

Transcriptional Control of the Human Aggrecan Gene

by

Andrea Layyous

PhD Thesis

Kennedy Institute of Rheumatology

Imperial College London

Table of Contents

List of Figures	4
List of Tables	7
Acknowledgements	9
Abstract	10
Abbreviations	11
Introduction	13
1 <i>Transcription</i>	13
1.1 Introduction to transcription	13
1.2 Components of the eukaryotic transcriptional apparatus	15
1.3 The transcription cycle	32
1.4 The evolution of transcriptional control across eukaryotic species	36
2 <i>Skeletal development</i>	39
2.1 Introduction to skeletal development	39
2.2 Endochondral ossification- formation of the primary ossification centres (POCs).	40
2.3 Synovial joint formation	42
2.4 Postnatal development of bones- development of secondary ossification centres (SOCs)	44
3 <i>Aggrecan</i>	58
3.1 Overview	58
3.2 Aggrecan protein	59
3.3 The Aggrecan Gene	62
3.4 The proteolysis of aggrecan	73
3.5 The aggrecan knockout mouse	75
4 <i>Osteoarthritis</i>	77
4.1 Overview	77
4.2 The development of osteoarthritis	78
4.3 Understanding Osteoarthritis pathogenesis and testing therapeutic ideas	81
5 <i>Aims of PhD project</i>	84
Materials and Methods	85
1 <i>Isolation of nucleic acids:</i>	85
1.1 Isolation of plasmid DNA	85
1.2 Isolation of genomic DNA from embryonic and adult tails	86
1.3 Oligonucleotides	86
2 <i>Standard manipulations of DNA</i>	87
2.1 Polymerase chain reaction	87
2.2 Digestion of DNA by restriction enzymes	88
2.3 Agarose gel electrophoresis	88
2.4 Isolation and purification of DNA from agarose gels	88
2.5 Recombinant DNA procedures	89
2.6 Transformation of DH5 α E.coli	89
3 <i>Cloning vectors</i>	91
3.1 pBluescript	91
3.2 p β gal-Basic	91
3.3 RP11-343B18	93
3.4 RP11-1123E8	94
4 <i>Recombineering</i>	95
4.1 Making electro-competent EL350 or SW106 <i>E.coli</i> (for introducing BAC DNA, prior to recombineering):	95
4.2 Transforming a target BAC (RP11-343B18/RP11-1123E8) into electro-competent cells	95
4.3 Recombineering	96

<i>5 Generation of Constructs</i>	97
<i>6 Generation of transgenic mouse embryos and lines</i>	109
6.1 Transgenic media	109
6.2 Preparation of DNA for microinjection	109
6.3 Preparation of egg donors by superovulation	110
6.4 Injection of fertilised eggs.	110
6.5 Transfer to pseudopregnant recipients	110
<i>7 Transgenic mouse analysis</i>	112
7.1 Analysis of transgenic mouse embryos	112
7.2 Analysis of adult transgenic mouse lines	113
<i>8 Osteoarthritis models</i>	120
8.1 Surgical induction of OA through the destabilisation of the medial meniscus (DMM) model	120
8.2 Axial Loading of murine tibias	121
<i>9 Stock solutions</i>	124
9.1 Standard solutions	124
9.2 Bacterial culture and transformations	124
9.3 DNA preparation	124
9.4 Analysis of transgenic embryos	125
Results	126
<i>Chapter 1: Locating cis-acting sequences that regulate the cartilage-specific expression of the human aggrecan gene:</i>	126
1.1 Background	126
1.2 Results	130
1.3 Discussion	176
<i>Chapter 2: Conditional/inducible cartilage-specific model</i>	200
2.1 Background	200
2.2 Results	202
2.3 Discussion	211
<i>Chapter 3: Induction of osteoarthritis in the knees of line 39 ('Col1I4xA1Cre') and line 12 ('βAGG20')</i>	218
3.1 Background	218
3.2 Results	221
3.3 Discussion	241
Conclusions	252
Future Work	256
References	258

List of Figures

Introduction

Figure 1: Some core promoter motifs for transcription by RNA polymerase II	19
Figure 2: Mechanism of preinitiation complex assembly	26
Figure 3: Eukaryotic transcription factors and activation mechanisms	32
Figure 4: Multistages of the transcription cycle	37
Figure 5: Schematic diagram depicting the major steps in digital synovial joint formation	44
Figure 6 The major stages in the formation and development of long bones and epiphyses from stage 1 to stage 16a	47
Figure 7: Diagrammatic representation of the general structure of adult human articular cartilage to show the zones, regions and relationship with subchondral bone	53
Figure 8: Schematic representation of an aggregate of aggrecan, with detail of monomer	59
Figure 9: Diagram illustrating the evolution of osteoarthritis over time	78

Methods

Figure 10: pBluescript II KS (+/-) vector	92
Figure 11: p β gal-Basic vector	93
Figure 12: Mechanism of axial loading	123

Results Chapter 1

Figure 13: Paraffin section of mouse embryos at different stages of development stained with the BR1 antibody	132
Figure 14: Schematic vista analysis of the aggrecan gene including ~20kb of non-coding DNA upstream of the TSS (-20kb - +72kb relative to the TSS)	133
Figure 15: Schematic vista analysis of highly homologous sequence within intron 14 of the human aggrecan gene	134
Figure 16: Diagram representing the recombineering strategy used to clone the large ~20kb sequence (-20,054kb to +375bp) from the human Aggrecan gene into the pBS KS+ vector.	137
Figure 17: 2 representative E14.5 wholemount embryos expressing β -galactosidase (blue staining) driven by the β AGG20 promoter (-20,054bp - +375bp).	139

Figure 18: Histological sections of a blue stained transgenic embryo expressing β -galactosidase driven by the β AGG20 promoter (-20,054bp to +375bp).	140
Figure 19: Representative E16.5 wholemount embryo expressing β -galactosidase (blue staining) driven by the β AGG20 promoter (-20,054bp to +375bp).	143
Figure 20: ~400bp sequence with 83% identity between Human DNA (-10.7kb to 10.4kb relative to the human aggrecan TSS) and Rat DNA (-9.2kb to -8.7kb relative to the rat aggrecan TSS).	146
Figure 21: Diagram representing the recombineering strategy used to clone the deleted 7kb fragment into the p β gal-Basic vector upstream of the basal promoter and LacZ gene.	148
Figure 22: Representative E14.5 wholemount embryo expressing β -galactosidase (blue staining) driven by the β galAB+7 promoter (-20,054bp to -13,030bp fused to minimal promoter/-452bp to +375bp).	150
Figure 23: Histological sections of a representative β galAB+7 blue stained embryo.	151
Figure 24: Representative E14.5 wholemount embryo expressing β -galactosidase (blue staining) driven by the β galAB + cis 3 promoter (-10,703bp - -10,343bp fused to minimal promoter/-452bp - +375bp).	155
Figure 25: 4 representative E14.5 wholemount embryos expressing β -galactosidase (blue staining) driven by the β galAB ⁺⁹² +cis 3 promoter (-10,703bp to -10,343bp fused to minimal promoter including missing 92 base pairs/-452bp to +375bp).	158+159
Figure 26: Histological section of a β galAB ⁺⁹² +cis 3 blue stained embryos.	160
Figure 27: Diagram representing the recombineering strategy used to clone the 3kb intronic sequences (+5kb to +8kb) in an indirect way from the BAC RP11-343B18 into β AGG20 downstream of the ~20kb insert and upstream of the LacZ gene.	163
Figure 28: 3 representative E14.5 wholemount embryos expressing β -galactosidase (blue staining) driven by the β AGG20 + cis 6 promoter (-20,054bp to +375bp fused to +5,000bp to +7,813bp).	164
Figure 29: Tail tips of 5-week-old weaned pups stained with the substrate X-gal.	167+168
Figure 30.1 + 30.2: 30.1) Wholemount X-gal blue staining of heterozygous E12.5-E14.5 embryos from line 12 (with β AGG20 transgene). 30.2) Wholemount X-gal blue staining homozygous 4 month old adult mouse from line 12 (with β AGG20 transgene).	170
Figure 31: Paraffin sections of the line 12 (β AGG20 transgene) mouse tails (A) and hip (B) stained with the X-gal substrate and counterstained with the BR1 antibody.	172
Figure 32: Coronal sections of a representative 4 month old line 12 (β AGG20) mouse knee in which the hypertrophic blue-stained chondrocytes have been circled.	173
Figure 33: Sequence of the human aggrecan minimal promoter.	190

