ChIP on chip data 49
Evolutionarily conserved elements across the *Bmp2* locus 49
In silico analysis of DNase HS across the *Bmp2* locus 49
Direct comparison of DNase HS sites and evolutionary conservation 50
Nested RT-PCR analysis of lincRNA transcription 50
Sequencing of nested RT-PCR products 52
Results 52
MN7 cells recapitulate the endogenous *Bmp2* expression and FGF2 response of primary calvarial osteoblasts 52
In vitro, osteoblasts exhibit looping interactions between the *Bmp2* promoter and distant loci across the gene desert 54
A lincRNA 120kb upstream of *Bmp2* is actively transcribed in MN7 cells 57
Combination of bioinformatic and *in vitro* evidence supports lincRNA significance and identifies novel loci of interest for putative enhancer function 59
Discussion 61

IV. SUMMARY AND FUTURE DIRECTIONS 64

Analyzing DNase HS in murine osteoblasts 64
Further analysis of ECR1 response to FGF2 *in vitro* 65
Functional tests of putative enhancers 66
Identifying specific transcription factors that act on osteoblast enhancers 67
3C assays on additional (non-) osteoblast cell lines 67
In vivo analysis of looping interactions in different osteoblast populations and other *Bmp2*-expressing tissues 68
Independent verification of looping interactions at the *Bmp2* locus 70
Summary 72
References 74

LIST OF TABLES

Table		Page
1.	Summary of pronuclear injections	33
2.	iCycler real-time PCR protocol	45
3.	TaqMan PCR primers for 3C analysis	48
4.	Nested RT-PCR primer sequences	51
5.	Effective primer combinations for nested RT-PCR	57

LIST OF FIGURES

Figure		Page
1.	Canonical BMP signaling pathway	2
2.	*Bmp2* conditional knockout phenotype	5
3.	Overview of lacZ-BAC transgene scan to identify *Bmp2* cis-regulatory elements	18
4.	Identification of the osteoblast enhancer ECR1	19
5.	3C strategy	21
6.	Expected 3C cross-linking profiles in the presence and absence of looping	24
7.	ECR1 alone does not recapitulate all sites of 3' lacZ-BAC expression in osteoblasts, but exhibits anatomical specificity	32
8.	PCR genotyping confirmation of transgenic embryos	34
9.	Summary of transgene expression patterns after ECR1 deletion compared to sites of expression previously observed with full-length 3' lacZ-BAC	36
10.	FGF2 fails to induce ECR1 transgene expression in fetal long bone osteoblasts	38
11.	qPCR primer design for 3C analysis	48
12.	Primer design for nested RT-PCR	51
13.	Induction of *Bmp2* expression following FGF2 treatment	53
14.	Looping interactions at the *Bmp2* locus in select osteoblast cell lines	55-56
15.	Splice variants of the lincRNA 120kb upstream of *Bmp2* expressed in MN7 cells	58
16.	Integrated 3C, DNaseHS, ChIP on chip analysis, and evolutionary conservation at the *Bmp2* locus	60

CHAPTER I

INTRODUCTION

Molecular physiology and function of BMP2

Bone morphogenetic proteins (BMPs) belong to the TGF-β superfamily of secreted signaling molecules that are essential not only for embryonic development but also postnatal life [1]. Following translation, BMPs are cleaved into propeptide and carboxy-terminal mature domains [2], [3], which then form homo- or heterodimers before they are secreted and bind to heteromeric Type I and Type II BMP receptors. Ligand-bound BMP receptors then phosphorylate Smad proteins, which consequently translocate into the nucleus where they interact with other transcription factors to stimulate the expression of target genes [1] (Figure 1).

Having originated prior to bilaterial evolution at least 700 million years ago [4], BMP signaling serves numerous critical functions amongst modern vertebrates, and *Bmp2* function is no exception. Expressed with remarkable temporal and spatial precision in the majority of cells, *Bmp2* already plays critical roles in pattern formation, morphogenesis, cell fate determination, and differentiation during early embryonic development, which is illustrated by the fact that *Bmp2* homozygous knockout mice die between E7.0 and E10.5 due to failure to close the preamniotic canal and defects in cardiac development [5]. Even beyond their indispensable role in early development, though, proper regulation of BMP signaling continues to be essential throughout all stages of life, and any alteration of the pathway has the potential to have significant

effects on health and disease. For example – amongst others -, genome-wide association studies have identified susceptibility loci for colorectal cancer near multiple genes of the BMP pathway, including BMP2 and BMP4 [6], [7], [8], a genetic association for BMP3 has been implicated in papillary thyroid cancer [9], and BMPR2 mutations are considered to be the most common cause of heritable pulmonary arterial hypertension [10], [11].

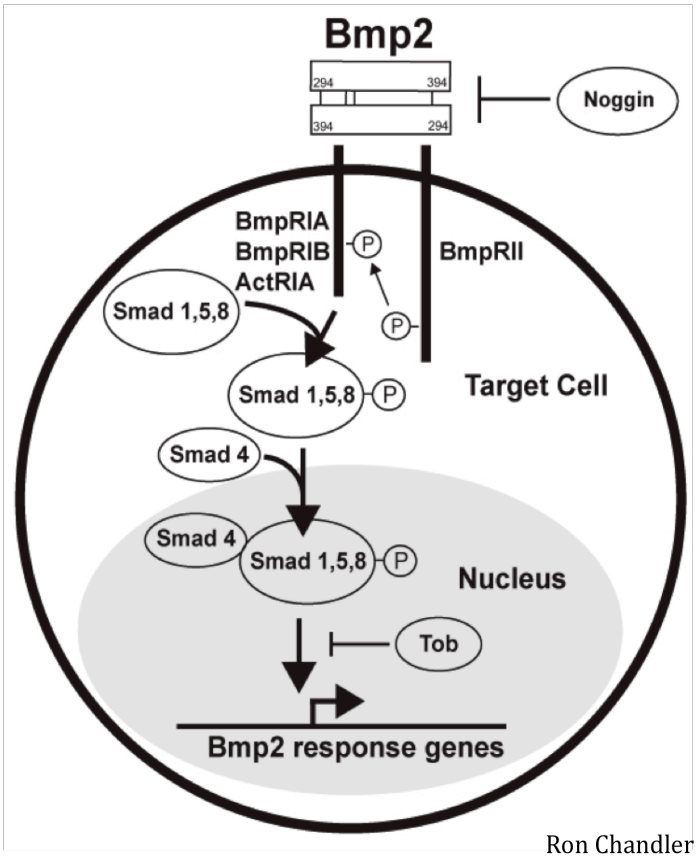


Figure 1. "**Canonical BMP signaling pathway.** BMPs bind heteromeric Type I and Type II receptors. Upon ligand binding, the Type II receptor phosphorylates the Type I receptors, which, in turn phosphorylate and activate downstream Smad proteins. Activated Smads translocate to the nucleus and activate or represses target gene transcription. Extra- and intracellular antagonists, such as Noggin or Tob, regulate BMP ligand binding or effector function." (Figure and legend courtesy of Ron Chandler, 2007 [12])

Bmp2 significance in bone

BMPs were first discovered by virtue of their osteoinductive properties when ectopic bone formation was induced by transplantation of protein fractions from demineralized bone [13]. However, it wasn't until more than 20 years later, that *Bmp2* in particular was identified as one of the main genes responsible for proper osteoblast differentiation and bone formation [14], [15].

During development, each bone in the mammalian skeleton is formed through one of two processes. Endochondral bones, which include the long bones of the limbs, ribs, vertebrae, the medial part of the clavicles, and certain parts of the occipital bone, develop from mesenchymal cells through a cartilage intermediate that is eventually replaced by ossifying bone. The facial skeleton, the majority of the calvarium, and the lateral part of the clavicle, on the contrary, are intramembranous bones, which are formed by immediate differentiation of mesenchymal condensations into osteoblasts without a cartilage intermediate [16]. The differentiation pathway from multipotent mesenchymal cells to mature osteoblasts proceeds through an osteoblast progenitor stage, and each stage along the pathway is characterized by different cellular markers. While the first specification step from mesenchyme to osteoblast progenitor requires expression of the transcription factor Runx2 (Cbfa1), and osteoblast progenitors exhibit low levels of *Col1a1* expression, the commitment of osteoblast progenitors to become functional osteoblasts is stimulated by the transcription factor *Osx.* Once the committed osteoblasts are capable of forming a mineralized bone matrix, they are considered functional osteoblasts and can be identified based on robust expression of mature osteoblast markers, such as *Osteocalcin, Osteopontin* (Spp1/Bsp), and high levels of *Col1a1* [16], [12]. Several of these markers

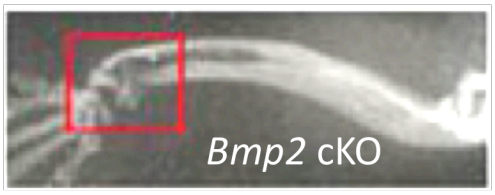
can be activated by signaling pathways downstream of *Bmp2* [17], [18], and the expression of both *Runx2* and *Osx* are stimulated by BMP2, which underscores the crucial role played by *Bmp2* in osteoblast lineage commitment and maturation.

During endochondral ossification, *Bmp2* is expressed in both hypertrophic chondrocytes and the osteogenic perichondrium [19], where its stimulatory effect on bone formation and osteoblast differentiation is balanced by the repressor activity of Noggin and Tob (Figure 1). Mutations in either of these repressors, which mimic human disease phenotypes, illustrate the precarious nature of this feedback mechanism. While *Tob* knockout mice display increased bone mass and bone formation rate [20], [21], overexpression of *Noggin* has the opposite effect, resulting in an osteoporosis phenotype with reduced bone formation and low bone mineral density [22], [23].

Similarly*, Bmp2* conditional knockout (cKO) mice that were generated with a prx1-cre driver, which deletes *Bmp2* from all limb bud mesenchyme derivative cells, illustrate the devastating effect of direct loss of *Bmp2* expression in bone [24]. Although initial bone formation can still proceed in the absence of *Bmp2,* adult *Bmp2* cKO mice not only suffer from spontaneous fractures due to significantly reduced bone mass but are also unable to initiate endogenous fracture repair, which highlights the vitally essential role of BMP2 in the healing process [15] (Figure 2).

Thus considering its complex functions in osteogenesis, it is not at all surprising that *Bmp2* is of great significance for the maintenance of bone health and has been implicated in a variety of human disease phenotypes. Indeed, not only have several association studies identified *Bmp2* as a candidate gene for conditions such as osteoporosis [25], [26], osteoarthritis [27], [28], and ossification of the posterior

longitudinal ligament (OPLL) [29], [30], but the functional consequences of *Bmp2* polymorphisms have also been demonstrated on a molecular level [31]. In addition to its role in the pathogenesis of disease, *Bmp2* has further been recognized as both a drug and a promising drug target for the treatment of bone disorders. For example, recombinant human BMP2 protein (rhBMP2) (INFUSE®, Wyeth Pharmaceuticals) has been FDA approved for use in fracture repair and spinal fusion [32], [33], [34], and statin drugs have been shown to upregulate *Bmp2* expression *in vitro* and *in vivo* [18], which consequently leads to an increase in bone formation and enhanced fracture healing [35], [36], [37], [38], [39], [40]. In view of the potential benefits of such clinical applications, it is imperative that studies of *Bmp2* and its effectors continue in order to further improve our knowledge of regulatory mechanisms that govern its osteogenic function.



Tsuji et al. (2006)

Figure 2. ***Bmp2* conditional knockout phenotype.** In the absence of *Bmp2,* cKO mice exhibit significantly decreased bone mass and spontaneous fractures which remain unresolved due to failure to initiate endogenous fracture repair. Shown is an x-ray of a 10-week old *Bmp2* homozygous conditional knockout animal. (Panel adapted from Fig. 2v of Tsuji et al., 2006 [15])

Bmp2, FGF2, and Runx2 interactions during osteogenesis

Upstream of *Bmp2,* fibroblast growth factor (FGF) signaling pathways, especially mediated by FGF2, play an important role in the activation of *Bmp2.* This effect is supported by the fact that *Bmp2* and *Fgf2* expression peak at the same stage of osteoblast differentiation [41], and *Bmp2* expression is reduced in *Fgf2* null mice [41]. Moreover, it has recently been demonstrated that the stimulation of *Bmp2* expression by FGF2 is mediated by Runx2 both *in vitro* and *in vivo* [42], which suggests the involvement of a feedback mechanism, as *Runx2* expression can also be stimulated by BMP2 [43]. In addition to enhancing *Bmp2* expression, FGFs are likewise known to be important for the regulation of Runx2 during skeletal development [44]. They activate MAP kinase pathways resulting in the phosphorylation of Runx2 at the carboxy-terminus, which enhances Runx2 binding, and consequently increases transcription of target genes [16], [43]. Considering the great complexity of interactions involving Runx2 and its early role in mesenchymal cell specification, it is not surprising that this transcription factor of the runt protein family is referred to as the "master organizer of gene transcription" in osteoblast differentiation [43]. While severe skeletal defects are observed in homozygous *Runx2* null mice that completely lack both endochondral and intramembranous ossification because osteoblasts fail to differentiate [45], [46], the phenotype of *Runx*+/- mice closely mirrors that observed in cleidocranial dysplasia in humans, which is caused by mutations of *Runx2* [45]. These mutant phenotypes and their direct relevance to human health further justify the title of "master organizer" and highlight the importance of understanding the interactions between Runx2 and its various effectors, including *Bmp2.*

Long-range regulation of gene expression in evolution and development

It is estimated that only about 1.5% of mammalian genomes are protein-coding, however 5% of genomic DNA have been evolutionarily conserved ([47] and sources cited therein). This obvious discrepancy naturally raises questions regarding the function of the remaining 3.5% of conserved sequence that do not fall under the traditional definition of a gene, but can be answered by the well-established fact that these conserved non-coding regions can provide the necessary space for important regulatory elements. Considering the need for many cells with identical genetic material to orchestrate the development of an entire organism while differentiating into various distinct tissues with vastly diverse functions, each cell requiring its own unique program of gene expression, such a seemingly disproportionally large amount of regulatory sequences is certainly not at all implausible. While numerous examples of complex gene regulation have been reported in the literature, though, many questions remain regarding the identification and evolutionary significance of regulatory elements as well as their mechanism of action.

Long-range cis-regulatory elements in development and human disease

Cis-regulatory elements are indispensable for faithful maintenance of accurate expression patterns of genes that govern vertebrate development, and the mutation of any given element can cause severe developmental defects, including various human genetic diseases. For example, members of the bone morphogenetic protein family, such as *Bmp5*, play crucial roles in early stages of skeletal development. While null mutations cause a short ear phenotype and significant skeletal abnormalities ([48] and sources cited therein), mutations in single regulatory elements that act over long distances have been

shown to disrupt individual patterns of *Bmp5* expression [49], consistent with the modular function of enhancers at this locus. Similarly, elimination of an intronic *shh* enhancer in zebrafish inhibits expression of this crucial signaling molecule in the notochord and floor plate [50], whereas deletion of a distant limb-specific enhancer element causes loss of *Shh* expression in the developing limb bud and consequent degeneration of the distal portion of the mouse limb [51].

Another regulatory element controlling *Shh* expression in the developing limb and fin is conserved in humans where point mutations in the element, which is located in an intron of the *Lmbr1* gene at a 1Mb distance from the *Shh* promoter, have been associated with preaxial polydactyly due to ectopic *Shh* expression because they seem to disrupt a repressor function [52]. In Van Buchem disease, a non-coding deletion is thought to remove a *SOST*-specific regulatory element, which significantly reduces the level of *SOST* expression in adult bone and causes defects in bone metabolism [53]. While these previous two cases identify a fairly straightforward genotype-phenotype relationship with complete penetrance, quite the opposite is the case in Hirschsprung disease [54]. It is but one example of multigenic inheritance and demonstrates how non-coding mutations in regulatory elements, such as a *RET* enhancer in this instance, can significantly contribute to disease susceptibility, but do not inevitably cause the disease; thus, mutations in regulatory elements exhibit a range of genetic characteristics no different from those in protein-coding regions.

DNA sequence comparison identifies evolutionarily conserved elements

Given the significant role played by regulatory elements in the early development of humans as well as other vertebrates, one cannot neglect the problem of identifying

such elements within the large genomic context. While a universal genetic code and recurring structural features greatly facilitate the identification of protein-coding genes once genomic sequences for a species of interest become available, non-coding sequences with regulatory function cannot readily be recognized based on their nucleotide sequence alone. In order to circumvent this problem and begin to elucidate the informational content of intergenic DNA, sequence comparison between related species is a commonly used approach based on the premise that sequences conserved during evolution are more likely to be functional than those that have not been subject to selective pressure against mutations.

In recent years, rapid advances in genomic sequencing have provided a wealth of data that is maintained and annotated in public databases such as NCBI and the Ensembl and UCSC Genome Browsers, and provide new opportunities for investigators to address questions of conservation in vertebrate and mammalian evolution. However, in order to be able to make efficient use of the increasing amount of sequence information, easy-to-use bioinformatic tools need to be readily accessible. This was accomplished by the development of resources such as MultiPipMaker and Vista, which allow investigators to align and analyze multiple genomic sequences as well as to visualize the results in a user-friendly format [55], [56], [57], [58]. Woolfe et al. and Miller et al. provide just two examples of how genome-wide multiple alignments can be used to identify evolutionarily conserved elements and develop hypotheses about their significance in both protein-coding and non-coding regions [59], [60]. While most sequence comparisons are performed between distantly related species because conservation across larger distances is generally considered more significant, though, Fisher et al. argue that regulatory

elements can be effectively identified based on conservation across moderate evolutionary distances, whereas the comparison of too divergent species might in turn prevent the recognition of conserved functional elements that have accumulated too many mutations over time [61]. Considering the accompanying evidence that human RET enhancers function in zebrafish despite lack of sequence similarity, this theory certainly has to be taken into account when analyzing sequence comparison data.

In vitro and *in vivo* approaches to functional analysis of putative regulatory elements

In light of the large amount of data generated by new *in silico* methods, the quest for efficient functional assays that reliably test the role of evolutionarily conserved elements *in vitro* and *in vivo* is becoming more important than ever. In 1998, studies of regulatory elements controlling *Bmp5* expression in mouse required radiation- and chemical mutagenesis-generated alleles as well as laborious cloning procedures in order to generate overlapping reporter gene constructs that cover the putative regulatory region [49]. Despite yielding reliable results for a single gene of interest, however, such methods are rather impractical for the study of multiple genes requiring the analysis of large numbers of gene fragments. More efficient enhancer screening strategies were first reported in zebrafish by Müller et al., who co-injected putative enhancers with reporter constructs and evaluated the mosaic expression in transient transgenic embryos by accumulated expression maps [50]. This convenient method of creating transgenics, which takes advantage of the large number of embryos produced by zebrafish and the fact that injected DNA fragments tend to integrate at the same genomic location, has since been used successfully in other studies (e.g. [59]), but the significant degree of mosaicism in fish transgenics remained a cause for concern. Consequently, Fisher et al. developed an

alternative transgenic assay in zebrafish based on the Tol2 transposon, which circumvents this problem [61].