Results Chapter 2

Figure 34: Schematic diagram of the mT/mG construct before and after Cre-mediated recombination	201
--	-----

Figure 35: Bioluminescence imaging of luciferase activity in an adult transgenic mouse from the ColII4xA1Cre transgenic line	204
Figure 36: Ubiquitous mT and mG labelling of fixed hip sections in 'mT/mG'/ColII4xA1Cre' double transgenic mice (A1-3, B1-3 and C1-3) and an 'mT/mG' single transgenic (D1-3) post Cre-mediated recombination (1 week after final injection of tamoxifen).	206+207
Figure 37 : Ubiquitous mT and mG labelling of fixed hip (A1-3, B1-3), knee (C1-3, D1-3) and muscle (F1-3) sections in 'mT/mG'/ColII4xA1Cre' double transgenic mice and hip sections of an 'mT/mG' single transgenic (E1-3) post Cre-mediated recombination (2 week after final injection of tamoxifen).	208-210

Results Chapter 3

Figure 38: Basic anatomy of the knee.	219
Figure 39: Wholmount operated + unoperated knees from line 12 (β AGG20 line) expressing the β AGG20 transgene (blue staining) only on the medial side of right operated knees 4 weeks post-surgery (black arrows).	222
Figure 40 (A-E): A) Coronal sections of representative line 12 (β AGG20) operated (DMM) right knees (R) and contralateral unoperated left knees (L) 4 and 8 weeks after surgery that have been stained overnight with X-gal. B) High magnifications of coronal sections of the knees in figure 40A showing the lateral side of articular cartilage C) High magnifications of coronal sections of the knees in figure 40A showing the medial side of articular cartilage. D) High magnifications of coronal sections of the knees in figure 40A showing the lateral side of the tibial growth plate. E) High magnifications of coronal sections of the knees in figure 40A showing the medial side of the tibial growth plate.	224-230
Figure 41 (A-E): A) Coronal sections of a representative line 12 (β AGG20) loaded right knee (A2) and contralateral unloaded left knee (A1) that have been stained overnight with X-gal and counterstained with eosin or the β -galactosidase antibody. B) High magnifications of coronal sections of the knees in figure 41A showing the lateral side of articular cartilage. C) High magnifications of coronal sections of the knees in figure 41A showing the medial side of articular cartilage. D) High magnifications of coronal sections of the knees in figure 41A showing the lateral side of the tibial growth plate. E) High magnifications of coronal sections of the knees in figure 41A showing the medial side of the tibial growth plate.	233-237
Figure 42: Results of bioluminescence imaging of luciferase activity in adult transgenic mice from the ColII4xA1Cre transgenic line before and after DMM surgery	240

Conclusions

Figure 43: Proposed action of ' <i>cis</i> -regulatory elements' <i>cis</i> 1-6.	253
Figure 44: Schematic of hypothesised human aggrecan gene regulation	255

List of Tables

Introduction

Table 1: Histologic stages in long bone and epiphseal development.	48
Table 2: Extracellular matrix components of cartilage	56+57
Table 3: Some of the characteristics of the exons/introns of the human aggrecan gene	63

Methods

Table 4: List of constructs generated.	98
Table 5: Table representing the names of the primers used, their sequences and the generated products of each primer pair	107+108

Results Chapter 1

Table 6: Summary of mice transgenic and expressing β galAB (minimal promoter/ -452bp to +375bp).	136
Table 7: Summary of transgenic and expressing mice driven by the β AGG20 promoter (-20,054bp to +375bp).	141
Table 8: Average levels of expression for β AGG20 transgenic mice after overnight staining at room temperature.	141
Table 9: Summary of transgenic and expressing mice driven by the β AGG20 promoter (-20,054bp to +375bp).	144
Table 10: Average levels of expression for β AGG20 transgenic mice at E16.5 after overnight staining at room temperature	144
Table 11: Summary of transgenic and expressing mice driven by the β AGG13 promoter (-13,030bp to +375bp).	147
Table 12: Summary of transgenic and expressing mice driven by the β galAB+7 promoter (-20,054bp to -13,030bp fused to minimal promoter/ -452bp to +375bp).	152
Table 13: Average levels of expression for β galAB+7 transgenic mice after overnight staining at room temperature	152
Table 14: Summary of transgenic and expressing mice driven by the β AGG10 promoter (-20,054bp to -13,838bp fused to -4,182bp to +375bp).	153
Table 15: Summary of transgenic and expressing mice driven by the β galAB + <i>cis</i> 3 promoter (-10,703bp to -10,343bp fused to minimal promoter/-452bp to +375bp).	155

Table 16: Average levels of expression for β galAB+ <i>cis</i> 3 transgenic mice after overnight staining at room temperature	156
Table 17: Summary of transgenic and expressing mice driven by the β AGGAB ⁺⁹² + <i>cis</i> 3 promoter (-10,703bp to -10,343bp fused to minimal promoter <i>including missing 92 base pairs</i> /-452bp to +375bp).	161
Table 18: Average levels of expression for β AGGAB ⁺⁹² + <i>cis</i> 3 transgenic mice after overnight staining at room temperature.	161
Table 19: Summary of transgenic and expressing mice driven by the β AGG20 + <i>cis</i> 6 promoter (-20,054bp to +375bp fused to +5,000bp to +7,813bp).	165
Table 20: Average levels of expression for β AGG20 + <i>cis</i> 6 transgenic mice after overnight staining at room temperature	165
Table 21: Summary of mice transgenic and expressing all generated constructs for microinjection.	174
Table 22: Average levels of expression for all transgenic mice after overnight staining at room temperature	174+175

Results Chapter 3

Table 23: Average levels of expression of β -galactosidase in different compartments of the operated and unoperated knees of β AGG20 transgenic mice after overnight staining at room temperature	223
Table 24: Average levels of expression of β -galactosidase in different compartments of the loaded and unloaded knees of β AGG20 transgenic mice using overnight staining at room temperature with X-gal and β -galactosidase antibody staining	232

Acknowledgements

I am heartily thankful to my supervisors Dr George Bou-Gharios and Professor Jeremy Saklatvala, whose encouragement, guidance and intellectual support from the beginning to the very end enabled me to develop a scientific expertise in the subject of aggrecan. I would also like to make special reference to Tapan Bhattacharyya and Dr Phuong Vo for sharing their immense experience and insight but most of all for their friendship, without which I would have found it very difficult to finish this PhD. I would also like to thank my parents Najeeb and Jill who through my childhood and study career, have always encouraged me to follow my heart and inquisitive mind in any direction this took me. My parents provided me with a loving home, one where an academic mind was celebrated, and never lost faith in me. Last but not least, I would like to thank my fiancé Jean-Philippe for his ongoing love, support, patience and ability to make life a better place through all the tough times.

Abstract

Osteoarthritis (OA) is a degenerative joint disease and is the leading cause of physical disability in industrialised nations. Cartilage has received the most attention in the study of OA due to articular chondrocytes acting as potential instigators of disease. These cells are responsible for the anabolic-catabolic balance required for matrix maintenance due to their ability to synthesise the structural components of the extra-cellular matrix along with matrix-degrading proteases. In order to better understand the mechanism by which this balance is shifted in OA, it would be useful to investigate the roles of the genes that are expressed in these cells. One of the most powerful tools to do this would be an inducible and chondrocyte-specific system that utilises a cartilage-specific promoter, such as the aggrecan promoter.