In mice, transgenic mosaicism is less common, and soon after the first co-injection studies in zebrafish, a co-injection reporter gene approach using Bacterial Artificial Chromosomes (BACs) was successfully employed by diLeone et al. in order to screen large DNA regions surrounding *Bmp5* for regulatory function [48]. Similarly, the efficient use of BACs has also been demonstrated by Spitz et al. who identified a DNA element that drives expression of several unrelated genes at the HoxD locus in a transposon based targeted enhancer-trap approach [62]. Due to its unusual complexity, the regulatory landscape surrounding the HoxD cluster has inspired numerous studies, which are not limited to the BAC methodology mentioned above, but have also given rise to less conventional strategies such as the STRING approach, which is designed to induce large chromosomal rearrangements by a combination of meiotic and targeted recombination [63]. Although the STRING approach is significantly more time-intensive and best suitable to answer very specific types of questions concerning large regulatory control regions, an important common feature of the methods mentioned here is the fact that they avoid the manipulation of embryonic stem cells.

In addition to *in vivo* studies, various *in vitro* assays can also represent a very efficient and powerful means to analyze potential regulatory function of evolutionarily conserved elements. This has been successfully demonstrated in a study by Holohan et al. that utilized ChIP-array analysis in order to identify CTCF binding sites in the Drosophila genome [64] as well as by Grice et al. who used luciferase and electrophoretic mobility shift assays to test putative RET enhancers [65]. Ultimately, these examples illustrate the

vast array of functional assays available to date and hopefully underscore the importance of selecting the most efficient approach suitable to the particular problem at hand and system of study, which often includes a combination of methods *in silico, in vitro,* and *in vivo.*

The role of cis-regulatory elements in genome architecture

As more and more vertebrate genomes were sequenced, whole genome comparisons between species became feasible, and a comprehensive study comparing the human and pufferfish genomes performed by Woolfe et al. provided valuable insight into the role and characteristics of conserved non-coding sequences [59]. Particularly striking is the trend that the great majority of sequences that have been extremely highly conserved across this large evolutionary distance are found in close proximity to so-called trans-dev genes that play important roles in transcription and developmental regulation. Moreover, many of these trans-dev genes are located in gene deserts, whose existence can be rationalized by the presence of cis-regulatory elements, which exert selective pressure against the collapse of such intergenic regions. In fact, this hypothesis is strongly supported by the findings of Nobrega et al. who showed that several evolutionary conserved elements located in the gene desert surrounding the human DACH gene exhibit enhancer function *in vivo* [66]. When such genomic regions surrounding trans-dev genes are compared across different vertebrates, however, it is important to note that while the order of conserved elements as well as the physical linkage to neighboring genes are generally conserved between human and teleost fish, the absolute distances between two genes, that can often lie several hundred kilobases apart in humans, are significantly decreased in less complex vertebrates [66], [59]. A

comparative study of genomic architecture in *Drosophila* and *C. elegans* by Nelson et al. showed that this general trend even holds true in invertebrates where intergenic distances are positively correlated with a gene's regulatory complexity. However, in absolute terms, more complex animals have significantly larger gene loci of high regulatory complexity than less complex organisms [67].

Genome complexity, conservation, and regulatory functions

In order to explain the larger amount of non-conserved non-coding DNA surrounding genes in more complex organisms, one could speculate on possible implications about the evolution of vertebrates and invertebrates. Viewed in this context, the data seems to suggest that the non-conserved regions in more complex organisms appear to be under selective pressure similar to conserved elements because they contain additional regulatory functions; however, the corresponding functional elements are less readily detectable because they are unique to each higher organism. An extensive analysis of evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes by Siepel et al. supports this idea [47]. It not only underscores the significance of regulatory elements by showing that many of the most highly conserved non-coding elements are actually more conserved than the coding regions themselves, but it also demonstrates that the overall fraction of the genome that is conserved as well as the fraction of coding regions among conserved elements decrease with increasing biological complexity of the organism. In fact, according to this study, only 3%-8% of the human genome is conserved (as compared to 47%-68% of the *S. cerevisiae* genome), which is consistent with other reports that describe 5% conservation in mammalian genomes ([47] and sources cited therein). Since the same estimates suggest that only 1.5% of a

mammalian genome is actually protein coding, this leaves a disproportional amount of conserved, non-coding sequences, whose function is still largely unknown, but seems likely to be responsible for specific regulatory functions corresponding to higher biological complexity of these mammalian genomes.

Alternate models for mechanisms of enhancer and repressor function

While investigations are ongoing about the identification and functional characterization of regulatory elements in different genomes, another subject of continued discussion is the mechanism by which enhancers and repressors function to control gene expression. Since various regulatory elements have been mapped either within introns [54], [65], [50], upstream [52], [66], [51], [62], or downstream [48], [49] of the corresponding genes, their location with respect to the gene under control would seem to be relatively irrelevant to its function, if it can be assumed that the same global mechanism applies to all such elements. Furthermore, the underlying mechanism has to be capable of functioning efficiently over large distances, because - while some elements may lie in close proximity to the promoter - it is not at all uncommon for enhancers of highly regulated trans-dev genes, such as *HoxD*, *DACH*, *Shh*, and *Bmp5*, to be separated from their corresponding gene by several hundred kilobases [48], [49], [66], [51], [62]. In one of the most extreme cases, as mentioned above, a regulatory element proposed to repress *Shh* expression in the developing limb even lies within an intron of an apparently unrelated gene, *Lmbr1*, 1Mb upstream of the transcription start site of *Shh* [52].

Both of the two primary models under debate, based on a looping and a scanning mechanism, respectively, fit these general criteria. In brief, the looping model, as further discussed in detail below, hypothesizes that a chromosome's conformation changes such

that long stretches of DNA assume so-called “loops” in order to establish sufficient physical proximity between genes and distant regulatory elements to allow transcription-factor complexes to form between them ([68] and sources cited therein). The scanning model, on the other hand, proposes that transcription factors track along the DNA from regulatory elements until they reach the transcription start site ([69] and sources cited therein), and thus fundamentally differs from the looping hypothesis in that it does not require any direct contact between enhancer-associated factors and the promoter. While there is no general consensus favoring one model over the other, results of independent studies that used an RNA fluorescence *in situ* hybridization technique called RNA TRAP [70] and a Chromosome Conformation Capture (3C) methodology [71], respectively, to examine the spatial organization of the beta-globin locus in mouse, already seemed to favor a mechanism involving close physical contact early on. Since then, this tendency has further been heavily supported by advanced applications of the 3C strategy (i.e. 5C) that show complex three-dimensional organization of entire genomes [72], as well as 3C evidence of *inter*-chromosomal interactions between promoters and enhancers [73], [74]. Neither of these analyses, however, provides evidence that would definitively rule out at least a partial contribution of a scanning mechanism to enhancer function in some instances. Therefore, the only unanimous conclusion seems to be that chromatin structure plays an essential role in the function of regulatory elements [70], [69], [71]; but ultimately, further investigation will be necessary to fully elucidate the mechanistic details of this process and determine to what extent - if at all - a scanning mechanism may still be involved.

Bmp2 gene regulation

Similar to related BMP genes, such as *Bmp5, Bmp4,* and *Gdf6,* [49], [75], [76], *Bmp2* is located in a large gene desert, where it is surrounded by approximately 1.5 Mb of non-coding DNA, which, as described above, is a phenomenon common to many so-called 'trans-dev' genes that play important roles in development and transcriptional regulation. Since they require multiple cis-regulatory elements to control their spatiotemporal expression patterns during development, their presence and functional significance exert selective pressure against the collapse of these large intergenic regions and thus shape the genomic landscape [67]. The indispensable role of these cis-regulatory elements for normal gene function is illustrated by the severe consequences resulting from their mutations [52], [51], [54], [53]. They are remarkable not only due to their large number and high level of specificity, but also in terms of the large distances across which they are known to function. While the analysis of such large DNA regions poses certain difficulties, the use of overlapping bacterial artificial chromosomes (BACs) for the identification of regulatory elements in transgenic assays is a very powerful method that has been proven successful at a variety of gene loci [48], [77], [62].

In the Mortlock lab, Ron Chandler recently employed a BAC scanning approach to survey a ~400kb region surrounding the *Bmp2* gene for regulatory function in transgenic mice [12], [19]. As illustrated in Figure 3, the two BACs chosen for this assay share a ~55kb overlap region, which includes the entire transcription unit, and each extends ~200kb 5' or 3' of *Bmp2.* Each of the BACs was modified by insertion of an IRES- lacZ / Neo (βgeo) cassette into exon 3 of the *Bmp2* gene, which encodes the carboxy-terminal mature domain of the protein, and transgenic mouse lines were

generated by pronuclear injection. Extensive analyses of transgenic embryos from both lacZ-BAC lines demonstrated that transgene expression recapitulates endogenous *Bmp2* expression in a variety of tissues, which could be assigned to one of three regions depending on whether only the 5' BAC, only the 3' BAC, or both BACs were capable of driving expression in a given tissue. An overview of representative results is shown in Figure 3. While a comprehensive overview of the observed cis-regulatory functions across all various *Bmp2*-expressing tissues as previously discussed [19], [12] is not essentially relevant in great detail to this study, it is particularly worth noting here that an interesting partition of transgene expression was observed in bone. While the 3' BAC showed evidence for enhancer function in osteoblast progenitor cells of both endochondral and some intramembranous bones, the 5' BAC was only able to drive expression in hypertrophic chondrocytes.

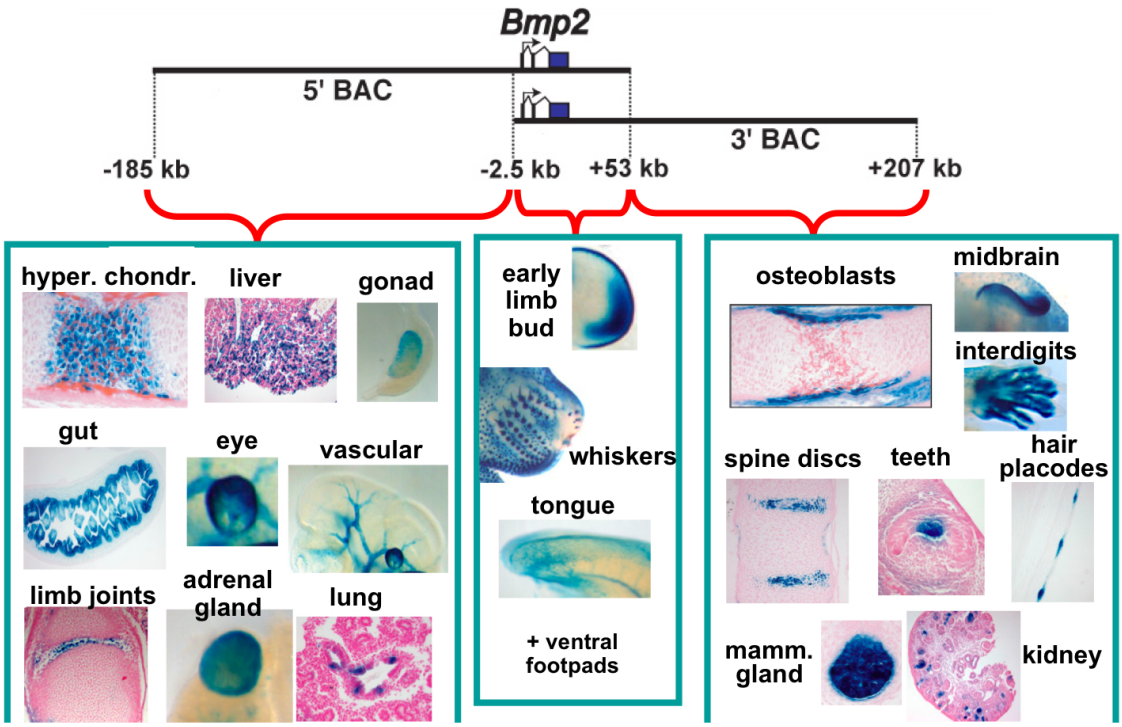


Figure 3. **Overview of lacZ-BAC transgene scan to identify *Bmp2* cis-regulatory elements.** Locations of 3'- and 5'-BACs are shown relative to *Bmp2*, and panels below illustrate select tissues where transgene expression is driven by elements in the corresponding BAC regions. Overall, transgene expression patterns recapitulate most sites of known endogenous *Bmp2* expression. (Figure derived from data by Chandler et al., 2007 [12], [19])

Consequently, a deletion-based approach, as described in Figure 4, was employed in order to narrow down the location of the osteoblast control element to one of four ~40kb regions. This strategy eventually lead to the identification of a 656bp evolutionarily conserved region (ECR1) in the deletion 3 region, 156kb downstream of *Bmp2,* that is able to drive lacZ expression in osteoblast progenitors of embryos carrying a transgene with ECR1 cloned upstream of a hsp-lacZ cassette.

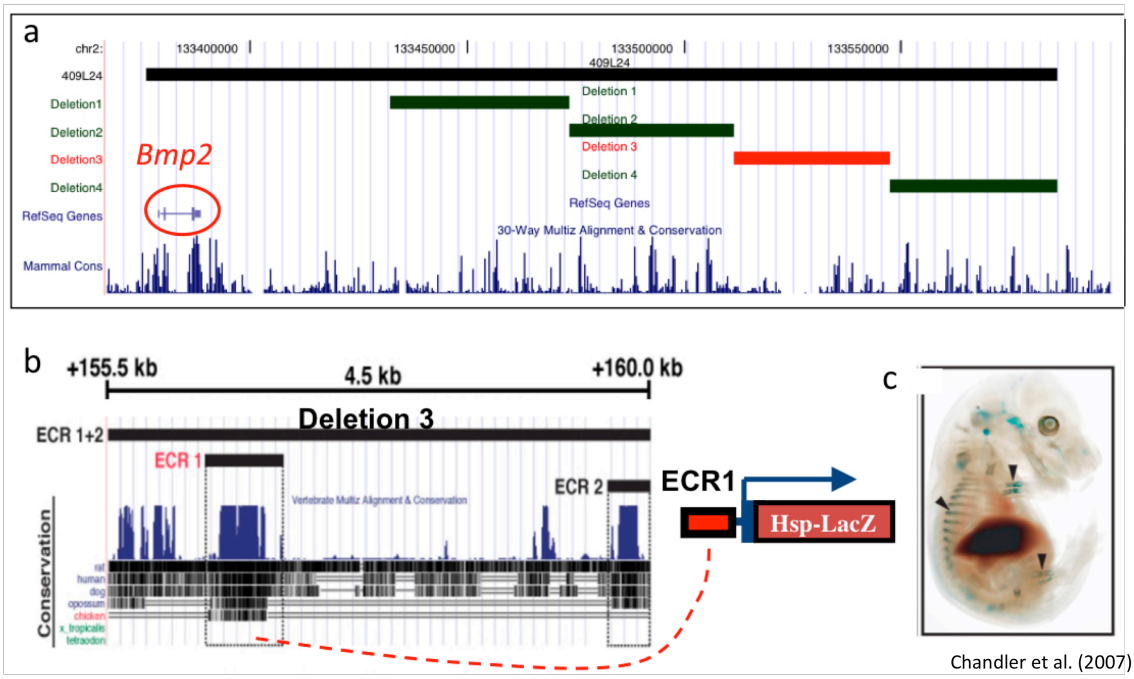


Figure 4. **Identification of the osteoblast enhancer ECR1.** (a) On the UCSC Genome Browser, the locations of the 3' BAC (black) and the 4 deletion regions (green, red) are shown relative to Bmp2. (b) ECR1, a 656bp element, is the only part of the deletion 3 region that is conserved to chicken. (c) When ECR1 was cloned upstream of a hsp-LacZ cassette, it was able to drive transgene expression in osteoblast progenitors. (Panels b and c adapted from Chandler et al., 2007 [19])

Chromosome Conformation Capture

Chromosome conformation capture (3C) was first described by Dekker et al. in 2002 [78] and has since become an established technique for the detection of both intra- and interchromosomal interactions between DNA elements. While large structural polymorphisms and changes in the nuclear localization of chromosomes can be studied by microscopy techniques, conformational changes at the resolution of a single gene locus spanning only a few hundred kilobases cannot yet be visualized by current technology [79]. However, structural dynamics at the gene level are believed to play a vital role in transcriptional regulation. Once general transcription factors have bound to the promoter region, they have to join specific activator and coactivator proteins at the enhancer locus in order to form the preinitiation complex that is required for RNA-polymerase II to be able to initiate transcription of the gene [80]. This type of interaction becomes of particular interest at complex gene loci where numerous tissue- and time-specific enhancer elements can be distributed both upstream and downstream at often large distances from the promoter, because it suggests that the intervening DNA has to form a 'loop' in order to allow direct communication between these distant elements.

Designed to detect this looping mechanism, the 3C technique measures cross-linking frequencies between different regions of a gene locus or between different chromosomes, which is thought to be indicative of functional interactions between cross-linked DNA elements. As illustrated in Figure 5, chromatin in cells of interest is first cross-linked by formaldehyde in order to form a covalent bond between interacting loci and then digested with a suitable restriction enzyme [81]. While the choice of restriction enzyme depends on both the desired resolution and the distribution of restriction sites

across the regions of interest, six-cutters, such as HindIII, EcoRI, and BglII are commonly used for this purpose. In the next step, the free ends of restriction fragments are ligated to one another. In a naked DNA sample, this process would create hundreds of thousands of different ligation products resulting from ligation between any two random fragments. However, the key assumption of 3C is that fragments that are in close proximity to one another, because their physical association has been retained in the cross-linking step, will preferentially ligate to each other. Consequently, the 3C sample will be enriched in the ligation product composed of two restriction fragments that are bound to one another by proteins such as those that make up the preinitiation complex. After reversal of the cross-links, removal of the proteins, and purification of the DNA, ligation products can then be detected by PCR with primers that amplify across the ligation sites.