Aggrecan, one of the major structural components of cartilage, is a large aggregating chondroitin sulphate proteoglycan. This PhD set out to further our understanding of the transcriptional regulatory mechanisms that govern the chondrocyte-specific expression of the gene. A series of constructs containing various combinations of non-coding aggrecan DNA (mostly upstream of the transcriptional start site and including the first untranslated exon) were used to generate transgenic mouse embryos and an adult line in which the cartilage-specific expression of the transgene indicated the presence of *cis*-regulatory elements. These studies have identified three putatively important regions. One serves as a basal/core promoter at and around the transcriptional start site (TSS) (containing what could possibly be a non-essential 92bp sequence depending on the developmental stage of the mouse). Another located 10kb upstream of the TSS can serve as an enhancer, and finally another located within a 7kb region further upstream. One or more of these elements seem to respond to injury as indicated by increased promoter activity in the OA-induced knees of the adult line containing all three regulatory regions.

Using a combination of our data alongside published work (Han and Lefebvre, 2008), a cartilage-specific Cre recombinase line was generated and tested for specificity and continual gene expression under the induction of OA. The work presented in this thesis can now be used to contribute to knowledge on the pathogenesis of OA.

Abbreviations

ATP	Adenosine triphosphate
bp	Base-pair
ChiP	Chromatin immunoprecipitation
dpc	Days post coitum
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
<i>E.coli</i>	<i>Escherichia coli</i>
EMSA	Electrophoretic mobility shift assay
GTF	General Transcription factor
HAT	Histone acetyltransferase
HDAC	Histone deacetyltransferase
HMT	Histone Methyltransferase
kb	Kilobase-pair
LCR	Locus control region
Mb	Megabase-pair
MED	Mediator
MMP	Matrix metalloproteinases
mRNA	Messenger ribonucleic acid
µg	microgram
µl	microlitre
ml	millilitre
ng	nanogram
OA	Osteoarthritis
PCR	Polymerase chain reaction
PIC	Preinitiation complex

Pol II	RNA polymerase II
SDS	Sodium dodecyl sulphate
TF	Transcription factor
TAF	TBP associated factor
TBP	TATA-box binding protein
TIMP	Tissue inhibitors of matrix metalloproteinases
TSS	Transcription start site
X-gal	5-bromo-4-chloro-3-indolyl-8-D-galactopyranoside

Introduction

1 Transcription

1.1 Introduction to transcription

The regulation of the spatio-temporal expression of genes in eukaryotes is central to processes such as development and differentiation, as well as homeostasis. Eukaryotic genes are complex and are organised within genomes that are in turn organised within compact nucleoprotein structures (chromatin). Studies over the past 35 years have indicated that different genes are regulated through several mechanisms, primarily transcription (Roeder, 2003).

Although approximately 9 years have passed since the human genome sequence was first mapped (Lander et al., 2001), much of the genome still remains unannotated. A mere 1.5% of the human genome is responsible for protein coding while 45% is composed of repetitive DNA, which are mainly remnants of retrotransposons accumulated over thousands of years of evolution. The remaining ~50% of the human genome, which is made up of non-coding sequences that were previously referred to as 'junk' DNA, is now thought to contain key regulatory elements that are crucial for the execution of the elaborate expression programs in the multiple cell types of the human body (Heintzman and Ren, 2009). Therefore, it is the DNA sequences in and around specific gene promoters that provide the important information which controls the time, location and level at which specific genes are transcribed. This information is generally thought to consist of three parts and can also be referred to as *cis*-regulatory DNA sequences: the core/basal promoter; the region proximal to this promoter; and finally the more

distal regulatory elements such as enhancer sequences; silencer elements and insulators. It is now widely accepted that tissue specific transcriptional activation of genes requires three distinct categories of transcription factors (*trans*-acting factors) to interact and assemble at the *cis*-regulatory DNA sequences of these genes. The first of these factors to be purified and characterised both biochemically and genetically were the transcriptional activators/repressors (also called transacting/transcription factors). Next came the 'basal factors' or the 'general transcription factors' (GTFs) and finally, the 'coactivators' (Meisterernst et al., 1991) or 'mediators' (Flanagan et al., 1991).

The 'basal' or 'core' transcriptional apparatus, independent of activators and coactivators, is sufficient for recognition of and low levels of accurate transcription from, common core promoter elements *in vitro* (basal transcription) (Roeder, 1996). However, the general consensus has always been that although this 'basal' transcriptional apparatus is nearly universal in its role as a highly conserved 'driver', it is the sequence-specific transacting factors (activators and repressors) that serve as the master switches in turning transcription of specific genes on or off in a cell- and gene- specific manner. This notion has always been accepted due to the high conservation of the components of the 'core machinery' amongst all species from yeast to humans. Transacting factors, on the other hand, can show major differences from organism to organism and their expression reflects cell-to-cell variation. It is now known that this model is overly simplified and that members of the 'core transcriptional apparatus' also serve as 'engines' that drive cell-specific gene expression too (Kuras et al., 2000), (Xiao et al., 2006), (Yan et al., 2005).

1.2 Components of the eukaryotic transcriptional apparatus

1.2.1 RNA Polymerases – RNA polymerase II transcription cycle

Eukaryotes use three functionally distinct classes of nuclear RNA polymerases (Roeder and Rutter, 1969): RNA polymerase I (which transcribes large ribosomal RNA genes); RNA polymerase II (which transcribes protein-coding and small structural RNA genes) as well as RNA polymerase III (which transcribes tRNA, 5SRNA and other small structural RNA genes) (Weinmann and Roeder, 1974). Each polymerase contains its own distinct composition of subunits (14, 12 and 17 subunits in RNA polymerase I, II and III respectively).

Eukaryotic RNA polymerase II typically has 12 subunits (RPB1-RPB12) and is incapable of specific promoter recognition (Myer and Young, 1998). RNA polymerase II transcribes mRNA, but that is only its most basic role. While transcribing mRNA, it also scans for damaged DNA and modifies the surrounding chromatin. It also acts, through its largest subunit, as a platform for several necessary mRNA processing factors.

The largest subunit of eukaryotic RNA polymerase II, RPB1, contains an unusual C-terminal domain (CTD) that is made up of repeats of the heptapeptide consensus sequence N-Tyr 1 - Ser 2 -Pro 3 -Thr 4 -Ser 5 -Pro 6 -Ser7 -C, of which there are 52 copies in mammals (Corden et al., 1985). The amino-acids in these repeats are subject to reversible post-translational modifications (phosphorylation, glycosylation and *cis-trans* isomerisation) (Phatnani and Greenleaf, 2006), (Meinhart et al., 2005). These modifications control many aspects of the functional state of pol II. These include the assembly of a multisubunit enzyme, its transportation to the nucleus, its localisation within subnuclear domains or onto DNA templates and finally, its ultimate destruction. It is also worth noting that lysine residues in the

several non-consensus CTD repeats that are also found in mammals can be ubiquitinated in response to DNA damage (Somesh et al., 2005).

This CTD serves as a scaffold for the recruitment and interaction of a large number of nuclear factors. This recruitment of transcription and processing factors throughout different stages of the transcription cycle is closely linked to the modification state of the CTD, in particular the position of the phosphorylated serines. RNA polymerase II is recruited to the promoters of active genes in a hypo-phosphorylated state. However, studies have shown that it is phosphorylated by TFIIF-mediated phosphorylation at Ser5 (Ser 5P) when it passes through the 5' end of a gene whereas Ser2 phosphorylation is prominent when the enzyme reaches the 3' end of a gene (Komarnitsky et al., 2000). These phosphorylated repeats help recruit and bind capping factors and 3' processing factors to the 5' end of newly transcribed RNA and the poly(A) sites at the 3' end respectively (Ho and Shuman, 1999). When pol II is in the middle of a gene, both Ser2 and Ser5 are phosphorylated.

Other interacting partners include the mediator complex (to regulate the initiation of transcription) as well as several histone methyltransferases. Rather than carrying all these factors throughout the transcription cycle, the CTD interacts differently with each factor at the correct time. The biological role of CTD phosphorylation remains to be fully elucidated but it is now thought that different phosphorylation patterns throughout the 52 repeats including additional phosphorylation of Ser7 in some of the repeats may allow for gene-specific regulation through the dynamic modification of DNA templates and processing of nascent RNA transcripts (Chapman et al., 2007).

1.2.2 *Cis*-regulatory sequences

1.2.2.1 *Core/basal promoters*

The core promoter lies at the heart of transcription and can be widely diverse in its structure and function. There are two known modes of transcription initiation – focused and dispersed (Juven-Gershon et al., 2006a), (Juven-Gershon et al., 2006b). In focused initiation, transcription starts at a single nucleotide or within a narrow region of multiple nucleotides. However, in dispersed initiation, there are multiple weak start sites that are found over a wider region of about 50-100 nucleotides. In complex organisms such as humans, around 70% of genes have dispersed core promoters which are typically found in CpG islands (genomic regions that contain a high frequency of CpG sites). It is generally thought that focused promoters are linked to inducible genes whereas dispersed promoters, are used for constitutively expressed genes.

The study of focused core promoters has led to the discovery of a set of common DNA sequence elements/patterns that are associated with these promoters. The most commonly known sequence motifs are the TATA box, BREu (upstream TFIIB recognition element), Inr (initiator), MTE (motif ten element), DPE (downstream promoter element), DCE (downstream core element), and XCPE1 (X core promoter element 1). These elements have unique qualities, which are linked to the expression of downstream genes. They can co-exist in the same promoter. However, specific combinations are more common than others with some patterns being complementary to one another. For example, a TATA-dependent core promoter with TATA+Inr motifs has different properties than a DPE-dependent core promoter with Inr+DPE motifs. In contrast to this, TATA, BRE, DPE and MTE motifs are not associated with dispersed promoters (Sandelin et al., 2007), (Carninci et al., 2006). Therefore, it is likely that

there are important and fundamental differences in the mechanisms of transcription between focused and dispersed promoters.

The following are examples of core promoter motifs involved in RNA polymerase II transcription:

1- The TATA box is probably the best-known transcription factor binding site and the first to be discovered. It is associated with strong tissue-specific promoters and often co-exists with initiator (Inr)-like sequences at the transcriptional initiation site. Although it is a well-known core promoter motif, it is only present in approximately 10-15% of mammalian core promoters (Kim et al., 2005), (Cooper et al., 2006);

2- The BRE element, (TFIIB recognition element) can either increase or decrease transcription rates in eukaryotes. Two separate recognition sites have been found: The BREu element, which lies upstream of a subset of TATA boxes and the BREd element, which lies immediately downstream of the TATA box. (Deng and Roberts, 2005);

3-The Inr is possibly the most common core promoter motif and encompasses the transcription start site (TSS). Although the initiator element can co-exist and act synergistically with the TATA box, it is in fact independent of it. The TATA and Inr motifs are the only known core promoter motifs that can alone recruit the pre-initiation complex (described in Introduction section 1.2.3.1) and initiate transcription (Smale and Baltimore, 1989);

4-The DPE (downstream promoter element) typically lies downstream of the TSS/Inr in TATA-less promoters of *Drosophila melanogaster*. It is thought to have a similar function to the TATA box in directing the pre-initiation complex to a nearby TSS. In some cases, however, the TATA, Inr and DPE motifs can be found in the same core promoter (Burke and Kadonaga, 1997), (Kutach and Kadonaga, 2000);

5- The MTE (motif ten element) is located immediately upstream of the DPE and like it, is conserved from *Drosophila* to humans. The MTE functions together with the Inr but can act independently of the DPE as well as the TATA box. However, there is synergy between the MTE and DPE as well as between the MTE and TATA box (Lim et al., 2004). Figure 1 shows details including the locations and consensus sequences of all the above-mentioned functionally *cis*-regulatory motifs.

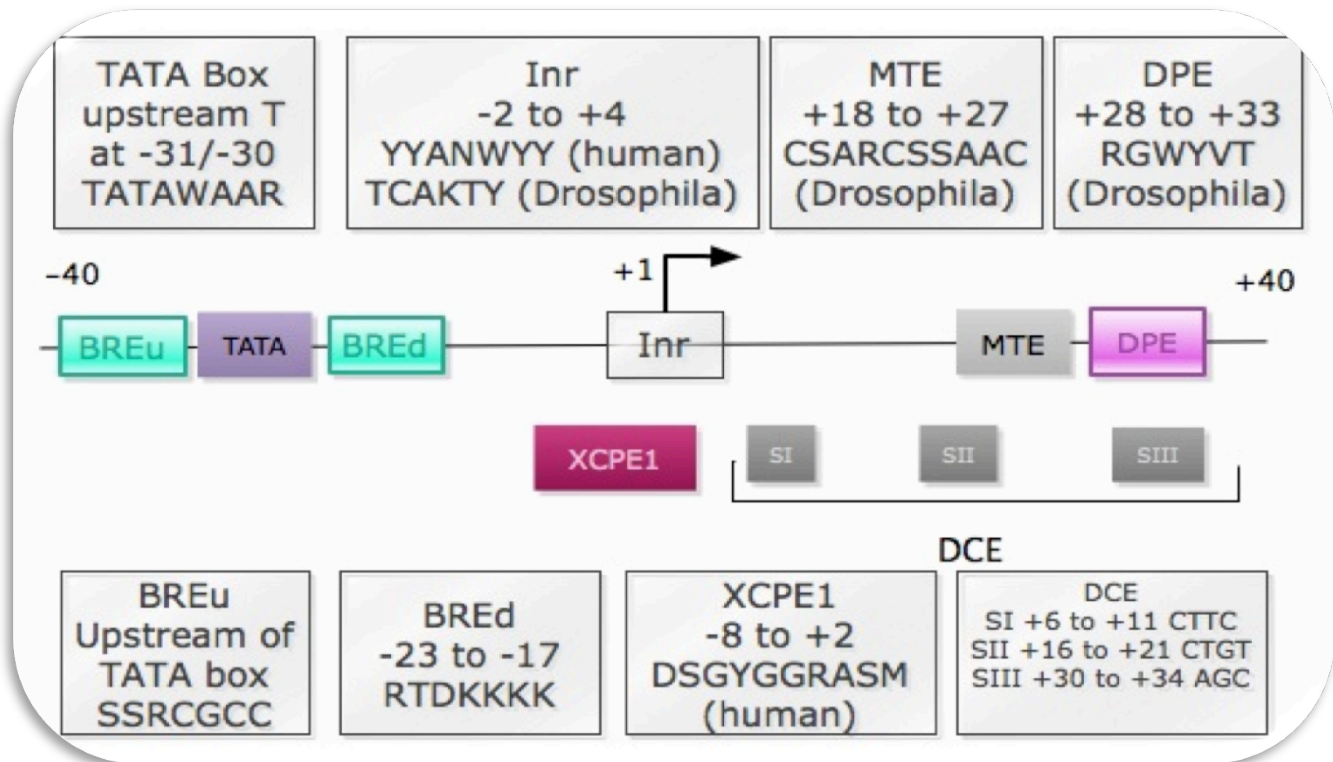


Figure 1: Some core promoter motifs for transcription by RNA polymerase II. This diagram is roughly to scale. The motifs are presented as IUPAC consensus symbols: N, any nucleotide; R, A or G (purine); S, C or G; V, A or C or T; W, A or T; Y, C or T (pyrimidine).

1.2.2.2 Distant regulatory elements

The other classes of *cis*-regulatory elements that are more distal to the core/basal promoters of genes include enhancers, silencers and insulators. Whereas **enhancers** act to recruit a complex array of transcription factors with chromatin-modifying functionalities that facilitate gene transcription (Lee and Young, 2000), **silencing elements** (silencers) bind proteins and/or modify chromatin structure to inhibit the transcription of genes (Ogbourne and Antalis,

1998). **Insulators**, can provide further regulation by restricting the spread of heterochromatin and impeding transcriptional enhancers from activating unrelated nearby promoters (West et al., 2002).

1.2.2.2.1 Enhancers

DNA exists in cell nuclei in a highly ordered compact form called chromatin (Felsenfeld and Groudine, 2003). The structural units of chromatin are nucleosomes, which are composed of 146 base pairs of DNA wrapped around a single histone octamer composed of two histone H2A-H2B dimers and one H3-H4 histone tetramer (Luger et al., 1997). In order for transcription to occur, the transcriptional machinery must first overcome the negative effects that chromatin has on the expression of genes namely, the inability of protein-DNA interactions required for transcription to occur because of the compaction of DNA. One of the main roles of enhancer elements is to recruit transcription factors that promote the decondensation of repressed chromatin. Another is to facilitate the assembly of the transcriptional machinery at gene promoters (Maston et al., 2006). Enhancer sequences can contribute to a gene's overall spatiotemporal pattern of expression in an additive fashion (Visel et al., 2009). An example of a gene that is controlled by multiple enhancer sequences with different specificities is the mammalian β -globin gene (Bender et al., 2001).