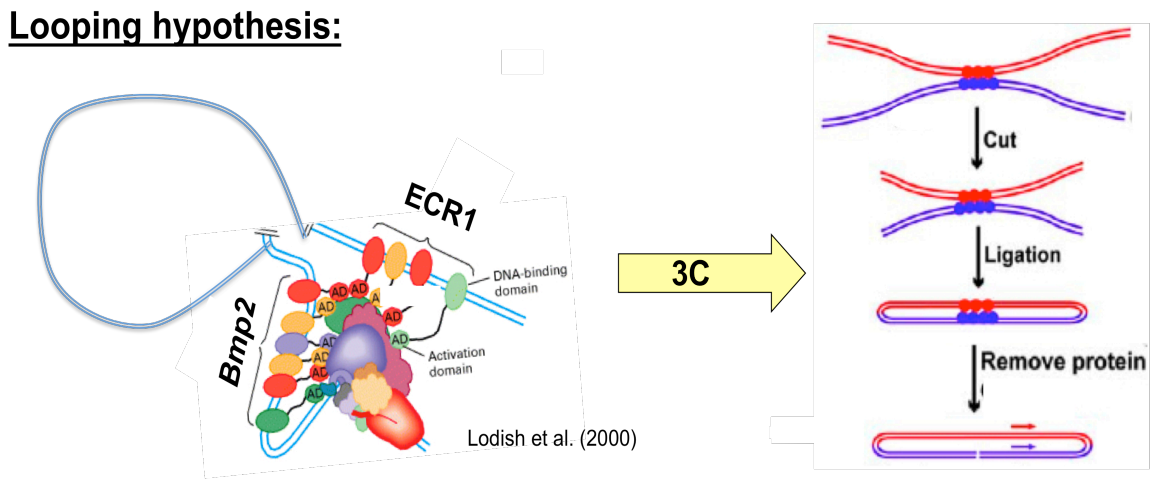


Figure 5. **3C strategy.** After formaldehyde cross-linking, the chromatin is cut with a restriction enzyme and the ends of restriction fragments are ligated. Once protein has been removed, ligation products are detected by PCR amplification across the ligation site. (Figure modified from Lodish et al., 2000 [80] and Lomvardas et al., 2006 [73])

While early 3C studies used conventional PCR and gel electrophoresis to quantify and compare ligation products of interest [78], [79], [71], [73], in 2007, Hagège et al. proposed a TaqMan-based qPCR approach that provides significantly more accurate measurements [82]. Moreover, recent studies have begun to combine competitive PCR on the Sequenom platform, sequencing, and microarray-based comparative genome hybridization to achieve optimal sensitivity and further improve quantification of 3C products [83], [84]. Besides accurate quantification, however, a critical aspect of every 3C assay is the inclusion of appropriate controls in order for 3C results to be interpreted correctly. The importance of: (1) a control template containing equal amounts of all possible ligation products in order to normalize for differences in primer efficiencies, (2) establishing the frequency of random interactions across the genomic locus of interest, and (3) evaluation of interactions at an unrelated locus that is presumed to maintain a constant chromosome conformation *if* multiple different cell types or conditions are compared, have been reviewed in detail by Dekker et al. [81]. Eventually, the objective is to detect local peaks of cross-linking frequency against a background of low cross-linking frequencies at the locus of interest, which provides evidence for functional interactions between DNA elements at the respective locations. Figure 6 provides an example of expected 3C profiles both in the presence and absence of looping.

While 3C had originally been developed in yeast [78], Tolhuis et al. not only applied the method to a mammalian system but also demonstrated its significance for the comparison of gene regulatory mechanisms in different tissues [71], which provides valuable evidence for the dynamic nature of chromosome conformation. Lastly, after 3C had proven effective in detecting intrachromosomal interactions at relatively isolated

gene loci, the remarkable power of this technique has been further validated by its successful implementation for the study of gene associations that span up to 40Mb [85] as well as tissue-specific interchromosomal interactions [73], [74]. More recently, Simonis et al. adapted the general strategy to a 4C (chromosome conformation capture-on-chip) technology for higher throughput applications, which allows for genome-wide analysis of interactions between DNA elements and provides insight into the overall spatial organization within the nucleus [86], [87], [88], [89], while Dostie et al. developed 5C (chromosome conformation capture carbon copy) as further variation and expansion of the 3C approach for large-scale applications that uses ligation-mediated amplification (LMA) combined with microarray or ultra-high-throughput DNA sequencing to quantify large numbers of inter- and intrachromosomal interactions and thus reveal highly complex networks between distant DNA elements across entire genomes [90], [91], [92], [93], [72]; all of which provide invaluable insight into the enormous complexity of nuclear organization, as reviewed in detail by de Wit et al. [94].

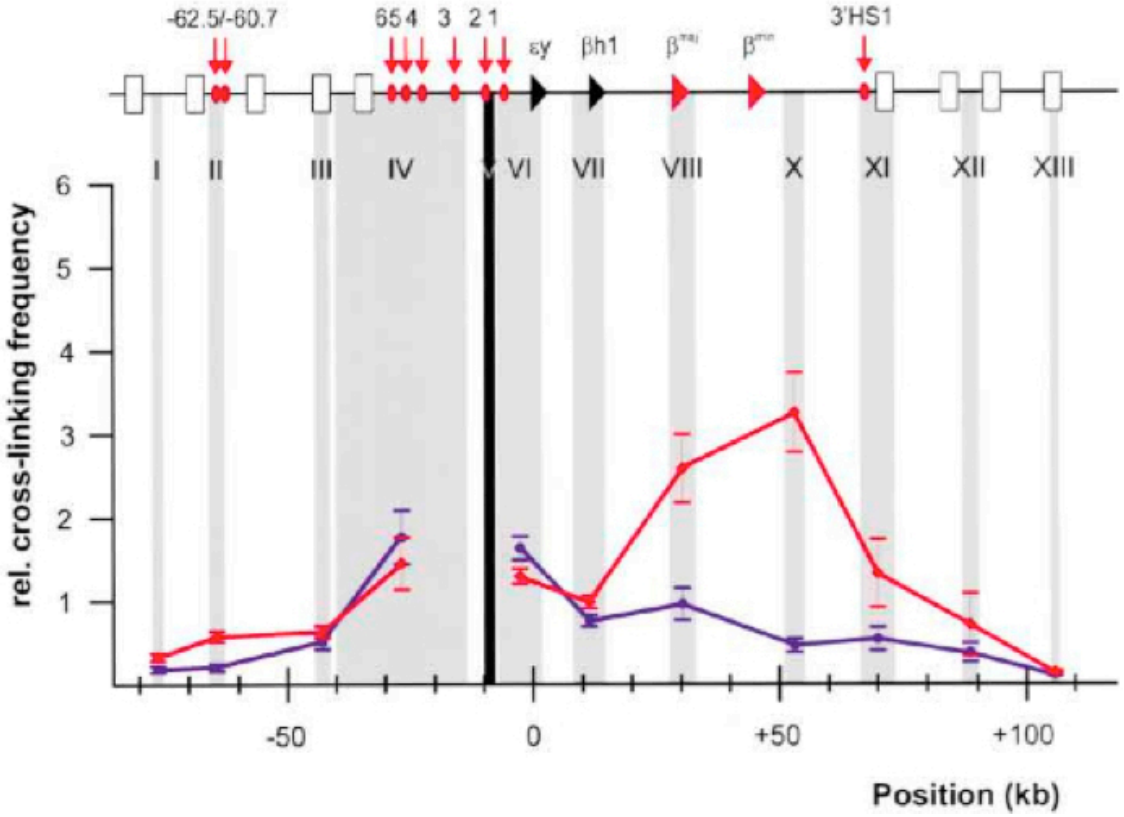


Figure 6. **Expected 3C cross-linking profiles in the presence and absence of looping.** The example above illustrates cross- linking frequencies observed at the β-globin locus. The vertical black bar indicates the constant fragment, whereas grey bars represent candidate interacting fragments. Blue and red curves illustrate cross-linking in brain and fetal liver, respectively, with an obvious peak observed in liver, where β-globin is robustly expressed, whereas the 3C profile in brain is typical for loci where no looping occurs. (Figure from Tolhuis et al. [71])

CHAPTER II

COMPLEXITY OF DISTANT *BMP2* ENHANCERS IN BONE

Introduction

Although transgenic analysis in mouse has shown that the 656bp evolutionarily conserved element ECR1 possesses robust enhancer function in osteoblasts [19], [12], we currently do not have enough information to determine whether it is actually essential for osteoblast-specific *Bmp2* expression in all bones or if additional enhancer elements might also be involved in regulating *Bmp2* in this complex cell type, as has already been shown to be the case in the example of *Bmp5* regulation in bone [95], which exhibits striking similarities. Consequently, if additional enhancers are responsible for driving *Bmp2* expression in other osteoblast populations, it remains to be determined if these elements act independently of one another and complement each other according to the specific population of osteoblasts they function in, or if they are able to interact to some extent such that they either require cooperation or exhibit a level of redundancy that would enable them to completely compensate for one another. This question is of particular interest in the context of upstream signaling pathways that might act on *Bmp2* enhancers in osteoblasts, because studies have shown that *Bmp2* gene expression is stimulated by FGF2 in both endochondral [41] and intramembranous [42], [96] bones *in vivo* as well as in cultured osteoblasts [42]. In regards to ECR1 in particular, results of previous luciferase assays in MC3T3-E1cells describe the direct stimulation of ECR1 enhancer activity by FGF2 treatment [97], which in turn raises the question whether the same

signaling mechanism could be responsible for the activation of separate enhancers of *Bmp2*, or if distinct upstream activation pathways might be necessary to achieve such a remarkable degree of anatomical specificity and thus contribute yet another layer of complexity to the system.

Methods

ECR1hspLacZ transgenic mouse line

The ECR1hspLacZ moue line used for these studies was originally created by Ron Chandler in the Mortlock lab, and carries a plasmid transgene consisting of the ECR1 enhancer cloned upstream of an hsp-LacZ cassette, as previously described [12], [19]. While formerly being maintained in heterozygous condition, though, they have since been bred to homozygosity, such that all embryos sacrificed for the harvest of long bones and subsequent explant culture and FGF2 treatment described here were guaranteed to be transgenic.

ECR1 deletion from 3' LacZ-BAC

The 656bp ECR1 sequence was deleted from the 3' lacZ-BAC transgene mentioned above [12], [19] with the BAC recombineering strategy using GalK selection that has been described in detail by Warming et al. [98]. In brief, the 3' lacZ-BAC was grown in SW102 cells, in which the GalK gene has been deleted from the otherwise intact galactose operon, and ECR1 was replaced with a GalK cassette by homologous recombination. The homology arms used to flank the GalK cassette in this first step were

ECR1+GalK-F: 5'-TCCATGTGTTAGGCAAACACTGTTGTAAGCCACGCTGCTTTTGAATGCTTCCTGTTGACAATTAATCATCGGCA-3' and ECR1+GalK-R: 5'-AGAAAACAAGATTTTAGGGTCTGGAAATTAGACCCAAGATGGAGTATTAGTCAGCACTGTCCTGCTCCTT-3', and recombinant clones were identified based on their ability to grow on minimal media plates supplemented with galactose as the only carbon source. Next, the GalK cassette was removed by homologous recombination with a double-stranded oligonucleotide (top strand = 5'-GGAGACCATTGCTGATGTTCCATGTGTTAGGCAAACACTGTTGTAAGCCAATTTCCAGACCCTAAAATCTTGTTTTCTCATTAAAAAAAAAACTTTTTCC-3') homologous to the immediate flanking sequences surrounding GalK. Successful recombinants in which ECR1 had been deleted seamlessly were eventually selected for on 2-deoxy-galactose (2-DOG) containing minimal media plates with glycerol as the only carbon source.

Purification of BAC DNA for pronuclear injection

BAC DNA was prepared for pronuclear injection by NucleoBond AX 100 column purification (Clontech) according to the manufacturer's protocol and resuspended in microinjection buffer (MIB) (10mM Tris-HCl [pH 7.4]; 0.15mM EDTA [pH 8.0]). Resulting DNA samples were quantified by band densitrometry following restriction digest and pulsed-field gel electrophoresis, and temporarily stored at 4°C until being diluted to 1ng/µl in MIB and filtered through 0.22µm Millex-GV syringe filters (Millipore) the day before pronuclear injection.

Generation of transient transgenic mice

Twelve rounds of pronuclear injections and oviduct transfers were performed by the Vanderbilt Transgenic Core Facility, where BAC DNA was injected as uncut, circular molecules into C57BL/6J x DBA/2J F1 hybrid embryos. Foster dams were sacrificed and embryos dissected at E15.5.

Yolk sac DNA preps

Genomic DNA of all harvested embryos from pronuclear injections was isolated from yolk sacs, which were carefully preserved during dissection, by overnight digestion with 6.25µl Proteinase K (20mg/ml) in 500ml Digestion Buffer (10mM Tris- HCL [pH 8.0], 100mM NaCl, 10mM EDTA [pH 8.0], 0.5% sodium dodecyl sulfate), followed by multiple rounds of (phenol-) chloroform extractions and ethanol precipitation with sodium acetate. DNA pellets were then washed with 70% ethanol before resuspension in TE [pH 7.4] and stored at -20°C.

X-gal staining

By and large, all x-gal staining for lacZ expression was completed as previously described with Wash buffer (0.1M sodium phosphate buffer [pH 7.3], 2mM MgCl, 0.01% deoxycholate, 0.02% Nonidet P-40) [49] and x-gal staining solution (0.8mg/ml x-gal, 4mM $K_3Fe(CN)_6$, 4mM $K_4Fe(CN)_6 \cdot 3\ H_2O$, and 0.1M Tris [pH 7.4]) [77], with the exception that specimens were always pre- and post-fixed in 10% Neutral Buffered Formalin. While all isolated bones - both calvaria and long bones - were treated the same as whole mount embryos of the same age, cultured primary osteoblasts were stained for

48 hours at 37°C, pre-fix time was reduced to 10-15min at room temperature, and all washes were kept to a minimum and performed with extreme care in order to minimize the loss of poorly adherent cells.

PCR genotyping

In addition to x-gal staining, all embryos resulting from pronuclear injections were genotyped by PCR of yolk sac DNAs with primers amplifying the CAM (F: 5'-GGAAATCGTCGTGGTATTCACTC-3'; R: 5'-TCCCAATGGCATCGTAAAGAAC) and LacZ (F: 5'-TTTCCATGTTGCCACTCGC-3'; R: 5'-AACGGCTTGCCGTTCAGCA-3') loci of the BAC transgene multiplexed with *Gdf5* primers (F: 5'-TGGCACATCCAGAGACTAC-3'; R: 5'-TGGAGAGAAATGAAGAGGC -3') as positive control in 10X pH9 buffer (100mM Tris HCl [pH 9], 250mM KCl, 15mM $MgCl_2$). Further PCR amplification across the ECR1 locus was performed with primers ECR1+164bp-F: 5'-CTTGCCTAGAGGCATCTCCA-3' and R: 5'-CAGGAAACTTTTTAAGGCGAAA-3' in order to independently confirm the absence of the osteoblast enhancer from the BAC transgene.

Fetal long bone osteoblast isolation and culture

Pregnant female mice that were mated with homozygous ECR1hspLacZ males were euthanized by CO_2 at E17.5, and pups were dissected out in sterile 1X DPBS without $MgCl_2$ or $CaCl_2$. After dissection of radius and ulna, the epiphyses were removed whenever explants were to be used for prolonged explant culture and osteoblast isolation,

and bones washed twice in 1X DPBS with vortex agitation between washes. The resulting bone fragments were resuspended in digestion solution (1-2mg/ml CollagenaseII) and incubated at 37°C for 30-60 minutes while shaking in order to remove any remaining muscle and connective tissue. As soon as fragments had settled down following this incubation, the supernatant was removed and bone fragments were washed three times with cold recovery media (10% FBS, 1% Pen/Strep, 1% Amphotericin B in low glucose DMEM [Invitrogen#11885-092]) prior to plating approximately 10 long bones per well in a 24-well plate, in order to eventually allow osteoblasts to migrate out of the explant.

For FGF2 treatment experiments, the primary cell cultures were first split after eight days in culture when migrated cells had essentially covered the entire well, and were allowed to continue to proliferate until they reached confluence at passage two. 24hr- FGF2 treatments were then performed using 20ng/ml FGF2 in serum-free DMEM in parallel with serum-starved, untreated control wells (as otherwise described in Chapter III). After 24hrs, RNA was isolated for RT-PCR analysis from one treated and untreated well, respectively, while identical wells, processed in parallel, were stained with x-gal.

Results

ECR1 enhancer activity alone is insufficient to recapitulate all osteoblast-specific transgene expression of the *Bmp2* 3' lacZ-BAC

Extensive analyses of the two *Bmp2* lacZ-BAC lines have previously shown that the 3' lacZ-BAC drives transgene expression in osteoblast progenitor cells in both endochondral and intramembranous bones [12]. Moreover, no osteoblast transgene

expression in either type of bone was observed in transgenic mice carrying the deletion 3 BAC, which in turn was a vital piece of evidence that lead to the identification of the osteoblast-specific enhancer ECR1 located in the deletion 3 region [19]. Since the ECR1hsplacZ transgene is both robustly expressed in long bones (Figure 7b - bottom), and sections through the mandible also revealed lacZ-positive osteoblasts in intramembranous bone [12] (data not shown), it initially appeared to recapitulate endogenous *Bmp2* expression in all osteoblast progenitor cells where the intact 3'BAC drives transgene expression.

Surprisingly, however, when calvaria were later dissected from ECR1hspLacZ embryos at E15.5 and stained for lacZ, it became very obvious that transgene expression is in fact absent from the majority of the bones in the skull, such as the parietal or frontal bones (Figure 7b - top), although embryos from the 3' lacZ-BAC line exhibit abundant transgene expression in these areas (Figure 7a). In fact, the only part of the skull where x-gal staining can be observed in ECR1hspLacZ embryos at E15.5 is a well-defined spot in the back of the head where the occipital bone will eventually be located (Figure 7b). Interestingly, the occiput is the only bone in the skull that is partially formed by endochondral ossification, and thus - from a developmental perspective - is actually more similar to long bones than to any other part of the skull. While we cannot yet exclude the possibility that the observed x-gal staining might be due to ectopic expression or an artifact of trapped lacZ substrate (and in case it is determined to be authentic, the exact location and timing of transgene expression in the base of the skull have yet to be determined by concurrent x-gal and alizarin red staining at different time-points in

embryonic development), the absence of expression in the cranial intramembranous bones at E15.5 is undoubtedly striking and justifies further exploration.