One of the most common features of enhancers is that they are scattered across large non-coding intervals (Nobrega et al., 2003). Some can also drive gene expression independently of their orientation (Alley et al., 1995). There are also exceptional cases of enhancers that can function at a distance of over 1 Mb upstream or downstream of their gene promoter target. An example of this is the Sonic hedgehog (*Shh*) gene whereby mutations in a conserved non-coding region in the 5th intron of the *Lmbr1* locus, which is 1 Mb away, can lead to mouse and human preaxial polydactyly with mirror-image digit duplications (Sagai et al., 2005). Enhancers can even be located within their target gene (intronic sequences) (Krebsbach et al., 1996).

Each enhancer can integrate and initiate regulatory prompts and is generally believed to have a binary on/off output. Genes can be regulated in a cell-specific manner due to enhancers containing binding sites for specialised sets of transcription factors (TFs) (Introduction section 1.2.3.2), which include activators and repressors. Therefore, a gene is only activated, or indeed repressed, in cell types where these factors reside. TFs bind to short regulatory consensus sequences (Stormo, 2000), which usually contain specified positions (base pairs) that are highly discriminative and some which are more dispensable.

Although the main function of enhancers as binding platforms for transcription factors is widely accepted, the details of their structural necessities are less known. Two models exist which embody two extremes of structural necessities: The first is that of the **enhanceosome model**. This model is based on the interferon- β enhancer, which functions as a specific assembly platform for precise protein complexes with rigid requirements on binding-site composition, order, orientation and spacing (Merika and Thanos, 2001). Given all these qualities, enhanceosomes would display near-perfect conservation across their entire lengths amongst different species, which as discussed in Introduction section 1.4, isn't always true for *cis*-regulatory sequences. The second model is that of the **billboard enhancer**. This type of enhancer can integrate regulatory output from individual trans-factors which can be bound in any order, orientation and spacing. As long as the sum of activating and repressing inputs remains constant, the individual factors could change. In reality, enhancers lie in between these two extremes.

1.2.2.2.2 Silencers

Silencers generally share most of the properties that are ascribed to enhancers. They are sequence-specific elements that confer a negative effect on the transcription of a gene. Most silencers can function independently of orientation and distance from the core promoter

sequence. However, some have been identified to have position-dependence. In terms of their location, they can be within a proximal promoter, as part of a distal enhancer or as an independent distal regulatory module (situated upstream of the TSS, in an intron or in a 3' untranslated region of a gene). They can act synergistically and cooperate in binding to DNA.

In contrast to enhancers, which act as binding sites for activators, silencers are binding sites for the negative transcription factors called repressors. Repressor function may, in some cases, require the recruitment of negative cofactors (Introduction section 1.2.3.3) called corepressors. There are a number of models that have been suggested for repressor function. In the first model, repressors may function by blocking the binding of a neighboring activator or by directly competing for the same site. In the second model, a repressor may inhibit activators and/or general transcription factors (Introduction section 1.2.3.1) from accessing a promoter through the establishment of a repressive chromatin structure by the recruitment of histone modifying activities or chromatin-stabilizing factors. Finally, in the third model, a repressor may block transcriptional activation by preventing the assembly of the preinitiation complex (Introduction section 1.2.3.1) (Maston et al., 2006).

1.2.2.2.3 Insulators/boundary elements

The main function of an insulator is to block genes from being affected by the transcriptional activity of nearby genes. In this way, insulators can limit the action of transcriptional regulatory elements to defined regions and divide the genome into individual domains of expression. Insulators are typically approximately 0.5kb – 3kb in length and function in a position-dependent, orientation-independent manner. They have two main characteristics: their ability to block enhancer-promoter communication, and their ability to stop the spread of repressive chromatin by acting as heterochromatin barriers. For some insulators, these characteristics can be separable (Recillas-Targa et al., 2002), (Maston et al., 2006).

The globin genes of vertebrates are found clustered within the genome. Mapping studies in erythroid cells have defined an open chromatin domain that contains the entire cluster. Within this domain, there are globin enhancers that temporally regulate these genes without affecting the expression of independently regulated genes flanking the globin domain. At the 5' end of the chicken globin domain is a folate receptor gene and its regulatory elements (HAS). Downstream of the globin cluster are the olfactory receptor genes (OR). The folate receptor genes are active when the globin genes are repressed and inactive when the globin genes become active during erythropoiesis. The OR genes, on the other hand, are not expressed in erythroid cells. Insulator elements are able to achieve this differential expression amongst the different genes within the cluster. The hypersensitive sites (5'HS4 and 3'HS1), which flank the active chicken globin domain, were proposed as candidate insulators due to their proximity between the differentially regulated promoters and regulatory elements. An element that resides within the human β -globin gene locus (Li et al., 1999a) is homologous to the chicken β -globin insulator, 5'HS4, which is the most well characterised insulator element in vertebrates. Molecular analyses of the 5'HS4 insulator in transgenic assays indicated that it is proficient in enhancer blocking as well as in protection against position effects (Valenzuela and Kamakaka, 2006).

1.2.2.3 Locus control regions (LCRS)

Other than basal/core promoter regions and the distal *cis*-regulatory elements described above, there are also **locus control regions** that facilitate the activation of a cluster of genes. Locus control regions (LCRs) are groups of regulatory elements that can organise an entire locus or gene cluster and are operationally defined as elements that direct tissue-specific, physiological expression of a linked *trans*-regulated gene. They work in a position-independent and copy-number-dependent manner. LCRs typically consist of multiple *cis*-acting elements (such as enhancers, silencers, insulators and nuclear-matrix or chromosome scaffold-attachment regions (MARs or SARs)) that are bound by transcription factors (tissue-

specific as well as ubiquitous), coactivators, repressors and/or chromatin modifiers. Although each of these components differentially affects the expression of a gene, it is their collective activity that functionally determines an LCR and dictates appropriate spatial/temporal gene expression. The most prominent characteristic of an LCR is its strong, specific enhancer activity. They frequently feature a cluster of nearby DNase I hypersensitive sites and are thought to provide an open chromatin domain for genes with which they are associated (Li et al., 2002).

LCRs have been identified in a broad spectrum of mammalian loci. The first and most-studied LCR to date is the mammalian β -globin LCR (Chakalova et al., 2005). The human β -globin locus has five genes that are differentially expressed throughout development, and are organised in order of their developmental expression. The β -globin LCR is located approximately 6-25kb upstream of the cluster of genes within the β -globin locus and drives high-level, erythrocyte lineage-specific expression of these genes. The activity of this LCR is orientation-dependent since inverting it destroys much of its function (Tanimoto et al., 1999).

1.2.3 *Trans*-regulatory molecules

1.2.3.1 *General transcription factors (GTFs)*

The transcription of a gene by the enzyme RNA polymerase II (pol II) requires the assembly of a core promoter along with a group of GTFs into a complex called the preinitiation complex (PIC) (Buratowski et al., 1989). On its own, RNA polymerase II is unable to recognise target promoters and so needs these general factors to achieve this function. Each of the GTFs bring specific functions to the transcription complex and have been broadly characterised (Thomas and Chiang, 2006).

TFIID is a TF, which is composed of TATA-binding protein (TBP) in a complex with at least 10 other TBP-associated factors (TAFs). It plays a significant role in the recognition of promoters and binds to specific DNA elements that can be found within the core promoter such as the TATA box, the DPE, the Inr and the MTE motifs. It does this with the help of TBP and TAFs (Chen and Hampsey, 2002), (Juven-Gershon and Kadonaga). Although TFIIA is not essential for the formation of the PIC, it can help stabilize the TFIID-promoter complex. TFIIIB can bind the TFIID-promoter complex by DNA recognition as well as direct interaction with TFIID. Along with other factors such as mediator, it supports the stable binding of TBP to the TATA-box of an activated gene (Li et al., 1999b), (Kuras and Struhl, 1999). It is essential for the assembly of the PIC since it is needed for promoter recognition, the selection of the TSS as well as the recruitment of pol II. Pol II, in turn, enters the PIC pre-bound to TFIIF, which prevents its non-specific binding to DNA. It does this by serving as an early 'checkpoint' for correct PIC assembly by the recognition of a correct TFIID-TFIIA-TFIIIB complex at the TATA box. TFIIF interacts with the assembled TFIID-TFIIA-TFIIIB complex through TFIIIB. After all the above-mentioned components of the PIC have bound, TFIIIE and TFIIH enter and complete this complex. Structural experiments that were performed on pol II revealed the presence of a flexible 'jaw' which closes around the DNA when the enzyme is associated with DNA at an active promoter (Cramer et al., 2001), (Gnatt et al., 2001). TFIIIE appears to bind directly to pol II at the position of this 'jaw' suggesting that it is involved in regulating its opening and closing (Leuther et al., 1996). TFIIH, on the other hand, is responsible for the phosphorylation of Ser 5 on the CTD of pol II. Due to the complex array of protein-DNA and protein-protein interactions, the PIC is a very stable structure (Figure 2).

It has been suggested that the assembly of the PIC possibly occurs in a single concerted recruitment event, which brings to the promoter a large preassembled 'pol II holoenzyme' consisting of the enzyme bound to the mediator complex as well as most, if not all, the GTFs.

However, previous *in vivo* and *in vitro* evidence have shown that the GTFs (such as TFIID, TFIIA and TFIIB) (Zawel et al., 1995) and the mediator complex (Svejstrup et al., 1997) actually stay behind at the promoter as pol II is engaging in transcript elongation. This would allow for the fast entry of new polymerases for re-initiation of the transcription of the gene. If the holoenzyme recruitment model were to hold true, it would be necessary to postulate different forms of the enzyme for 'first time' and 'ongoing' transcription of a gene.

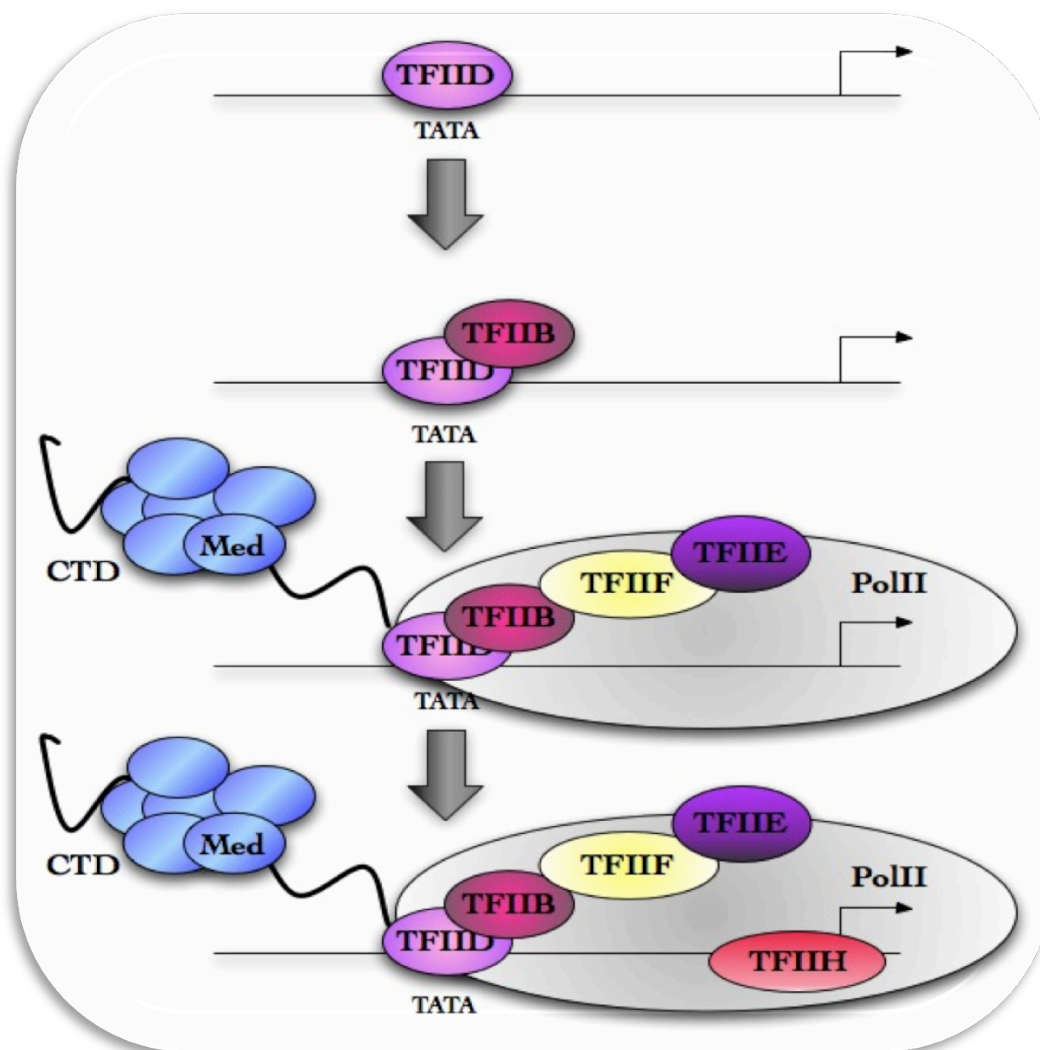


Figure 2: Mechanism of preinitiation complex assembly. The promoter DNA (straight black line) is shown with a TATA sequence upstream of the TSS (arrow). TFIID binds to the TATA element, followed by TFIIB. TFIIF and RNA polymerase II enter the complex together. The C-terminal domain (CTD) of RNA polymerase II (wavy line) binds to the mediator complex (Med) at this stage. TFIIE and TFIIH complete the complex.

1.2.3.2 *Transcription factors*

The recognition of DNA response elements by transcription factors is a key step in transcription (Georges et al.). The response elements can be present within the core promoter near the TSS, in a stretch of regulatory DNA several base pairs away or within an enhancer/silencer sequence (Gondor and Ohlsson, 2009). Transcription factors are part of a large group of easily identifiable gene regulatory proteins and represent between 5 and 10 percent of the coding capacity of human genomes. The relative ease with which these proteins can be characterised comes from early experiments, which delineated their modular structural architecture consisting of well-defined functional motifs. These include DNA binding, activation, oligomerization and regulatory domains (Babu et al., 2004). One primary function of the regulatory domain is to recruit cofactors (discussed in Introduction section 1.2.3.3) that carry chromatin-remodelling activities or can directly interact with pol II transcriptional machinery (Ptashne and Gann, 1997).

The classification of TFs is based on a common structural framework of the DNA binding domain (Narlikar and Hartemink, 2006). They are organised hierarchically and the main types include homeodomain TFs, POU TFs, basic helix-loop-helix TFs, leucine zipper TFs, zinc finger TFs, nuclear hormone receptor TFs and finally HMG box-containing TFs. The Arabidopsis transcription factor database (AtTFDB, <http://arabidopsis.med.ohio-state.edu/AtTFDB/>) lists 41 families. Each family consists of paralogous genes that encode structurally and functionally similar TFs that can recognize the same, or very similar, consensus DNA target sequences (Georges et al.).

The organization of TFs can often occur in a dimeric or tetrameric manner, and because each of the monomers binds to similar sequences, the DNA response elements in general consist of two half-sites. Although the half sites of a particular TF typically show high sequence similarity,

they can in fact have different sequences as well as different organization and spacers between them and can vary in their distances from TSS. It can thus be unclear what constitutes a specific TF target sequence. Examples of these TFs include the p53 (Menendez et al., 2009) and nuclear hormone receptor (Sandelin and Wasserman, 2005).