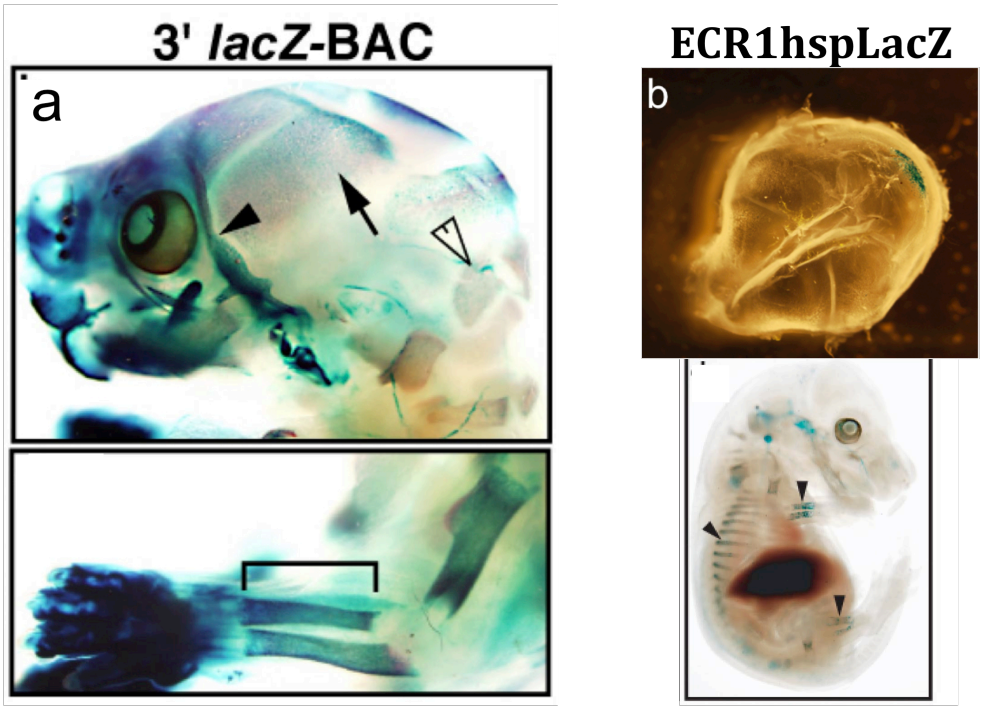


Figure 7. **ECR1 alone does not recapitulate all sites of 3' lacZ-BAC expression in osteoblasts, but exhibits anatomical specificity.** (a) *Bmp2* transgene expression in calvaria and long bones of 3' lacZ-BAC mice, 17.5 dpc. Numerous lacZ-positive cells are present in the parietal (arrows), zygomatic (arrowheads), and occipital bones (open arrowhead). (b) Transgene expression in ECR1hspLacZ embryos, 15.5 dpc. No lacZ-positive cells are found in cranial intramembranous bones. (All but top of panel b adapted from Chandler et al., 2007 [19])

ECR1 is essential for all 3'BAC specific transgene expression in bone

In order to investigate whether ECR1 is solely responsible for driving the transgene expression in osteoblasts that can be observed with the 3' lacZ-BAC, or if additional enhancers might be involved in osteoblast-specific *Bmp2* regulation, we deleted ECR1 from the 3' lacZ-BAC with the GalK recombineering strategy. Once successful deletion had been confirmed, BAC DNA was purified, and pronuclear injections were completed at the Vanderbilt Transgenic Core Facility in order to generate transient transgenic mice that were to be sacrificed and analyzed at E15.5. As summarized in Table 1, even twelve rounds of injections only yielded a combined total of four transgenic embryos, out of which only two were informative at the time-point of interest. Among the other two embryos that could be identified to be transgenic by PCR genotyping, one had died prematurely (i.e. had the appearance of ~E11.5 by the time it was dissected and littermates had reached E15.5), whereas the other one did survive until E15.5, but had obviously only inserted a presumably very small portion of the transgene, since the ECR1 locus at the more distant end of the BAC could not be detected by PCR, and the only LacZ expression that could be observed was ectopic.

Table 1. **Summary of pronuclear injections**

Rounds of pronuclear injections	12
Total # of surrogate dams	17
Total # of embryos harvested	38
Total # of transgenic embryos	4*

* while 4 embryos were genotyped to be transgenic, only 2 of these were actually informative at E15.5

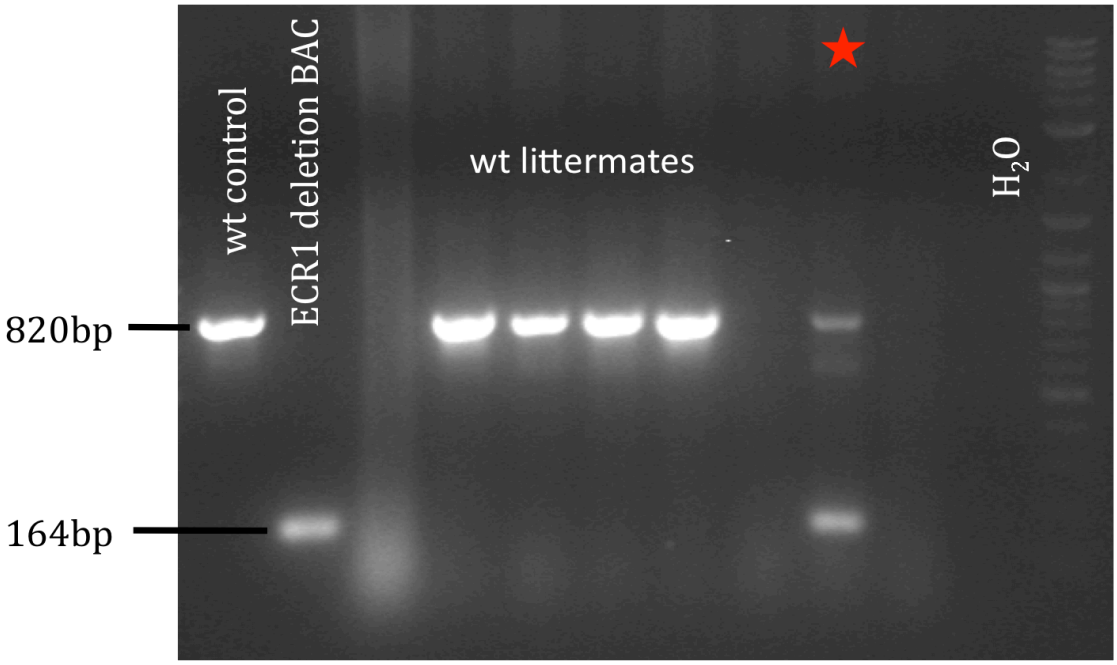


Figure 8. **PCR genotyping confirmation of transgenic embryos.** PCR products resulting from amplification with 'ECR1+164bp' primers as visualized by agarose gel electrophoresis. The first and second lane, respectively, illustrate fragment sizes expected from amplification of the wildtype (wt) genomic locus and ECR1 deletion BAC. Lanes 4-7 show genotyping results of wt embryos, while the red star in lane 9 indicates genotyping confirmation of the single transient transgenic embryo from this litter with PCR products representing both the endogenous, intact ECR1 locus *and* the randomly integrated ECR1 deletion BAC.

The transgene expression patterns observed in the two remaining transient transgenic embryos, whose integration of the ECR1 deletion BAC could clearly be confirmed by genotyping PCR (Figure 8), are illustrated in Figure 9. Overall, apparent mosaicism observed in both instances and the relatively weak x-gal staining, which suggests that the copy number of integrated BACs is presumably quite low and consequently increases the likelihood of their integrity being compromised, preclude definitive conclusions at this point in time. But nonetheless, the consistency of expression patterns amongst these two transgenic embryos, on the other hand, supports their validity. Upon direct comparison to the known sites of 3' lacZ-BAC expression, we detected LacZ staining in the majority of expected locations, except those associated with deletion regions 3 and 4 (Figure 9). While the lack of expression in the brain and spine discs – which had previously been assigned to the deletion 4 region - could possibly indicate fragmentation of the BAC transgene, PCR analysis has confirmed that in addition to LacZ and the CAM resistance gene in the BAC backbone, at least the immediate flanking sequences surrounding the ECR1 deletion are still present in this transgene (Figure 8).

Most importantly for our purposes, though, it has to be noted that no staining could be detected in any osteoblasts, neither in long bones nor calvaria, which not only mirrors the results previously obtained with the Deletion 3 BAC, but also suggests that ECR1 is indispensable for LacZ transgene expression in all osteoblasts, although in isolation, it is only sufficient to drive expression in a defined subset of primarily endochondral osteoblasts. Consequently, these experimental results further support the hypothesis that *Bmp2* expression in bone is regulated by multiple enhancers, which are likely to serve specialized roles in distinct osteoblast populations, although not all are

capable of functioning autonomously. Instead, the evidence seems to suggest that all may actually be co-dependent on ECR1 to some extent, since loss of ECR1 abolishes transgene expression in all osteoblasts, including intramembranous populations, such as those in calvaria, where ECR1 alone exhibits no activity.

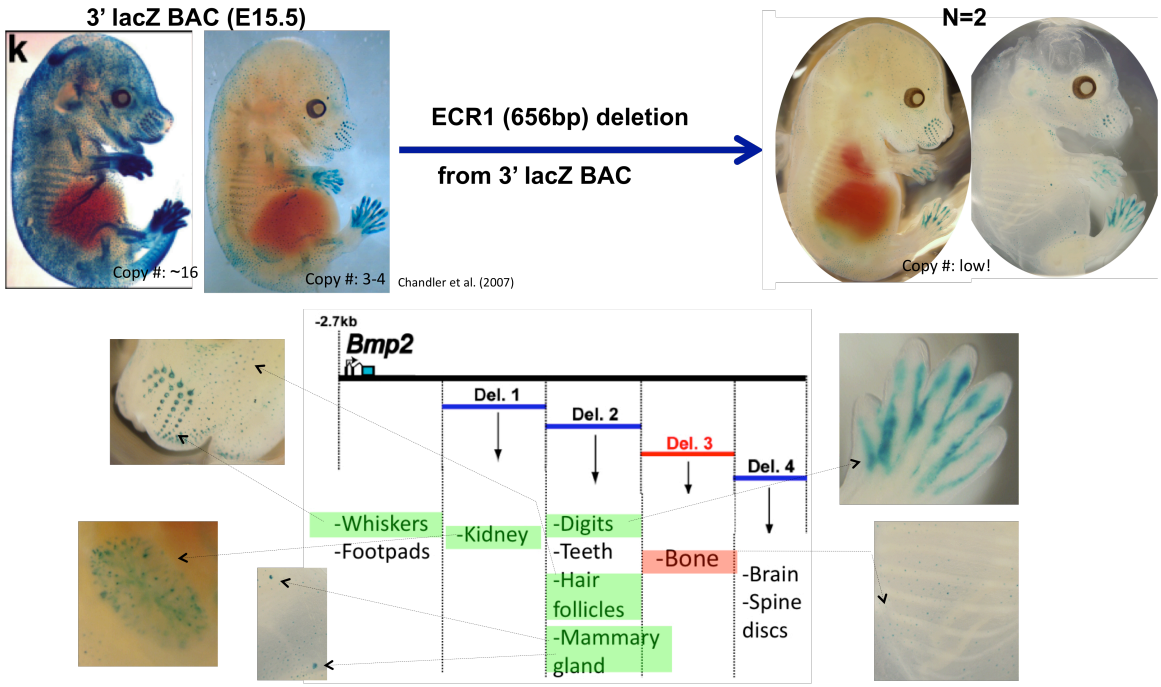


Figure 9. **Summary of transgene expression patterns after ECR1 deletion compared to sites of expression previously observed with full-length 3' lacZ BAC.** The bottom diagram lists sites of LacZ expression at E15.5 that have previously been associated to the indicated regions of the 3'lacZ BAC. Highlighted in green are those anatomical sites where transgene expression was similarly observed in transient transgenic ECR1 deletion BAC embryos at the same time point, whereas the enlarged rib image on the bottom right illustrates the complete absence of observable LacZ expression in bone. Whole mount images of the two transient transgenic ECR1 deletion BAC embryos (top right) in direct comparison to representative embryos from the intact 3'lacZ BAC transgenic line (top left) also underscore the difficulties of mosaicism and presumably low copy number that were encountered with the ECR1 deletion BAC transgenesis. (top left images of intact 3'lacZ BAC embryos adapted from Chandler et al., 2007 [12], [19])

ECR1 transgene expression cannot be induced by FGF2 *in vitro*

In order to test whether ECR1 enhancer activity can be stimulated by FGF2 and thus could also be involved in the induction of *Bmp2* transcription by this signaling pathway, we harvested and cultured long bones from ECR1hspLacZ transgenic embryos at E17.5 as described above until primary osteoblasts had migrated out of the explants and expanded sufficiently for FGF2 treatment to be initiated. After 24hrs, RNA was isolated for real-time RT-PCR analysis of both *Bmp2* and LacZ expression. As expected, *Bmp2* expression was greatly upregulated by FGF2 (Fig 10A, top), and the magnitude of induction was actually greater than that observed in MN7 cells and calvarial primaries (Figure 13). On the contrary, the already very low level of LacZ expression (Figure 10A, bottom - note different scales on the y-axis compared with the top graph) on average even decreased further with FGF2 treatment, which was consistent with the complete lack of observable x-gal staining in additional wells that had been treated in parallel.

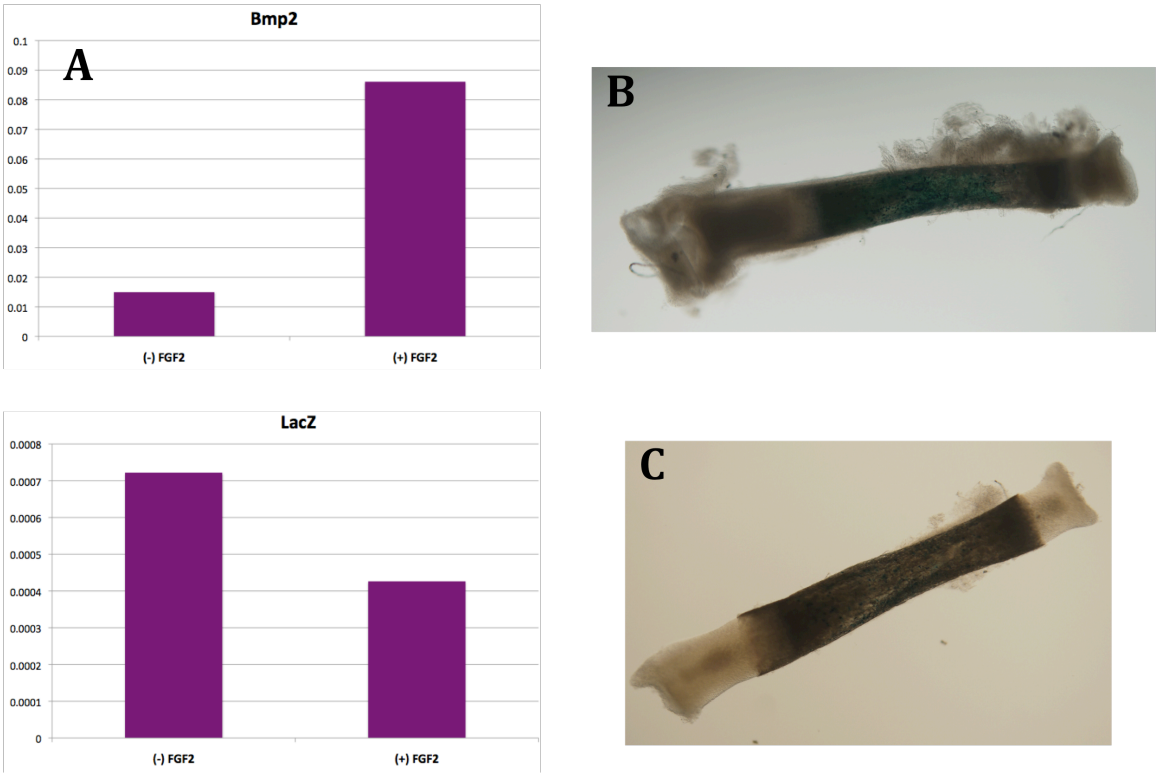


Figure 10. **FGF2 fails to induce ECR1 transgene expression in fetal long bone osteoblasts.** (A) *Bmp2* and *LacZ* expression, respectively, observed in ECR1hspLacZ transgenic long bone osteoblasts with (+) and without (-) 24hrs treatment with 20ng/ml FGF2 as measured by real-time RT-PCR and normalized to *Hprt.* (N=4). (B) E17.5 ECR1hspLacZ long bones stained with x-gal immediately after dissection and (C) following collagenase digestion.

Discussion

In light of the evidence presented in this chapter, we argue that ECR1 is an osteoblast enhancer that specifically drives *Bmp2* expression in long bones as well as some facial intramembranous bones, but fails to recapitulate endogenous expression in cranial intramembranous bones. Consequently, this suggests the presence of (an) additional osteoblast enhancer(s) located within the 3'BAC region, which is/are responsible for activation of *Bmp2* expression in distinct osteoblast populations, including, but not necessarily limited to those where ECR1 alone is inactive. Considering the obvious differences between endochondral and intramembranous ossification as well as the high complexity of gene regulation observed at the closely related *Bmp5* locus during skeletal development [95], the notion of two or more independent enhancers being required to control *Bmp2* expression during these distinct developmental processes is certainly plausible.

While, as mentioned above, the small sample size, apparent mosaicism, low copy number, and uncertainty about BAC integrity are certainly cause for concern in regards to the ECR1 deletion BAC transgenesis and ultimately preclude any definitive conclusions until these preliminary findings have been replicated in at least 3 transgenic embryos with robust and unambiguous lacZ expression, the consistency in expression pattern observed between the two transient transgenic embryos to date certainly supports the validity of our preliminary findings. Likewise, the existing results can still be additionally strengthened by further examination of the two transgenic yolk sac DNA samples with additional PCR analyses that might be able to confirm BAC integrity, e.g. by testing for the presence or absence of other unique parts of the BAC transgene, such as the junctions

between BAC backbone and insert. Similarly, transgene copy number could be determined by real-time PCR and would provide further insight into the likelihood of the integrated BAC transgene being intact, as previously discussed by Chandler et al. [99].

Ideally, though, additional pronuclear injections would be needed to further replicate our results such that $N \geq 3$, although judging by the poor yield of the previous twelve attempts, it may prove to be a rather expensive and time-intensive, yet inefficient endeavor if the same procedure is used. While the rate of transgenesis among the harvested embryos was within the expected range, conditions where initial survival rates of the pronuclear injections themselves are only between ~10 – 63% (based on those days for which detailed information was available) in addition to such problems as surviving embryos being lost in transfer, are certainly less than ideal, and suggest that it might be worthwhile to assure all conditions are optimized in order to improve overall efficiency and thus chances of obtaining informative data on future attempts.

The lack of ECR1 transgene expression in response to FGF2 described here in our study of transgenic long bone osteoblasts does not only argue against a direct involvement of ECR1 in the known stimulation of *Bmp2* expression by FGF2 signaling, but hence could also be interpreted as further evidence for the existence of additional osteoblast enhancer(s) which *are* stimulated by FGF2. Nonetheless, we cannot yet exclude the possibility that the seemingly negative results could merely be due to procedural difficulties in the isolation of this cell population, either. In fact, considering that most of the LacZ-positive cells appear to be concentrated primarily in the outermost layer(s) of these long bones, it would be plausible to suppose that the majority of the osteoblasts of interest actually could have already been lost during the initial collagenase

digestion, since a significant difference in staining intensity can be observed between bones that were immediately stained with x-gal after dissection (Figure 10B) and those stained after the suggested collagenase digestion (Figure 10C). Therefore these experiments might very well be worth repeating with modifications to the protocol in terms of the initial collagenase treatment after dissection. Furthermore, one also has to take into consideration that embryos ≥ E16.5 are susceptible to background staining in ossifying bone, and thus the intensity of staining observed in whole mounts and dissected long bones at E17.5 may in fact - perhaps greatly - overestimate the actual level of ECR1hspLacZ transgene expression in these bones at this particular time-point. In the future, this problem could possibly be circumvented by staining at 4°C instead of room temperature, however.