The earliest transcription factors- for genes transcribed by pol II- to be identified were those that acted through core promoter-specific DNA response elements. These included Sp1 (Dyran and Tjian, 1983), HSF (Parker and Topol, 1984) and USF (Sawadogo et al., 1988). Coupled with earlier experiments performed on the glucocorticoid receptor (Payvar et al., 1981) as well as others performed in 1985 on genetically identified yeast regulatory factors, these studies provided strong evidence for factors that regulate the activity of pol II by binding and interacting with distal DNA control elements. The original *in vitro* mechanistic experiments implicated TFIID as an assumptive target for activators (Sawadogo and Roeder, 1985).

1.2.3.3 Transcription co-factors

Although the general transcription apparatus is structurally very complex and transcriptional activators have tissue and DNA binding site specificity, other factors are required for the transcriptional machinery to be functionally active *in vivo*. These factors are defined as those that are required for the function of DNA-binding activators and not for basal transcription *per se* and do not show any independent site-specific binding. They have been named the 'coactivators' and are essential for 'activator' function. The coactivators fall into several nonmutually exclusive groups: 1- The factors intimately associated with the general transcriptional machinery, including the TFIIA, TAF components of TFIID, the SRB/MED components associated with RNA polymerase II, related complexes that catalyze nucleosomal histone acetylation, the SWI/SNF complex and related chromatin-remodeling complexes. 2- the general cofactors derived from the USA (upstream stimulatory activity) fraction in human

cells. In the early nineties, a crude fraction was isolated from human tissue that strongly enhanced stimulation of transcription by activators such as the E box-binding protein USF and the GC box-binding protein Spl in a purified transcription system. This co-factor fraction was termed USA and has been broken down into negative cofactors (termed NCs) as well as positive cofactors (termed PCs). Both act in concordance to selectively repress basal transcription while enabling activator-dependent transcription. 3- the factors that are either gene or cell type specific whose primary interactions are with DNA-bound activators, such as the B cell-specific OCA-B, the more pleiotropic CBP and p300 and nuclear hormone receptor-interacting factors (Roeder, 1998), (Kaiser and Meisterernst, 1996), (Hampsey, 1998). Some of the above factors in Introduction sections 1.2.3.3.1, 1.2.3.3.2 and 1.2.3.3.3.

1.2.3.3.1 The SRB/Mediator complex

The most significant of the above-mentioned co-activators is the 'mediator' or SRB/mediator complex. The human mediator is a counterpart of the yeast mediator with which it has strong conservation structurally and functionally. Human mediator was first discovered through a stable intracellular thyroid hormone-dependent association with thyroid hormone receptor (TR). This approximately 25-subunit complex (which was named TRAP TR-associated protein) was experimentally shown to be necessary for TR-dependent transcription from DNA templates by pol II and GTFs (Fondell et al., 1996). Subsequently, a number of site-specific DNA-binding regulatory factors have been shown to function through direct interactions with distinct Mediator subunits (encoded by *Srb* and *Med* genes) and the large complex is seen as a multivalent control panel, which has been proven to be the main passage for communication and interaction between the general transcriptional machinery (including Pol II and the GTFs) and the DNA-bound activators/repressors (Malik and Roeder, 2000).

1.2.3.3.2 Chromatin-Remodelling Complexes

Another group of cofactors are those that act through modifications of chromatin structure. Other than the DNA packaging functions of nucleosomal and higher order chromatin structures, there are other functions that are linked to gene regulation. These include imposing a barrier to the general transcription machinery that otherwise exhibits intrinsic potential to transcribe histone-free DNA templates through interactions at core promoters (reviewed in (Clapier and Cairns, 2009)). They were discovered through earlier studies in which purified pol II and GTFs couldn't interact with DNA templates that had assembled into chromatin, whereas the complete formation of a PIC or a TFIID-promoter complex prior to chromatin assembly actually enabled and facilitated subsequent transcription initiation (Workman and Roeder, 1987), (Lorch et al., 1987). Furthermore, histone depletion experiments that were performed in yeast confirmed a general repressive function for histones and demonstrated that some factors act mainly to overcome this repression, whereas others bring about a level of activation beyond that seen by histone depletion (anti-repression plus net activation) (Han and Grunstein, 1988). The SWI/SNF complexes are one class of cofactor proteins that modify the chromatin structure non-covalently in an ATP-dependent fashion. Once recruited to enhancer elements, these proteins can reposition specific nucleosomes along the DNA which enables core promoters to be exposed (Kingston and Narlikar, 1999).

1.2.3.3.3 Chromatin modifying enzymes

Another class of cofactors remodels chromatin structure by introducing covalent modifications to the N-terminal tails of histones. It is now known that a wide variety of modifications that occur on the N-terminal tails of histones are coupled with, and presumably effect specific gene activations and repressions. These include acetylation, deacetylation, methylation, phosphorylation and ubiquitylation and are brought about by acetyltransferase, deacetylase, methyltransferase, phosphorylase and ubiquitin enzymes respectively. For example, the enzyme histone acetyltransferase (HAT) transfers an acetyl group onto specific lysine amino

acids on the histone tails. This addition diminishes the positive charge of the histone, which leads to the loosening and weakening of the electrostatic bonds between it and the DNA around it, opening the chromatin and enabling the transcriptional machinery to access regulatory sites. A number of HATs have been identified that catalyze the acetylation of histones at specific residues. The protein complexes that catalyze histone acetylation include PCAF, CBP, p300, GCN5, TR-RAP complex and others (Rosenfeld et al., 2006). The functional antagonist of HATs are the histone deacetylases (HDACs), which in contrast, remove acetyl groups and prevent further acetylation, thus maintaining a closed chromatin structure (Strahl and Allis, 2000), (Lachner and Jenuwein, 2002), (Berger, 2002), (Vaquero et al., 2003).

It is important to note that a second level of regulation occurs not on the histone proteins, but rather directly on the DNA. The methylation of DNA, which is indeed the classic epigenetic alteration, involves the addition of a methyl group (CH_3) onto the cytosine residues of DNA (Bird, 1986). This event is associated with the silencing of transcription. This inhibitory effect works in two ways. First, densely methylated DNA can interrupt transcription factor binding which automatically silences expression. The second way is a little less direct and more suitable to regions with more dynamic changes in gene expression. The presence of methyl groups on selected cytosine residues attracts a class of proteins known as methylated-DNA binding proteins, which can further go on to attract an entire cluster of proteins that form a repressor complex (HDACs present in this complex) (Bird, 2002), (Klose and Bird, 2006).

Figure 3 illustrates the interactions that occur between the *cis*-regulatory and *trans*-regulatory elements discussed above in order to achieve efficient activation of a gene.