CHAPTER III

INTEGRATION OF EVOLUTIONARY CONSERVATION, RT-PCR, 3C, DNaseHS, AND ChIP ON CHIP DATA HIGHLIGHTS NOVEL CANDIDATE ENHANCERS OF *BMP2* IN OSTEOBLASTS

Introduction

While MC3T3-E1 cells, which are derived from murine calvaria, are commonly used as an osteoblast cell line of choice and have been employed for previous studies of *Bmp2* regulation, it is obvious that the baseline level of endogenous *Bmp2* expression is actually extremely low in this cell line [42], [12]. Moreover, it has to be noted that neither of the primer pairs used to study *Bmp2* transcripts in the RT-PCR reactions referenced above span any introns [42], [12], which means that one cannot exclude the possibility that the observed PCR products resulted from amplification of genomic DNA. In the Mortlock lab, first evidence for the lack of endogenous *Bmp2* expression in MC3T3-E1 cells was accrued when a ChIP on chip pilot experiment failed to detect any H3K9 acetylation of histones at the *Bmp2* locus, although numerous other osteoblast-specific genes that are known to be expressed in MC3T3-E1 cells were clearly marked by this histone modification (S. Pregizer et al., unpublished data). This result was then further verified when MC3T3-E1 culture was repeated under the same ChIP on chip conditions, and RT-PCR revealed no *Bmp2* expression over the course of 15 days (data not shown), thus highlighting the need to identify a more suitable osteoblast cell line for further *in vitro* studies of *Bmp2* regulation.

Amongst the *in vitro* studies designed to explore the *Bmp2* regulatory landscape from new angles, 3C was one of high priority. Especially considering the long distances that separate functional DNA elements at the *Bmp2* locus, we reasoned that the chromosome conformation would need to adapt dynamically to the specific requirements of particular enhancers in different tissues, such that the intervening DNA forms a loop whenever a given element, such as ECR1, joins the preinitiation complex at the *Bmp2* promoter during transcriptional activation in osteoblasts. As reviewed above, 3C is a well-studied technique whose power to detect looping has been extensively validated in a variety of *in vitro* and *in vivo* systems [78], [100], [73], [101], [83], [102], [71], [84] and has not only the ability to provide invaluable insight into the molecular basis of known enhancer function, but might even have the potential to identify the location(s) of novel regulatory elements.

Genomic DNA is associated with histones, which facilitate compact packaging into chromatin fibers within the nucleus. However, chromatin does not exist in one rigid configuration, but in fact its structure is highly dynamic in nature, as it can vary significantly between tissues and at different times in development based on the transcriptional needs of a cell. While chromatin in the condensed state is generally considered to be inactive, it has long been known that functional DNA elements such as actively transcribed genes and the corresponding regulatory elements are found in regions of decondensed, open chromatin, which is necessary for the complex transcription machinery to be able to access the DNA and initiate transcription [103], [104]. Since this open chromatin structure with its relative lack of histones makes DNA more susceptible to digestion by endonucleases such as DNase I, DNase hypersensitivity (HS) assays can

be used as a tool to screen for enhancer elements, which has been successfully employed at several loci of genes that are associated with a complex regulatory landscape, such as the globin loci or TH2 cytokine genes [105], [106].

Similarly, histones are frequently subject to posttranslational modifications with various acetylation, methylation, phosphorylation, or ubiquitination patterns that are considered to be characteristic of the functional role and activity of the associated genomic DNA. Such epigenetic associations can be taken advantage of for the functional characterization of loci of interest by chIP on chip or chIP-seq assays, for example, and has frequently been applied to comprehensive genome-wide analyses [107], [108]. Therefore it seems reasonable to hypothesize that information on both DNase HS sites and histone modifications across the *Bmp2* locus could greatly augment and complement our understanding of the regulatory landscape.

Methods

Neonatal calvarial osteoblast isolation

1-4 day old pups from CD1 mice were sacrificed by decapitation and calvaria stored in cold HBSS after dissection. Once all calvaria from a litter had been collected, HBSS was replaced by digestion solution (2mg/ml Collagenase, 0.5mg/ml Trypsin in HBSS), and fibroblasts were removed from calvaria by an 18-minute incubation at 37°C. Once the supernatant from the first round of digestion had been discarded, digestion was repeated five more times with 15-minute incubations, and 10% FBS was used to inactivate the enzymes after each round. The combined supernatants from the last five

digestions containing calvarial osteoblasts were then centrifuged for 5 min at 1500 rpm, and the pellet was resuspended in α- MEM + 10% FBS, 1% Pen/Strep, and 1% Amphotericin B. Lastly, the cell suspension was passed through a 70μm nylon membrane before being transferred to 10cm culture dishes.

Screen of osteoblast cell lines for appropriate *Bmp2* expression

Primary calvarial osteoblasts, MN7, MC3T3, and U-33 cells were each cultured in α-MEM (+10% FBS) and treated with 20ng/ml FGF2 as described in Figure 13. On Day 0, i.e. at confluence, and in 24 hour intervals thereafter, total RNA was isolated with the 'SV Total RNA Isolation' kit (Promega) according to manufacturer's instructions, and the 'SuperScript® III Reverse Transcriptase' kit (Invitrogen) was used to synthesize cDNA with oligo(dT) primers. Eventually, levels of *Bmp2* and *Hprt* expression were

Table 2. **iCycler real-time PCR protocol**

Cycle 1: (1X)		
Step 1:	50.0 °C	for 02:00.
Cycle 2: (1X)		
Step 1:	95.0 °C	for 08:30.
Cycle 3: (40X)		
Step 1:	95.0 °C	for 00:15.
Step 2:	60.0 °C	for 01:00.
Data collection and real-time analysis enabled.		
Cycle 4: (1X)		
Step 1:	95.0 °C	for 01:00.
Cycle 5: (1X)		
Step 1:	55.0 °C	for 01:00.
Cycle 6: (81X)		
Step 1:	55.0 °C-95.0 °C	for 00:10.
Increase set point temperature after cycle 2 by 0.5 °C		
Melt curve data collection and analysis enabled.		

measured by real-time RT-PCR with SYBR Green on an iCycler (Bio-Rad), with the following primers: Bmp2-F: 5'-AGGCGAAGAAAAGCAACAGA-3', Bmp2-R: 5'-GGGGAAGCAGCAACACTAGA-3'; Hprt-F: 5'-AAGCTTGCTGGTGAAAAGGA-3', and Hprt-R: 5'- TTGCGCTCATCTTAGGCTTT-3' according to the protocol in Table 2.

Chromosome Conformation Capture

All chromosome conformation capture assays were essentially performed as described by Hagège et al. [82] with the following modifications. In order to achieve sufficient digestion efficiency, chromatin was always cross-linked with 1% formaldehyde/10%FBS/DPBS only. Similarly, the more stringent lysis buffer containing NP-40 (10mM Tris-HCl, pH 7.5; 10mM NaCl; 0.2% NP-40; 1X complete protease inhibitor [11836145001 Roche]) was used for a 15-30 minute incubation on ice, and the cell suspension was both thoroughly pipetted up and down and subjected to at least 25 strokes with a dounce homogenizer in order to assure proper cell lysis. Before and after two rounds of digestion with 400 U of HindIII (100U/µl) for at least 8 hours each, 10µl of undigested and digested sample were collected in order to allow for thorough analysis of digestion efficiency. All digestion steps were performed at 37°C in either a microplate shaker at 900 rpm or a more slowly rotating apparatus, since both methods had been shown to achieve roughly comparable digestion efficiencies. De-crosslinking and proteinase K digestion were divided into two separate steps with de-crosslinking at 65°C overnight preceding proteinase K digestion at 55°C. Finally, DNA was purified by sequential (phenol-)chloroform extractions and ethanol precipitation.

qPCR analysis of 3C products and control samples:

Restriction enzyme digestion efficiency was assessed by SYBR Green real-time PCR with primers on opposite ends of HindIII restriction sites, e.g. 'promoterR': 5'-GTCTTGAGGCAGCCAATGGT-3' and 'promoterR_out': 5'-CCAGAATCTCCCTGCTTTCA-3' or 'ECR1L': 5'-CCTCTCGAACTGTTTCCTGG-3' and 'ECR1L_out': 5'-CCCTGGAGATGTTTGCTGAT-3'. If the digestion efficiency was deemed acceptable, the overall DNA concentration in the 3C sample was quantified by SYBR Green real-time PCR with internal primers, such as F: 5'-TTGTTTAGCTCCCCATGTCC-3' and R: 5'-TTGTCCAGTATTGTTGCCAGA-3' at the *Bsp* locus alongside serial dilutions of pure genomic reference DNAs of C57BL/6J and DBA/2J F1 mice obtained from the Jackson Laboratory. Lastly, 3C ligation products were quantified by TaqMan real-time PCR with the probe: 5' -FAM-CTGAGCCCTCCTCCCC-3' and primers specific for HindIII fragments across the *Bmp2* locus (Table 3), which had been designed with Primer Express® software for the 7900HT real-time PCR system. In order to eventually be able to calculate relative crosslinking frequencies across the locus that were adjusted for differences in primer efficiencies, the exact same TaqMan assays were run on control templates that had previously been generated by HindIII digestion and random ligation of 3'- and 5'-BAC DNAs, respectively, in parallel with the 3C samples themselves on every new plate.

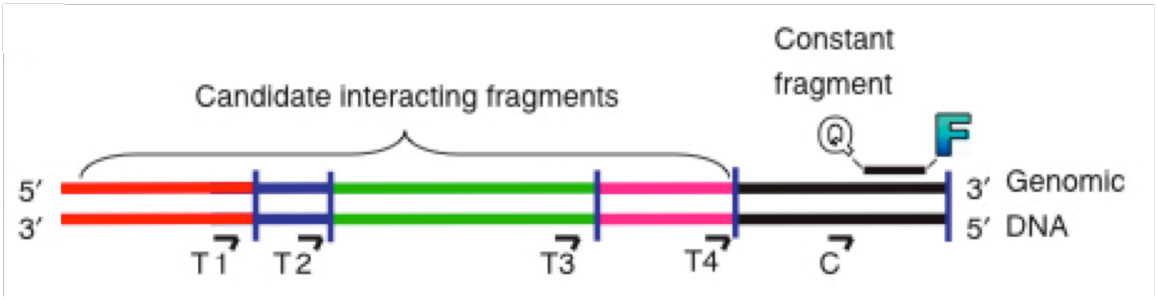


Figure 11. **qPCR primer design for 3C analysis.** Depicted are the locations of the TaqMan probe above the constant fragment (black segment, i.e. *Bmp2* promoter in our case) and candidate interacting fragments (colored segments) with primers (T1-T4) designed to amplify across the future ligation sites. Restriction sites that will be used in the 3C assay are depicted as small vertical bars in blue. (Figure and legend modified from Hagège et al. [82])

Table 3. **TaqMan PCR primers for 3C analysis**

3C_5'fw13	CTAAGGAAAAAAATCCAAGTGAAGCT
3C_5'fw54	TCAGGCTTGAAAGCAAATGTCTT
3C_5'fw70	CACAGAGGCACTAGAGCAAAGC
3C_5'fw89	GATCAGCATCCCTGTGAACAAA
3C_5'fw100	GGAAATTCTTTGCTGTTTTTCTAACA
3C_5'fw130	GAGGAGACATCATATCTTTCAAAAGCT
3C_5'fw170	TTTCCCCTGTCTAGTAGTATGTATAGTTTTTC
3C_3'fw200	CAAATGGGAACAGCAGTTAAGAAA
3C_3'fw219	GGAGGTCTTTGTGTGGGTTTTT
3C_3'fw238	CGAATGAAACTGGAAAATATGCTAAG
3C_3'fw259	CTGCTCCAATGGATAGAAACTTCTT
3C_3'fw281	TCAGTCATCTTGGATTTCATGGA
3C_3'fw302	TCCCAGCAGGGAACAAGCT
3C_3'fw321	CTGGAGCAGGCTTTGAAGCT
3C_3'fw340	GACAAGGCACTGGTCAAGCTT
3C_3'fw360	AAAGGCGGCAAACATAAAGC
3C_3'fw380	TGCAGGAGAGATTTAATAGTCAATGG
3C_promoter-rv188	GTCTTGAGGCAGCCAATGGT

ChIP on chip data

ChIP on chip assays on select osteoblast cell lines and neonatal calvaria were carried out by Dr. Steven Pregizer from the Mortlock lab. The custom oligo chip used for this purpose covers 49 bone related genes with ~40bp density, and chromatin immunoprecipitation was performed as previously described [109] with antibodies against histone modifications acH3K9, meH3K4, and variant H2AZ before samples were sent to Nimblegen for hybridization. Once the resulting data had been analyzed and peaks identified, detailed results were obtained from Dr. Pregizer in order to create custom tracks displaying all peak locations across the *Bmp2* locus on the UCSC Genome Browser.

Evolutionarily conserved elements across the *Bmp2* locus

For the purpose of this study, we focused on PhastCons Vertebrate Conserved Elements from 30-way Multiz Alignment as displayed on the Mouse July 2007 (NCBI37/mm9) Assembly of the UCSC Genome Browser.

***In silico* analysis of DNase HS across the *Bmp2* locus**

Since comprehensive screens for DNase HS across large genomic regions such as the entire *Bmp2* locus are expensive and would require the application of high-throughput technologies, we decided to take advantage of existing DNase HS data on primary human osteoblasts from the ENCODE project, which has recently become available on the UCSC Genome Browser and is referred to as "Open Chromatin by DNaseI HS from ENCODE/OpenChrom(Duke University)". Upon examination of the data across the

portion of the human *BMP2* gene desert that corresponds to our ~400kb BAC region of interest at the murine locus, we chose to focus on those peaks of the osteoblast density signal that reach $y \geq 0.02$, which translates to the top 15 loci with the strongest DNase HS signal. Once the exact coordinates of those peaks were determined with the Table Browser, corresponding loci in the mouse assembly (July 2007 (NCBI37/mm9)) were identified with the former "Convert", now renamed "LiftOver", tool of the UCSC Genome Browser. Finally, a new custom track was created in order to facilitate visualization and comparison of the transposed DNase HS peaks to other features at the murine *Bmp2* locus.

Direct comparison of DNase HS sites and evolutionary conservation

In addition to visual inspection of the relationships and similarities between various features such as 3C looping profile and histone modifications across the *Bmp2* locus, the putative DNase HS sites were also directly compared to the above mentioned Vertebrate Conserved Elements by creating intersections between the two with the Table Browser, in order to highlight loci of direct overlap as further means to test their likely significance and thus validity of our inferred DNase HS predictions.

Nested RT-PCR analysis of lincRNA transcription

RNA had previously been isolated from MN7 cells at confluence (i.e. Day 0) and cDNA synthesized as described in Chapter II for the screen of various osteoblast cell lines. Primers to test for transcription of the ~21kb lincRNA upstream of *Bmp2* (Table 4) were based on the annotated structure of the Ensembl gene ENSMUSG00000086650 and

designed to amplify across each of the four introns as shown in Figure 12. In the first round of the nested PCR, four separate reactions were run in which primer fw1 was combined with each of the reverse primers, respectively (i.e. fw1+rv1, fw1+rv2, fw1+rv3, and fw1+rv4). The second round of PCR then was designed to test all possible nested primer combinations on each of the primary PCR products. All PCR reactions were performed for 40 cycles at 60°C annealing temperature.

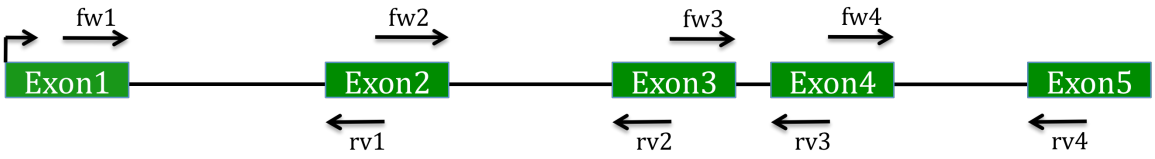


Figure 12. **Primer design for nested RT-PCR.** Shown are the relative locations and nomenclature of all primers tested in nested RT-PCR attempts to amplify transcripts of the lincRNA (ENSMUSG00000086650) upstream of *Bmp2*.

Table 4. **Nested RT-PCR primer sequences**

Bmp2-5'gene_fw1	AAACTGCACCGCCAGAAC
Bmp2-5'gene_fw2	AAGACCACAGTGGGAAATGG
Bmp2-5'gene_fw3	CATCTGAAGCACTCGGTGAA
Bmp2-5'gene_fw4	TGCTTCCTCAATGCTGTGAC
Bmp2-5'gene_rv1	AAAGCAGCTCCTCCATTGTT
Bmp2-5'gene_rv2	TTCACCGAGTGCTTCAGATG
Bmp2-5'gene_rv3	GTCACAGCATTGAGGAAGCA
Bmp2-5'gene_rv4	AACCAGAGCGCAGAGAACAT

Sequencing of nested RT-PCR products

In order to assure sufficient quality and quantity of substrates to be sequenced, the second round of all nested PCRs that had been shown to yield single bands was repeated in eight 50µl reactions each, which were combined and quantified by UV spectrophotometry such that 10µg of each DNA could be run out on an agarose gel and subsequently cleaned up with the 'QIAquick Gel Extraction Kit' (Qiagen) according to manufacturer's instructions. Thus purified DNA samples were then once again quantified by UV spectrophotometric analysis and submitted for Sanger sequencing at the Vanderbilt Sequencing Core with the respective fw primers of all samples as well as the rv primer for the fourth DNA sample.

Results

MN7 cells recapitulate the endogenous *Bmp2* expression and FGF2 response of primary calvarial osteoblasts

Since, as mentioned above, the MC3T3-E1 cell line that has commonly been used as osteoblast model system in culture had been proven to be unsuitable for studies of *Bmp2* expression in our lab, it was deemed necessary to carefully screen alternate cell lines in order to determine which one(s) would best mimic endogenous *Bmp2* expression patterns before proceeding with any further *in vitro* studies. Thus, we selected MN7 and U-33 cells – both murine cell lines derived from bone marrow stromal cells – as potential candidates and rigorously tested these two in comparison to both MC3T3-E1 and primary calvarial osteoblasts for levels of endogenous *Bmp2* expression as well as their response to FGF2. As expected, MC3T3-E1 cells once again failed to express any *Bmp2* transcript

detectable by real-time PCR, and even FGF2 treatment was unable to reactivate any significant amount of expression, which conclusively excluded them from consideration for any further experiments requiring *Bmp2* transcription. On the contrary, FGF2 treatment of both MN7 and U-33 cells significantly increased *Bmp2* expression and thus recapitulated the inductive effect of this growth factor on primary osteoblasts (Figure 13; U-33 data not shown). However, since the overall levels of *Bmp2* expression in U-33 cells (both in absolute copy number and relative to *Hprt)* were several orders of magnitude lower than in MN7 cells or primary osteoblasts (data not shown), MN7 cells, whose osteogenic potential has previously been confirmed by expression of alkaline phosphatase as well as Von Kossa staining [110], clearly stand out as the osteoblast cell line of choice for future *in vitro* studies.

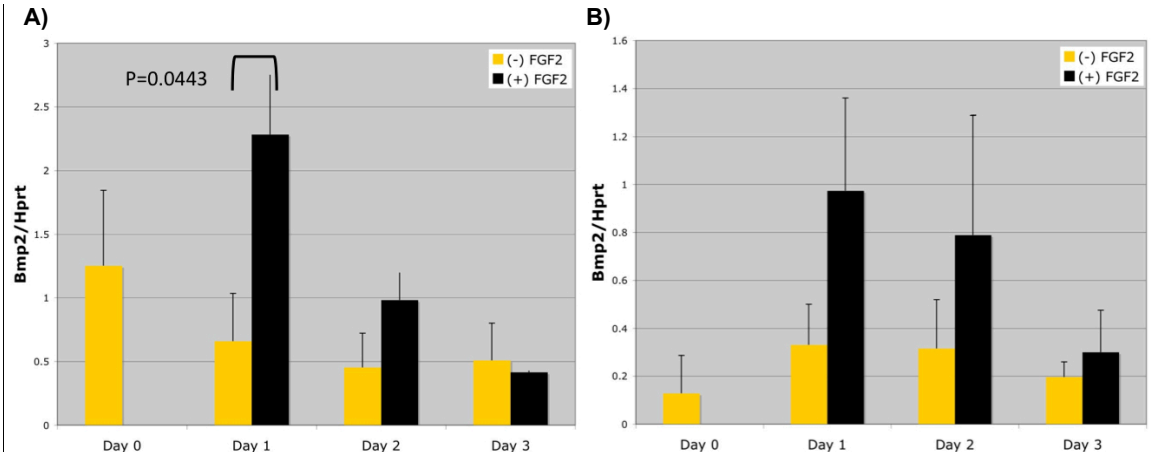


Figure 13. **Induction of *Bmp2* expression following FGF2 treatment.** MN7 cells (A) and primary calvarial osteoblasts (B) were grown to confluence in α-MEM +10% FBS, then medium was replaced with serum-free α-MEM either with (+) or without (-) 20ng/ml FGF2. After 24 hrs, medium was replaced with serum-free α-MEM in all plates, and cells were cultured for 48 more hours. Total RNA was isolated at all time-points, and *Bmp2* and *Hprt* expression were measured by real-time RT-PCR. Shown are the mean and SD of *Bmp2* expression normalized to *Hprt* from triplicate experiments, and p-value is based on paired t-test analysis.

In vitro, osteoblasts exhibit looping interactions between the *Bmp2* promoter and distant loci across the gene desert

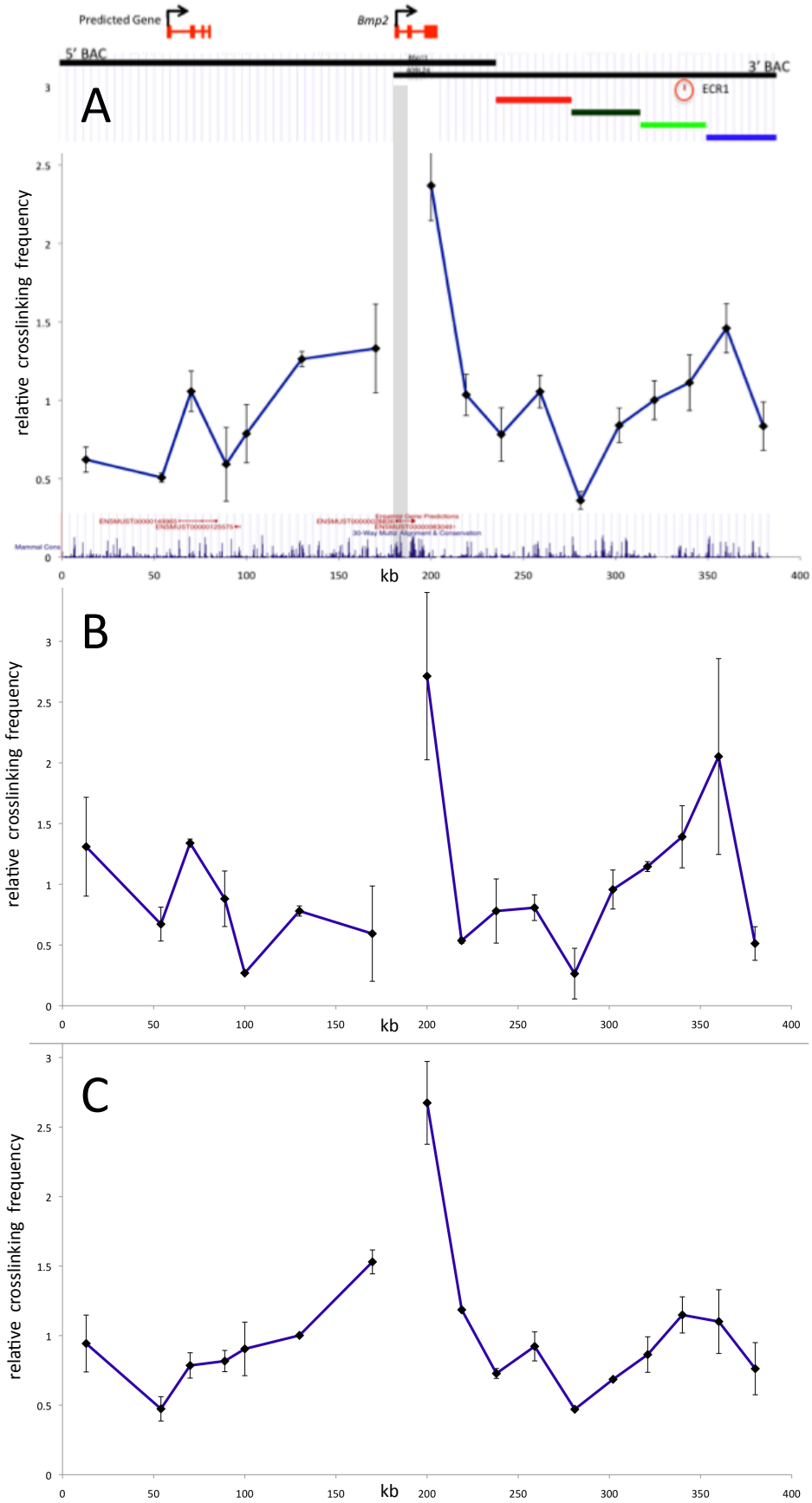
In order to test whether or not ECR1 physically interacts with the *Bmp2* promoter during transcriptional activation, we first performed 3C on MN7 cells, and later repeated the same analysis on MC3T3-E1 cells and FGF2-treated MN7s. Following formaldehyde crosslinking and HindIII digestion, as detailed above, the relative abundance of ligation products of interest was quantified by TaqMan PCR. While the *Bmp2* promoter served as the constant fragment where both the reverse primer ("3C_promoter-rv188", Table 3) and the TaqMan probe anneal, the 17 forward primers it was tested against were designed to cover the entire 400kb BAC region at approximately 20kb intervals. Once relative crosslinking frequencies across the locus had been determined by comparing the abundance of ligation products in the 3C sample to those detected in the 3'- and 5'-BAC control templates, their distribution across the *Bmp2* locus in MN7 cells identified local peaks of interest, indicative of looping interactions with the promoter, in both the 3'- and 5'-BAC regions (Figure 14A).

While an increased crosslinking frequency with downstream loci certainly agrees with prior expectations based on the extensive transgene expression in osteoblasts observed with the 3' LacZ-BAC, it is nonetheless worth noting that the most pronounced local peak (observed with primer "3C_3'fw360", Table 3) actually does not perfectly correspond to ECR1 itself, but is located ~20kb downstream of this known osteoblast enhancer locus. Significantly more surprising, on the other hand, was the finding of a very distinct peak of crosslinking in the 5'BAC region, ~120kb upstream of *Bmp2,* given that the 5' LacZ-BAC is not known to have the ability to drive the same osteoblast-specific expression. However, when the genomic region under this 5' peak of

crosslinking frequency was examined more closely on the UCSC Genome Browser, we found that it happens to fall almost directly on the first exon of a supposed long intergenic non-coding RNA (lincRNA) (i.e. Ensembl gene ENSMUSG00000086650), which not only corroborates the validity of this rather unexpected 3C result, but also directs the attention to previously less scrutinized parts of the gene desert and underscores the value of combining evidence from various genomic features and *in vitro* data in order to identify novel putative enhancer loci.

Having solidly established the looping profile at the *Bmp2* locus in MN7 cells, the next logical step was to test whether the observed interactions represent a permanent characteristic of this locus, or to what extent looping may be dynamic and dependent on the transcriptional state of *Bmp2* under any given circumstances. Thus, the same 3C analysis was applied to both MC3T3-E1 cells and FGF2-treated MN7s, which did indeed support the latter hypothesis of the chromosome conformation's dynamic nature. While FGF2 treatment of MN7s caused a notable height increase of both crosslinking peaks 5' and 3' of *Bmp2* (Figure 14B), the crosslinking frequencies at the exact same loci have diminished so far in MC3T3-E1 cells that they cannot even be distinguished as peaks above the background any longer (Figure 14C), which is consistent with the lack of *Bmp2* expression in this cell line and consequently exemplifies the notion that DNA looping interactions are directly related to if not necessary for long-range gene regulation.

Figure 14. **Looping interactions at the *Bmp2* locus in select osteoblast cell lines.** Depicted are average relative crosslinking frequencies (+/- SEM) observed between the *Bmp2* promoter (grey vertical bar) and 17 HindIII restriction fragments across the 5'- and 3'-BAC regions in (A) MN7s (N=4), (B) MN7s + FGF2 (N=2), and (C) MC3T3s (N=2). Mammalian conservation and Ensembl gene prediction according to the UCSC Genome Browser, as well as the relative locations of deletions 1-4 (red, black, green, and blue bars) and ECR1 are indicated below and above the graph, respectively, in panel (A).


Predicted Gene
Bmp2
5' BAC
3' BAC
ECR1
A
relative crosslinking frequency
kb
B
relative crosslinking frequency
kb
C
relative crosslinking frequency
kb

A lincRNA 120kb upstream of *Bmp2* is actively transcribed in MN7 cells

In light of our 3C data that suggest looping interactions between the *Bmp2* promoter and an intergenic locus 120kb upstream of *Bmp2* that coincides with the first exon of a lincRNA annotated on the UCSC Genome Browser, it was imperative to ascertain whether or not this particular RNA is even transcribed in the same osteoblast cell line that the 3C assays had been performed on in order to be able to assess the relevance of this annotation for our purposes. After the first attempt at RT-PCR with all possible primer combinations (Figure 12 and Table 4) yielded only multiple non-specific bands, we resorted to a nested PCR approach. Using this method, four primer combinations (Table 5) successfully yielded single bands at or very close to the expected size after the second round of PCR.

Table 5. **Effective primer combinations for nested RT-PCR**

Primary PCR Primers	Secondary PCR Primers
fw1+rv2	fw2+rv2
fw1+rv3	fw2+rv2
fw1+rv4	fw2+rv2
fw1+rv4	fw3+rv3

Upon sequencing to confirm the exact identity of these PCR products, we discovered three different splice variants between the second and third exon alone – two of which include an additional exon of slightly varying size that is missing from the Ensembl gene annotation – and one transcript spanning exon 3 and 4, which also slightly

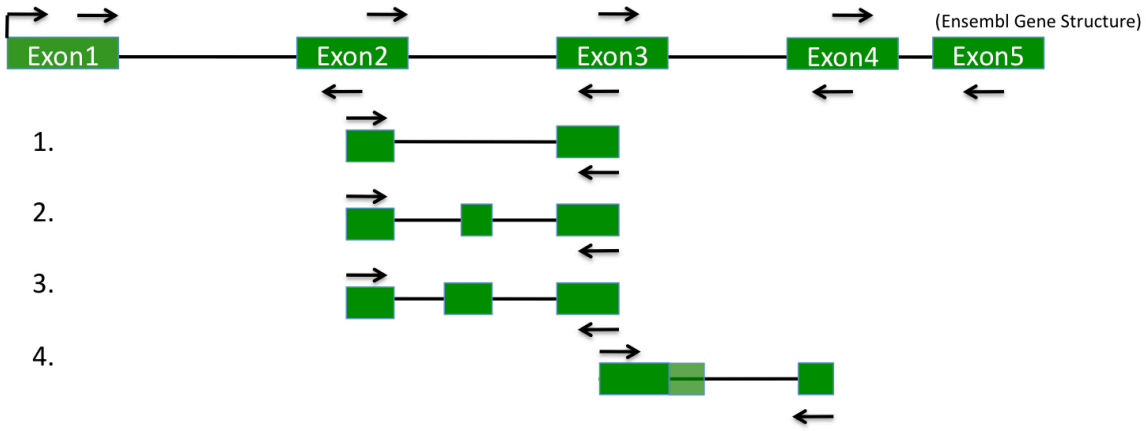


Figure 15. **Splice variants of the lincRNA 120kb upstream of *Bmp2* expressed in MN7 cells.** Schematic representations of sequencing results from the four distinct transcripts that could successfully be amplified from MN7s by nested RT-PCR are shown aligned below the annotated structure of Ensembl gene ENSMUSG00000086650.

deviates from the published annotation by a small elongation of the third exon, however (Figure 15). Although neither exon 1 nor 5 were included in any of the four successfully amplified transcripts, it is by no means evidence of their complete absence in MN7 cells, since all of the primary PCR reactions utilized the fw1 primer located in exon 1, and similarly, two primer combinations included the rv4 primer in the fifth exon as well, which – especially considering the large size of the introns that span several kilobases – would be virtually impossible to yield any useful PCR products from genomic DNA alone.

Combination of bioinformatic and *in vitro* evidence supports lincRNA significance and identifies novel loci of interest for putative enhancer function

As described above, DNase HS sites from human primary osteoblasts were mapped onto the murine *Bmp2* locus, whereas the order and relative distances of all elements was preserved, and all exhibit very high degrees of sequence conservation, including three that are 100% identical. Upon direct comparison, nine out of the 15 putative DNase HS sites in mouse show at least partial overlap with vertebrate conserved elements. When yet another layer of complexity was added with ChIP on chip peak data representing histone modifications and variants in different osteoblastic cells, the number of loci of interest was further narrowed, such that three DNase HS loci close to the two 3C peaks (Figure 16) eventually stand out as most likely candidate enhancers. While all three candidates are located in immediate vicinity (i.e. within <50bp) of either H2AZ and/or meH3K4 marks in MN7 cells and/or calvaria, though, only the most 3' one of these three putative DNase HS sites actually exhibits a short, 9bp, direct sequence overlap with a ChIP on chip peak - namely H2AZ in MN7s. Since both MN7s and calvaria are osteoblast populations known to exhibit robust *Bmp2* expression, these results could very well be explained by histones being displaced from the DNA in the precise region of DNase hypersensitivity, which would deplete any modified histone signal. Alternatively, it is also possible that the lack of significant overlap could represent a mere artifact of slight differences between the underlying peak calling algorithms, and consequently is of no particular concern with regards to the potential functional significance of these loci.

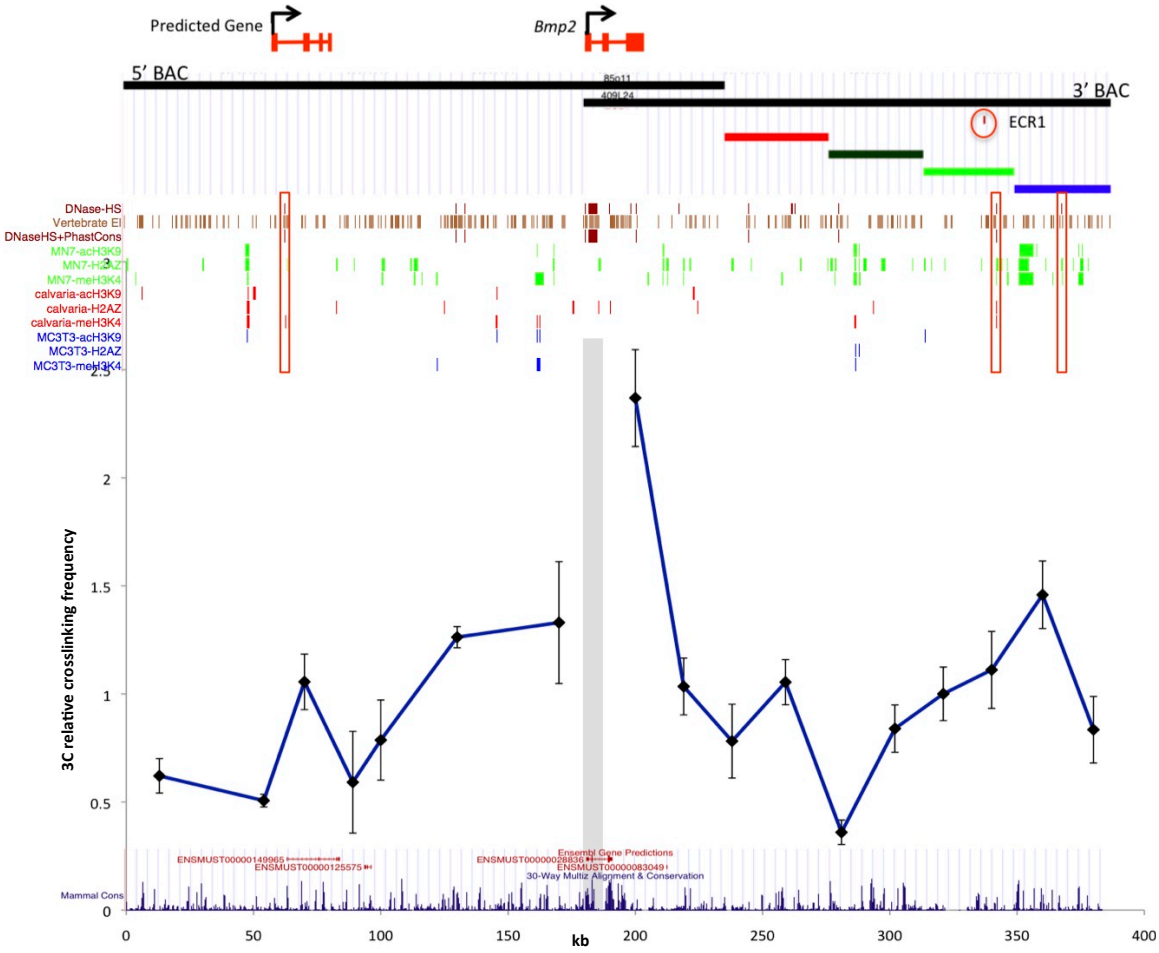


Figure 16. **Integrated 3C, DNaseHS, ChIP on chip analysis, and evolutionary conservation at the *Bmp2* locus.** Depicted are average relative crosslinking frequencies (+/- SEM) from triplicate 3C experiments observed between the *Bmp2* promoter (grey vertical bar) and 17 HindIII restriction fragments across the 5'- and 3'-BAC regions in MN7 cells. Above the graph, UCSC Genome Browser custom tracks display the overlap of DNaseHS sites (ENCODE data from human osteoblasts mapped onto the murine locus) with vertebrate conserved elements compared to select histone modifications and remodeling signatures observed in various osteoblasts by chIP on chip analysis. As reference, the locations of the 5'- and 3'-BACs, deletions 1-4 (red, black, green, and blue bars), and ECR1 are indicated on top.

Discussion

While the exact reason for the lack of *Bmp2* expression observed in our recent experience with MC3T3-E1 cells remains unknown, one plausible explanation is that they have simply become too differentiated with repeated passaging, such that *Bmp2* expression has already been reduced to a barely detectable baseline level, which in fact is a theory that agrees with caveats regarding the loss of osteoblast phenotype resulting from prolonged culture that has already been addressed in previous reports [12]. Regardless of the underlying cause, however, the important conclusion to be drawn from these observations is the fact that these subclones of MC3T3-E1 cells are not suitable for our purposes that require an *in vitro* system in which *Bmp2* is robustly expressed. Although some preliminary data was generated in MC3T3-E1 cells, we feel strongly that the current phenotype of these cells is an important indication against continuing to use them as a model system that is meant to recapitulate *in vivo* osteoblast conditions. Instead, they do however serve another important role as negative control, as has already been demonstrated in our 3C assays.

In fact, one of the most valuable conclusions from our 3C studies is that looping interactions across an individual gene locus are indeed dynamic - as evidenced by the MC3T3-E1 and MN7+FGF2 results -, since they directly correspond to transcriptional and thus enhancer activity. This finding in turn underscores its potential value for the identification of functionally significant elements at a given locus of interest, which is illustrated, for example, by both of the 3C peaks found in the *Bmp2* gene desert. At the distant end of the 3'BAC region, still downstream of ECR1, nothing had previously given us any reason to suspect regulatory function at any particular locus until the 3C data

prompted a closer look at other features in this region. In fact, if not for the overlap with conserved DNase HS sites and histone marks in this region, the two candidate loci around the 3C peak downstream of ECR1 (Figure 16) may otherwise have never been identified as regions of interest. Similarly, the 3C data was instrumental in highlighting a potential regulatory function in the 5' BAC region. Although in hindsight one could argue that the lincRNA upstream of *Bmp2* could have been recognized as potential sign of functional significance, it is rather unlikely that it would have elicited much interest on this end of the BAC region if it was not for the 3C peak that originally directed our attention its way and eventually prompted the discovery of an impressive accumulation of multiple other features (i.e. DNase HS site, evolutionary conservation, as well as H3K4 methylation in calvaria and H2AZ in MN7s), that all converge around the first exon of the lincRNA.

Given this plethora of evidence, it is highly unlikely to be purely coincidental, but, on the contrary, is extremely consistent with the rapidly growing understanding of the prevalence and significance of noncoding RNAs (ncRNAs). As the GENCODE consortium has, for instance, annotated over 3,000 long noncoding RNAs (lncRNAs) across the human genome [111], [112], [113], various types of ncRNAs have been receiving increased attention and have been reviewed extensively in recent years (e.g. [114], [113]). While ncRNAs are known to be transcribed from both promoters and enhancers, lncRNAs, including lincRNAs, in particular have been shown to have cis-regulatory function [115], and studies of the *HOXA* cluster, for example, have even reported a direct association between chromosomal looping and lincRNA expression already [116]. Thus, considering that the fundamental premise underlying the looping hypothesis is direct physical contact between promoters and distant regulatory elements

through various components of the transcription initiation complex, which consequently brings a participating enhancer in very close proximity with RNA polymerase II as well, it suggests itself that some degree of transcription could easily occur from an enhancer locus coincidentally with active transcription from the promoter itself. Conversely it follows that, similar to the presence of 3C peaks, active transcription of a lincRNA could in turn function as strong indicator for the presence of a regulatory element at the corresponding locus.

Overall, our findings across the *Bmp2* locus highlight not only the 3C methodology itself, but moreover its integration with other genomic characteristics as a very powerful and efficient tool to screen large genomic regions in order to identify putative enhancers, although it of course remains to be seen whether or not the regulatory potential of these candidate loci can ultimately be verified in functional assays.

CHAPTER IV

SUMMARY AND FUTURE DIRECTIONS

Analyzing DNase HS in murine osteoblasts

Having identified putative DNase HS sites at the mouse *Bmp2* locus by conversion of ENCODE data from human osteoblasts and further solidified loci of particular interest through the integration of various other forms of *in vitro* evidence, a logical next step to verify the significance of these candidate loci would be to directly analyze their chromatin structure in a DNase hypersensitivity assay on MN7 cells and/or primary mouse osteoblasts. While comprehensive screens for DNase HS across large genomic regions such as the entire *Bmp2* locus are expensive and would require the application of high-throughput technologies, the preliminary data would allow us to concentrate our analysis to use a more targeted approach where the DNase hypersensitivity would be quantified by SYBR Green real-time PCR [117], which has been shown to be not only more practicable than the conventional Southern blotting technique, but also superior in such critical aspects as resolution and accuracy of quantification.

In order to facilitate the detection of fragments of interest in these DNase HS assays, we could take advantage of transgenic mice carrying the 3'- or 5' lacZ-BAC, since primary osteoblasts isolated from either of these lines include multiple copies of the respective transgene, which in turn would greatly increase the intensity of PCR signals. Lastly, considering that active promoters are generally located in regions of open

chromatin and we know that *Bmp2* is robustly expressed in primary osteoblasts, the *Bmp2* promoter, which is in fact extremely hypersensitive to DNase I in human osteoblasts, can be used as positive control for these assays.

Further analysis of ECR1 response to FGF2 *in vitro*

While a previous study has reported ECR1 upregulation by FGF2 *in vitro* [97], one cannot neglect the fact that the luciferase assay results reported it that particular study were only based on a single, at that point unreplicated experiment and were also performed in MC3T3-E1 cells, which however fail to express *Bmp2* in our experience. Thus it would be particularly important to thoroughly replicate the above-mentioned findings in other osteoblasts in order to validate the proposed interaction between ECR1 and FGF2. Since our data has shown that MN7 cells not only express much higher baseline levels of *Bmp2* than more differentiated MC3T3-E1 cells but also exhibit robust *Bmp2* induction by FGF2 (Figure 13), MN7s would naturally be the first cell line of choice in which to attempt to replicate the earlier luciferase assays, although further studies in additional cell lines as well as primary osteoblasts would be equally interesting in order to investigate whether ECR1 response to FGF2 is dependent on factors such as the endogenous activity of the enhancer in a given osteoblast population, or if its response is even independent of baseline expression levels of *Bmp2*, as the previous findings in MC3T3-E1 cells would suggest.

For this purpose, we already established two stably transfected MN7 cell lines with the ECR1- and minimal *Bmp2*-promoter luciferase constructs (as described for the above-mentioned study in MC3T3-E1 cells [97]), respectively, each co-transfected with

an RFP-Hygromycin selectable marker, which only have yet to be analyzed in luciferase assays under varying concentrations of FGF2. Eventually, these luciferase assays could be expanded by strategic mutagenesis of the ECR1 construct, whereas a systematic deletion analysis would not only allow us to determine exactly which part(s) of the 656bp element are essential for enhancer activity, but it could also identify the exact location of the FGF2-responsive element or test for the significance of candidate transcription factor binding sites based on *in silico* predictions.

Functional tests of putative additional osteoblast enhancers

Following extensive *in vitro* and *in silico* analyses to identify candidate osteoblast enhancers at the *Bmp2* locus, the quasi gold-standard would of course be to test the functionality of each putative cis-regulatory elements *in vivo* by creating transgenic mouse models. Since this approach is naturally very time-intensive and can often be cost-prohibitive, though, a viable alternative *in vitro* would be to utilize additional luciferase assays similar to the ones mentioned above. Thus, one could initially screen a larger number of DNA element of interest, once they have been cloned into the same luciferase plasmid in place of ECR1, for their ability to upregulate luciferase expression - with or without FGF2 treatment, if desired – and compared to activity of the minimal *Bmp2* promoter alone. Select element(s) that show particular promise in these *in vitro* enhancer assays could still be chosen for confirmation in transgenic analysis. In case it were necessary or preferable to investigate the functionality of a particular element *in vitro* within a more authentic context in regards to its genomic environment and location relative to *Bmp2*, one could alternatively transfect any *Bmp2* lacZ-BAC into MN7 cells

after targeted deletion or modification of a locus of interest in order to determine its effect on β-galactosidase activity.

Identifying specific transcription factors that act on osteoblast enhancers

Studying upstream regulation of *Bmp2* in calvarial osteoblasts, Choi et al. have previously shown that FGF2 acts through the transcription factor Runx2 to activate *Bmp2* expression [42]. Moreover, subsequent studies by Ron Chandler not only identified a candidate Runx2 binding site at the ECR1 locus by TRANSFAC analysis, but also demonstrated an interaction between ECR1 and Runx2 in a ChIP assay in MC3T3-E1 cells [12], suggesting that ECR1 might play an important role as intermediary in this pathway. Thus, it is imperative to try and replicate these findings in order to determine, in particular, whether or not ECR1 can similarly be shown to bind Runx2 in MN7 cells and primary osteoblasts, and/or which, if any, other transcription factors might be involved. At the same time, ChIP assays could also be used to test if any transcription factor binding can be enhanced by FGF2 treatment, and eventually, similar assays can of course be applied to any other osteoblast enhancer(s) as well, once their location is confirmed.

3C assays on additional (non-) osteoblast cell lines

As mentioned above, all 3C assays described here support the hypothesis that looping interactions are dynamic and correlate well with the transcriptional activity observed at the *Bmp2* locus; but so far, our studies have only examined a limited number of osteoblast cell lines and culture conditions, and in order to be able to reinforce our hypothesis, it would certainly be valuable to expand the analysis to additional osteoblast

and (non-) osteoblast cell lines with varying degrees of *Bmp2* expression. While other *Bmp2*-expressing osteoblasts would be expected to exhibit similar looping profiles to MN7s, non-osteoblast cell lines that do not express *Bmp2* would likewise be lacking any significant peaks of crosslinking frequency consistent with the MC3T3-E1 profile, if our hypothesis holds. Ultimately, though, perhaps the most powerful evidence could be obtained from MN7s in which *Bmp2* expression has been transcriptionally repressed, as any decrease or loss of the previously observed crosslinking peaks would compellingly underscore the presumably highly dynamic nature of chromosome conformation.

In vivo analysis of looping interactions in different osteoblast populations and other *Bmp2*-expressing tissues

Considering that the studies presented here have not only provided overwhelming evidence for the hypothesis that ECR1 cannot be the only enhancer controlling *Bmp2* expression in bone, but have also demonstrated that chromosome conformation is dynamic and can readily adjust to the specific transcriptional needs of a locus at any given time, one might actually be able to utilize 3C as an efficient means to identify the location of any additional osteoblast enhancer(s). Since the chromosome conformation might be expected to differ between different osteoblast populations in order to allow for optimal tissue- or anatomical-specific communication between the promoter and the respective enhancer, 3C assays performed on primary osteoblasts isolated from calvaria and long bones, respectively, could help identify the exact location(s) of any additional osteoblast enhancer(s) based on the distinctive looping profiles found to be characteristic of a given osteoblast population.

Likewise, results from the previous BAC transgenesis have provided strong evidence for the existence of many more tissue-specific enhancers in this ~400kb region surrounding the *Bmp2* gene [19], [12] (Figure 3), but except for ECR1 and a digit-specific element in the 3' region that is more proximal to *Bmp2* than ECR1 [118], none of these other elements have been localized with resolution greater than the ~40kb deletion regions. Since additional transgenic experiments to test more conserved regions in the two BACs for possible enhancer function would be time-consuming and expensive, it is important to consider alternative approaches that might facilitate future functional studies by identifying the locations of these cis-regulatory elements with greater precision. For reasons detailed above, 3C could potentially fit this purpose, as it is not only significantly more cost-effective than transgenic approaches, but also unbiased in that it does not rely on pre-determined criteria such as evolutionary conservation to select regions of interest. Once again, the underlying assumption of using 3C for this purpose would be the dynamic nature of the chromosome conformation, where the physical interactions between promoter and enhancers are specific to a given tissue or cell type and can readily change depending on transcriptional requirements and/or differentiation status of the cells, which has not only been suggested by our studies here, but also previously demonstrated to be the case at several other gene loci, such as the β-globin locus, *Ifng,* and *Foxl2* [71], [100], [74], [84].

In fact, both applications of the 3C methodology described above have already begun to be tested, as 3C assays for this very purpose have already been performed on both cultured, calvarial primary osteoblasts isolated at E17.5 and on chromatin isolated from embryonic liver tissue at E15.5, although the resulting samples have yet to be

analyzed by qPCR. In addition to primary long bone osteoblasts for direct comparison to calvarial primaries, other embryonic tissues that could be great candidates for 3C analysis would be kidneys or gut tissue, which, similar to liver, have important advantages of ease of precise dissection, consistent transgene expression in >91% of transgenic embryos observed for the respective lacZ-BACs, and the fact that previous analysis has shown *Bmp2* to be expressed in a relatively large percentage of cells within each of these tissues (Figure 3).

Independent verification of looping interactions at the *Bmp2* locus

In addition to 3C, ChIP with antibodies against components of the preinitiation complex, such as RNA polymerase II (RNA pol II) and a common coactivator, could not only serve as an independent means to test whether or not any given enhancer of interest physically interacts with the *Bmp2* promoter during transcriptional activation, but it also simultaneously has the potential to identify additional trans-factors that contributes to these enhancers' activities.

Coactivators such as p300 are defined as "proteins that do not bind DNA, but are required for transcriptional activation of gene expression" [43], [119]. The fact that they do not themselves bind DNA either at the promoter or at the enhancer, but yet are vital components of the large preinitiation complex that associates the two DNA elements with one another, can be used to our advantage when an alternative method is sought to support the model of looping between any given enhancer and the *Bmp2* promoter. In fact, Visel et al. have convincingly demonstrated with a high-throughput ChIP-seq analysis that p300 binding is indeed able to predict the location of enhancer elements

[120], which corroborates the rationale of this approach. Moreover, since the coactivator and histone acetyltransferase p300 is known to bind Runx2 and has especially been demonstrated to be critical for the activation of the *Osteocalcin* promoter [43], it qualifies as a strong candidate for a similar role in the any Runx2-mediated induction of *Bmp2* in osteoblasts. Although knowledge about the occupancy of transcription factors and coactivators itself is ultimately nothing but suggestive in nature, it also resembles 3C looping profiles in the point that its informative value could be increased significantly if it can be shown to be dynamic across various cell lines and/or under different conditions (e.g. with FGF2 treatment) and corresponds to variations in *Bmp2* expression as well.

Ultimately, if the looping hypothesis is supported by the association of RNA pol II with both the *Bmp2* promoter and osteoblast enhancers, the central location of this active polymerase raises the question whether we can detect evidence for transcription at the respective enhancer locus as well. Although little is known about the functional benefits of such transcription from regulatory elements, it has been well studied in *Drosophila* [121], [122], and both the recruitment of RNA pol II to known enhancers [123], [124] and the generation of non-coding RNAs [125], [126], [127] are phenomena that have been described in mammalian genomes as well. Thus, considering the close physical proximity between interacting elements in the large transcription initiation complex, it is very plausible that low levels of transcription from an enhancer locus accompany *Bmp2* activation, and if this association can indeed be detected *in vitro* or *in vivo*, the findings would add further invaluable insight into the molecular mechanisms that facilitate enhancer activity.

Summary

A member of the TGF-β superfamily of cytokines, BMP2 not only plays a critical role in pattern formation and morphogenesis during early embryonic development, but also promotes osteoblast differentiation and bone formation, making it a vital factor for the maintenance of bone health. Expanding on earlier studies of the regulatory landscape surrounding *Bmp2* by BAC transgenesis in mice that had lead to the identification of the osteoblast enhancer ECR1, the work presented here further scrutinizes the particular role of ECR1 during osteogenesis and provides striking evidence for the hypothesis that – despite its indispensable role for *Bmp2* expression in osteoblasts -, it is in fact not entirely autonomous, but requires additional enhancer activity to control the full scope of osteoblast-specific expression. Subsequent analysis of the chromosomal conformation during transcriptional activation does not only support this hypothesis and substantiate the dynamic nature of looping interactions at the locus, but combined with additional epigenetic characteristics also serves to highlight several distant loci within the gene desert as excellent candidates for putative enhancer function. Therefore, in addition to future functional tests of these candidate loci, another question of particular interest will then be to explore whether or not dynamic changes in chromosome conformation accompany targeted *Bmp2* regulation in different osteoblast populations and other tissues, as this could potentially prove to be an invaluable tool to efficiently identify the location(s) of tissue-specific regulatory elements.

Eventually - considering its critical role in osteogenesis and consequent significance for human health and disease, e.g. as risk factor for various bone disorders such as osteoporosis, which substantiates *Bmp2* as a primary target for therapeutic

intervention -, a clear understanding of the factors and mechanisms that control *Bmp2* expression in osteoblasts will be essential for the future development of therapies that could enhance bone growth and stimulate healing.

References

1. Chen, D., M. Zhao, and G. Mundy, *Bone Morphogenetic Proteins*, in *Growth Factors*. 2004. p. 233-241.

2. Ducy, P. and G. Karsenty, *The family of bone morphogenetic proteins.* Kidney Int, 2000. **57**(6): p. 2207-14.

3. Hogan, B.L., *Bone morphogenetic proteins: multifunctional regulators of vertebrate development.* Genes Dev, 1996. **10**(13): p. 1580-94.

4. Blitz, I.L. and K.W. Cho, *Finding partners: how BMPs select their targets.* Dev Dyn, 2009. **238**(6): p. 1321-31.

5. Zhang, H. and A. Bradley, *Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development.* Development, 1996. **122**(10): p. 2977-86.

6. Fernandez-Rozadilla, C., et al., *BMP2/BMP4 colorectal cancer susceptibility loci in northern and southern European populations.* Carcinogenesis. **34**(2): p. 314-318.

7. Tomlinson, I.P., et al., *Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer.* PLoS Genet. **7**(6): p. e1002105.

8. Houlston, R.S., et al., *Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33.* Nat Genet. **42**(11): p. 973-7.

9. Kim, Y.O., et al., *Polymorphisms in bone morphogenetic protein 3 and the risk of papillary thyroid cancer.* Oncol Lett. **5**(1): p. 336-340.

10. Rabinovitch, M., *Molecular pathogenesis of pulmonary arterial hypertension.* J Clin Invest. **122**(12): p. 4306-13.

11. Cogan, J., et al., *Role of BMPR2 alternative splicing in heritable pulmonary arterial hypertension penetrance.* Circulation. **126**(15): p. 1907-16.

12. Chandler, R., *REGULATION OF BMP2 TRANSCRIPTION IN BONE.* 2007.

13. Urist, M.R., *Bone: formation by autoinduction.* Science, 1965. **150**(698): p. 893-9.

14. Wozney, J.M., et al., *Novel regulators of bone formation: molecular clones and activities.* Science, 1988. **242**(4885): p. 1528-34.

15. Tsuji, K., et al., *BMP2 activity, although dispensable for bone formation, is required for the initiation of fracture healing.* Nat Genet, 2006. **38**(12): p. 1424-9.

16. Nakashima, K. and B. de Crombrugghe, *Transcriptional mechanisms in osteoblast differentiation and bone formation.* Trends Genet, 2003. **19**(8): p. 458-66.

17. Lecanda, F., L.V. Avioli, and S.L. Cheng, *Regulation of bone matrix protein expression and induction of differentiation of human osteoblasts and human bone marrow stromal cells by bone morphogenetic protein-2.* J Cell Biochem, 1997. **67**(3): p. 386-96.

18. Mundy, G., et al., *Stimulation of bone formation in vitro and in rodents by statins.* Science, 1999. **286**(5446): p. 1946-9.

19. Chandler, R., et al., *Bmp2 Transcription in Osteoblast Progenitors Is Regulated by a Distant 3' Enhancer Located 156.3 Kilobases from the Promoter*, in *Molecular and Cellular Biology*. 2007. p. 2934-2951.

20. Yoshida, Y., et al., *Negative regulation of BMP/Smad signaling by Tob in osteoblasts.* Cell, 2000. **103**(7): p. 1085-97.

21. Usui, M., et al., *Enhancing effect of Tob deficiency on bone formation is specific to bone morphogenetic protein-induced osteogenesis.* J Bone Miner Res, 2002. **17**(6): p. 1026-33.

22. Wu, X.B., et al., *Impaired osteoblastic differentiation, reduced bone formation, and severe osteoporosis in noggin-overexpressing mice.* J Clin Invest, 2003. **112**(6): p. 924-34.

23. Devlin, R.D., et al., *Skeletal overexpression of noggin results in osteopenia and reduced bone formation.* Endocrinology, 2003. **144**(5): p. 1972-8.

24. Bandyopadhyay, A., et al., *Genetic analysis of the roles of BMP2, BMP4, and BMP7 in limb patterning and skeletogenesis.* PLoS Genet, 2006. **2**(12): p. e216.

25. Styrkarsdottir, U., et al., *Linkage of osteoporosis to chromosome 20p12 and association to BMP2.* PLoS Biol, 2003. **1**(3): p. E69.

26. Zmuda, J.M., Y.T. Sheu, and S.P. Moffett, *The search for human osteoporosis genes.* J Musculoskelet Neuronal Interact, 2006. **6**(1): p. 3-15.

27. Valdes, A.M., et al., *Association study of candidate genes for the prevalence and progression of knee osteoarthritis.* Arthritis Rheum, 2004. **50**(8): p. 2497-507.

28. Valdes, A.M., et al., *Reproducible genetic associations between candidate genes and clinical knee osteoarthritis in men and women.* Arthritis Rheum, 2006. **54**(2): p. 533-9.

29. Wang, H., et al., *Association between two polymorphisms of the bone morphogenetic protein-2 gene with genetic susceptibility to ossification of the posterior longitudinal ligament of the cervical spine and its severity.* Chin Med J (Engl), 2008. **121**(18): p. 1806-10.

30. Wang, H., et al., *Association of bone morphogenetic protein-2 gene polymorphisms with susceptibility to ossification of the posterior longitudinal ligament of the spine and its severity in Chinese patients.* Eur Spine J, 2008. **17**(7): p. 956-64.

31. Fritz, D.T., et al., *A polymorphism in a conserved posttranscriptional regulatory motif alters bone morphogenetic protein 2 (BMP2) RNA:protein interactions.* Mol Endocrinol, 2006. **20**(7): p. 1574-86.

32. Sandhu, H.S. and S.N. Khan, *Recombinant human bone morphogenetic protein-2: use in spinal fusion applications.* J Bone Joint Surg Am, 2003. **85-A Suppl 3**: p. 89-95.

33. Murnaghan, M., et al., *Time for treating bone fracture using rhBMP-2: a randomised placebo controlled mouse fracture trial.* J Orthop Res, 2005. **23**(3): p. 625-31.

34. Visser, R., et al., *The effect of an rhBMP-2 absorbable collagen sponge-targeted system on bone formation in vivo.* Biomaterials, 2009. **30**(11): p. 2032-7.

35. Garrett, I.R., G. Gutierrez, and G.R. Mundy, *Statins and bone formation.* Curr Pharm Des, 2001. **7**(8): p. 715-36.

36. Garrett, I.R. and G.R. Mundy, *The role of statins as potential targets for bone formation.* Arthritis Res, 2002. **4**(4): p. 237-40.

37. Gutierrez, G.E., et al., *Transdermal application of lovastatin to rats causes profound increases in bone formation and plasma concentrations.* Osteoporos Int, 2006. **17**(7): p. 1033-42.

38. Garrett, I.R., et al., *Locally delivered lovastatin nanoparticles enhance fracture healing in rats.* J Orthop Res, 2007. **25**(10): p. 1351-7.

39. Tang, Q.O., et al., *Statins: under investigation for increasing bone mineral density and augmenting fracture healing.* Expert Opin Investig Drugs, 2008. **17**(10): p. 1435-63.

40. Bostan, B., et al., *Simvastatin improves spinal fusion in rats.* Acta Orthop Traumatol Turc. **45**(4): p. 270-5.

41. Huang, Z., et al., *The sequential expression profiles of growth factors from osteoprogenitors [correction of osteroprogenitors] to osteoblasts in vitro.* Tissue Eng, 2007. **13**(9): p. 2311-20.

42. Choi, K., et al., *Runx2 regulates FGF2-induced Bmp2 expression during cranial bone development*, in *Dev. Dyn.* 2005. p. 115-121.

43. Schroeder, T., E. Jensen, and J. Westendorf, *Runx2: A master organizer of gene transcription in developing and maturing osteoblasts*, in *Birth Defect Res C*. 2005. p. 213-225.

44. Marie, P.J., *Fibroblast growth factor signaling controlling osteoblast differentiation.* Gene, 2003. **316**: p. 23-32.

45. Otto, F., et al., *Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development.* Cell, 1997. **89**(5): p. 765-71.

46. Komori, T., et al., *Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts.* Cell, 1997. **89**(5): p. 755-64.

47. Siepel, A., et al., *Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes.* Genome Res, 2005. **15**(8): p. 1034-50.

48. DiLeone, R.J., et al., *Efficient studies of long-distance Bmp5 gene regulation using bacterial artificial chromosomes.* Proc Natl Acad Sci U S A, 2000. **97**(4): p. 1612-7.

49. DiLeone, R.J., L.B. Russell, and D.M. Kingsley, *An extensive 3' regulatory region controls expression of Bmp5 in specific anatomical structures of the mouse embryo.* Genetics, 1998. **148**(1): p. 401-8.

50. Muller, F., et al., *Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord.* Development, 1999. **126**(10): p. 2103-16.

51. Sagai, T., et al., *Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb.* Development, 2005. **132**(4): p. 797-803.

52. Lettice, L.A., et al., *A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly.* Hum Mol Genet, 2003. **12**(14): p. 1725-35.

53. Loots, G.G., et al., *Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease.* Genome Res, 2005. **15**(7): p. 928-35.

54. Emison, E.S., et al., *A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk.* Nature, 2005. **434**(7035): p. 857-63.

55. Frazer, K.A., et al., *VISTA: computational tools for comparative genomics.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W273-9.

56. Mayor, C., et al., *VISTA : visualizing global DNA sequence alignments of arbitrary length.* Bioinformatics, 2000. **16**(11): p. 1046-7.

57. Schwartz, S., et al., *PipMaker--a web server for aligning two genomic DNA sequences.* Genome Res, 2000. **10**(4): p. 577-86.

58. Schwartz, S., et al., *MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences.* Nucleic Acids Res, 2003. **31**(13): p. 3518-24.

59. Woolfe, A., et al., *Highly conserved non-coding sequences are associated with vertebrate development.* PLoS Biol, 2005. **3**(1): p. e7.

60. Miller, W., et al., *28-way vertebrate alignment and conservation track in the UCSC Genome Browser.* Genome Res, 2007. **17**(12): p. 1797-808.

61. Fisher, S., et al., *Conservation of RET regulatory function from human to zebrafish without sequence similarity.* Science, 2006. **312**(5771): p. 276-9.

62. Spitz, F., F. Gonzalez, and D. Duboule, *A global control region defines a chromosomal regulatory landscape containing the HoxD cluster.* Cell, 2003. **113**(3): p. 405-17.

63. Spitz, F., et al., *Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes.* Nat Genet, 2005. **37**(8): p. 889-93.

64. Holohan, E.E., et al., *CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex.* PLoS Genet, 2007. **3**(7): p. e112.

65. Grice, E.A., et al., *Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer.* Hum Mol Genet, 2005. **14**(24): p. 3837-45.

66. Nobrega, M.A., et al., *Scanning human gene deserts for long-range enhancers.* Science, 2003. **302**(5644): p. 413.

67. Nelson, C.E., B.M. Hersh, and S.B. Carroll, *The regulatory content of intergenic DNA shapes genome architecture.* Genome Biol, 2004. **5**(4): p. R25.

68. Muller, M.M., T. Gerster, and W. Schaffner, *Enhancer sequences and the regulation of gene transcription.* Eur J Biochem, 1988. **176**(3): p. 485-95.

69. Geyer, P.K. and V.G. Corces, *DNA position-specific repression of transcription by a Drosophila zinc finger protein.* Genes Dev, 1992. **6**(10): p. 1865-73.

70. Carter, D., et al., *Long-range chromatin regulatory interactions in vivo.* Nat Genet, 2002. **32**(4): p. 623-6.

71. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.

72. van Berkum, N.L. and J. Dekker, *Determining spatial chromatin organization of large genomic regions using 5C technology.* Methods Mol Biol, 2009. **567**: p. 189-213.

73. Lomvardas, S., et al., *Interchromosomal interactions and olfactory receptor choice.* Cell, 2006. **126**(2): p. 403-13.

74. Spilianakis, C.G., et al., *Interchromosomal associations between alternatively expressed loci.* Nature, 2005. **435**(7042): p. 637-45.

75. Chandler, K.J., R.L. Chandler, and D.P. Mortlock, *Identification of an ancient Bmp4 mesoderm enhancer located 46 kb from the promoter*, in *Developmental Biology*. 2009. p. 590-602.

76. Portnoy, M.E., et al., *Detection of potential GDF6 regulatory elements by multispecies sequence comparisons and identification of a skeletal joint enhancer.* Genomics, 2005. **86**(3): p. 295-305.

77. Mortlock, D.P., C. Guenther, and D.M. Kingsley, *A general approach for identifying distant regulatory elements applied to the Gdf6 gene.* Genome Res, 2003. **13**(9): p. 2069-81.

78. Dekker, J., *Capturing Chromosome Conformation*, in *Science*. 2002. p. 1306-1311.

79. Splinter, E., F. Grosveld, and W. de Laat, *3C technology: analyzing the spatial organization of genomic loci in vivo.* Methods Enzymol, 2004. **375**: p. 493-507.

80. Lodish, H.F., *Molecular cell biology*. 4th ed. 2000, New York: W.H. Freeman. xxxvi, 1084, G-17, I-36 p.

81. Dekker, J., *The three 'C' s of chromosome conformation capture: controls, controls, controls*, in *Nat Meth*. 2006. p. 17-21.

82. Hagège, H., et al., *Quantitative analysis of chromosome conformation capture assays (3C-qPCR)*, in *Nat Protoc*. 2007. p. 1722-1733.

83. Pomerantz, M., et al., *The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer*, in *Nat Genet*. 2009. p. 882-884.

84. D'Haene, B., et al., *Disease-causing 7.4 kb cis-regulatory deletion disrupting conserved non-coding sequences and their interaction with the FOXL2 promotor: implications for mutation screening.* PLoS Genet, 2009. **5**(6): p. e1000522.

85. Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription.* Nat Genet, 2004. **36**(10): p. 1065-71.

86. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.

87. Splinter, E., et al., *Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation.* Methods. **58**(3): p. 221-30.

88. van de Werken, H.J., et al., *4C technology: protocols and data analysis.* Methods Enzymol. **513**: p. 89-112.

89. van de Werken, H.J., et al., *Robust 4C-seq data analysis to screen for regulatory DNA interactions.* Nat Methods. **9**(10): p. 969-72.

90. Dostie, J., Y. Zhan, and J. Dekker, *Chromosome conformation capture carbon copy technology.* Curr Protoc Mol Biol, 2007. **Chapter 21**: p. Unit 21 14.

91. Dostie, J., et al., *Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements.* Genome Res, 2006. **16**(10): p. 1299-309.

92. Dostie, J. and J. Dekker, *Mapping networks of physical interactions between genomic elements using 5C technology.* Nat Protoc, 2007. **2**(4): p. 988-1002.

93. Ferraiuolo, M.A., et al., *From cells to chromatin: capturing snapshots of genome organization with 5C technology.* Methods. **58**(3): p. 255-67.

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

95. Guenther, C., et al., *Shaping Skeletal Growth by Modular Regulatory Elements in the Bmp5 Gene*, in *PLoS Genet*. 2008. p. e1000308.

96. Naganawa, T., et al., *Reduced expression and function of bone morphogenetic protein-2 in bones of Fgf2 null mice*, in *J. Cell. Biochem.* 2008. p. 1975-1988.

97. Jiang, S., et al., *Repressive BMP2 gene regulatory elements near the BMP2 promoter.* Biochem Biophys Res Commun. **392**(2): p. 124-8.

98. Warming, S., et al., *Simple and highly efficient BAC recombineering using galK selection.* Nucleic Acids Res, 2005. **33**(4): p. e36.

99. Chandler, K.J., et al., *Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression.* Mamm Genome, 2007. **18**(10): p. 693-708.

100. Eivazova, E.R. and T.M. Aune, *Dynamic alterations in the conformation of the Ifng gene region during T helper cell differentiation.* Proc Natl Acad Sci U S A, 2004. **101**(1): p. 251-6.

101. Miele, A., K. Bystricky, and J. Dekker, *Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions.* PLoS Genet, 2009. **5**(5): p. e1000478.

102. Splinter, E., *CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus*, in *Genes & Development*. 2006. p. 2349-2354.

103. Weintraub, H. and M. Groudine, *Chromosomal subunits in active genes have an altered conformation.* Science, 1976. **193**(4256): p. 848-56.

104. Stalder, J., et al., *Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I.* Cell, 1980. **20**(2): p. 451-60.

105. Paul, J., et al., *Distant sequences which regulate globin genes.* Prog Clin Biol Res, 1985. **191**: p. 29-48.

106. Lee, G.R., C.G. Spilianakis, and R.A. Flavell, *Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions.* Nat Immunol, 2005. **6**(1): p. 42-8.

107. He, Y., et al., *Genome-wide bovine H3K27me3 modifications and the regulatory effects on genes expressions in peripheral blood lymphocytes.* PLoS One. **7**(6): p. e39094.

108. Dai, Y., et al., *Genome-wide analysis of histone H3 lysine 4 trimethylation by ChIP-chip in peripheral blood mononuclear cells of systemic lupus erythematosus patients.* Clin Exp Rheumatol. **28**(2): p. 158-68.

109. Pregizer, S., et al., *Progressive recruitment of Runx2 to genomic targets despite decreasing expression during osteoblast differentiation.* J Cell Biochem, 2008. **105**(4): p. 965-70.

110. Mathieu, E., et al., *Establishment of an osteogenic cell line derived from adult mouse bone marrow stroma by use of a recombinant retrovirus.* Calcif Tissue Int, 1992. **50**(4): p. 362-71.

111. Derrien, T., et al., *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.* Genome Res. **22**(9): p. 1775-89.

112. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome Res. **22**(9): p. 1760-74.

113. Guil, S. and M. Esteller, *Cis-acting noncoding RNAs: friends and foes.* Nat Struct Mol Biol. **19**(11): p. 1068-75.

114. Lee, J.T., *Epigenetic regulation by long noncoding RNAs.* Science. **338**(6113): p. 1435-9.

115. Orom, U.A., et al., *Long noncoding RNAs with enhancer-like function in human cells.* Cell. **143**(1): p. 46-58.

116. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature. **472**(7341): p. 120-4.

117. McArthur, M., S. Gerum, and G. Stamatoyannopoulos, *Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR.* J Mol Biol, 2001. **313**(1): p. 27-34.

118. Dathe, K., et al., *Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2.* Am J Hum Genet, 2009. **84**(4): p. 483-92.

119. Roeder, R.G., *Transcriptional regulation and the role of diverse coactivators in animal cells.* FEBS Lett, 2005. **579**(4): p. 909-15.

120. Pennacchio, L., et al., *In vivo enhancer analysis of human conserved non-coding sequences*, in *Nature*. 2006. p. 499-502.

121. Akbari, O.S., et al., *Unraveling cis-regulatory mechanisms at the abdominal-A and Abdominal-B genes in the Drosophila bithorax complex.* Dev Biol, 2006. **293**(2): p. 294-304.

122. Feng, J., et al., *The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator.* Genes Dev, 2006. **20**(11): p. 1470-84.

123. Nerenz, R.D., M.L. Martowicz, and J.W. Pike, *An enhancer 20 kilobases upstream of the human receptor activator of nuclear factor-kappaB ligand gene mediates dominant activation by 1,25-dihydroxyvitamin D3.* Mol Endocrinol, 2008. **22**(5): p. 1044-56.

124. Kim, S., et al., *Activation of receptor activator of NF-kappaB ligand gene expression by 1,25-dihydroxyvitamin D3 is mediated through multiple long-range enhancers.* Mol Cell Biol, 2006. **26**(17): p. 6469-86.

125. Kapranov, P., et al., *Large-scale transcriptional activity in chromosomes 21 and 22.* Science, 2002. **296**(5569): p. 916-9.

126. *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.

127. Cheng, J., et al., *Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution.* Science, 2005. **308**(5725): p. 1149-54.