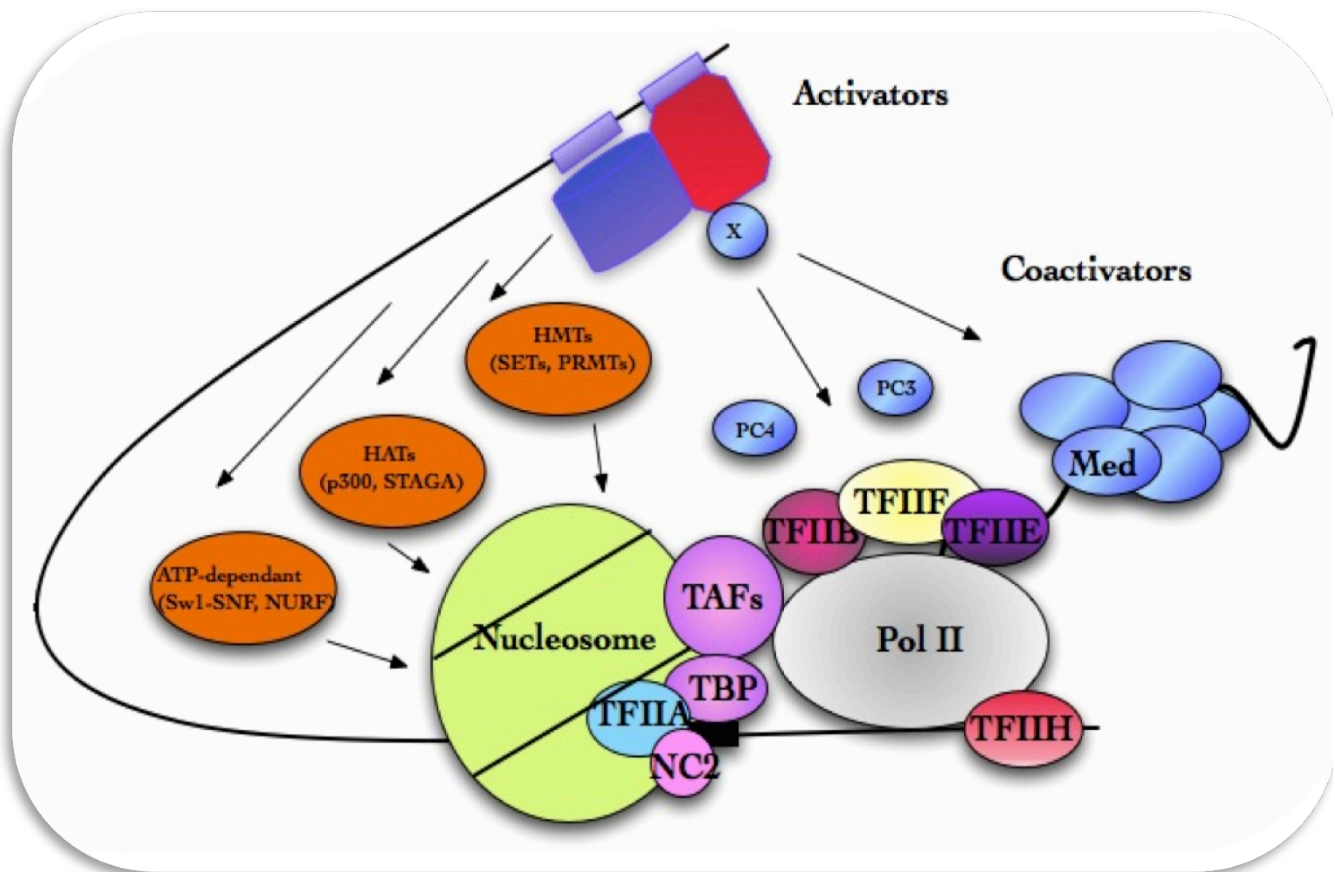


Figure 3: Eukaryotic transcription factors and activation mechanisms. The ultimate purpose of DNA-binding activators is to mediate Pol II recruitment and function. The effects of these activators on Pol II are mostly indirect and involve several interactions with diverse cofactors that modify chromatin structure and ease the recruitment of Pol II and general initiation factors. Chromatin remodelling and modifying factors including ATP-dependent factors, HATs and HMTs are shown in orange. Cofactors that act at the level of DNA are shown in shaded blue and include the Mediator, general positive cofactors such as PC3 and PC4, cell specific factors (X) such as OCA-B and OCA-S. The negative cofactor such as TBP-interacting NC2 is shown in pink.

1.3 The transcription cycle

The synthesis of eukaryotic RNA proceeds through multiple stages. These include the assembly of the **pre-initiation** complex, **initiation**, **promoter clearance**, **elongation** and finally **termination** of transcription. The **pre-initiation** of transcription starts with a transcriptional activator recruiting co-factors that serve as nucleosome remodellers (such as the SWI/SNF complexes) as well as those that act as nucleosome-modifying enzymes (such as HATs) to the nucleosomes of promoters of genes to be transcribed. This allows GTFs and Pol II to access the promoter (Fuda et al., 2009). Approximately 11-15 base pairs around the

TSS are unwound to form an 'open complex' that allows the single-stranded DNA template to eventually enter the active site of Pol II (Cramer, 2004).

During the **initiation** phase of transcription, a second activator binds to the promoter, interacts with and encourages the binding of the GTFs (such as TBP, TFIID, TFIIA and TFIIB), which eventually leads to the assembly of the PIC and recruitment of hypophosphorylated Pol II. It also recruits other co-activators such as the mediator, which can interact with the GTFs and increase expression levels (Kim et al., 1994). The DNA template is aligned in the active site of Pol II where nucleotides are paired with the template and are joined progressively to form the RNA transcript. The first ~8-15 nucleotides of RNA are formed within the pol II-promoter complex as, at this point, the transcription complex is functionally unstable and pol II requires the active participation of auxiliary TFs. However, once these initial bonds have formed, pol II can escape the promoter ('promoter escape') and form an early pol II-DNA transcription elongation complex from which all the GTFs rapidly dissociate (Pokholok et al., 2002). As opposed to 'promoter escape', '**promoter clearance**', which describes the availability of the promoter for binding of a new preinitiation complex, requires the physical exit of the enzyme and usually only occurs once the catalytic site has reached at least 60 base pairs downstream of the TSS. Only a fraction of initiation events lead to promoter escape/clearance, as there is often a tendency to release the transcribed RNA and produce truncated RNA transcripts since the polymerase is not yet committed to transcript elongation. This is called abortive initiation and commonly occurs in eukaryotes. Promoter clearance coincides with the phosphorylation of Ser 5 on the CTD of Pol II by TFIIH (Goldman et al., 2009), (Dvir, 2002). Like most of the remaining stages of transcription, promoter escape/clearance is an energy-dependent process that requires the consumption of adenosine triphosphate (ATP).

Following promoter escape, the RNA transcript continues to be stably bound in the transcription complex, but there are still instabilities that are associated with the newly formed early elongation complex (EEC) that differentiate it from a fully mature **elongation** complex. The tendency for transcript slippage persists until it reaches approximately 23 nucleotides in length. So the EEC undergoes continued adjustments- often accompanied by transcriptional pausing near the promoter- before becoming a fully productive elongation complex. Promoter-proximal pausing occurs when Pol II pauses in the 5' region of the transcription unit and only progresses efficiently into productive elongation upon stimulation by appropriate signals. This pausing is thought to function as a regulatory checkpoint that takes place before the EEC commits to productive elongation (Lis, 1998). DSIF (DRB sensitivity inducing factor (consisting of the elongation factors Spt4 and Spt5)) and NELF (Negative elongation factor) are negative factors that are thought to cooperate to repress transcription elongation and promote proximal pausing. Several other factors are required for the release of paused Pol II into productive elongation, after which Pol II continues through the rest of the gene. It is thought that the positive transcription-elongation complex (P-TEFb) phosphorylates DSIF, NELF and Ser 2 of the pol II CTD, which in turn relieves the paused polymerase of the negative effects of hypophosphorylated DSIF and NELF (Peterlin and Price, 2006). As pol II travels through the coding region, numerous factors (e.g FACT, Spt4, Paf1, TREX complexes, Spt6 and Swi/Snf) journey with it (Sims et al., 2004). FACT and Spt6 have been linked with histone disassembly and reassembly. FACT facilitates the displacement of H2A-H2B dimers throughout *in vitro* transcription and so promotes Pol II elongation through chromatin (Belotserkovskaya et al., 2003). Spt6, on the other hand, can interact physically with histones *in vitro* (particularly H3 and H4) and has histone chaperone activity (Bortvin and Winston, 1996). They both, however, are important for the elimination of false transcription initiation within the coding regions of the gene by promoting reassembly of histones. Paf1 might provide a platform for the association of complexes that modulate the structure of chromatin (Adelman et al., 2006). The multi-subunit

TREX complex is thought to ease transcription elongation by restricting the unsuitable pairing of the nascent RNA with the DNA template (Huertas and Aguilera, 2003). The Swi/Snf complex continues to remodel the chromatin of transcribed gene throughout elongation.

CTD phosphorylation appears to play a crucial role in pairing Pol II elongation to post-transcriptional steps such as mRNA capping, splicing, polyadenylation, export and chromatin modifications such as histone methylation. These processes are extremely important since the production of a mature messenger RNA (mRNA) involves the three main RNA processing events splicing, capping and polyadenylation. The process of splicing removes the non-coding introns from the transcribed RNA. Different protein isoforms can be translated by the selective inclusion of different coding sequences (called alternative splicing). The process of capping requires the addition of a methylated guanosine to the 5' end of the RNA through an unconventional 5' to 5' triphosphate linkage. This usually occurs during transcription of the first 40 nucleotides (Rasmussen and Lis, 1993). Polyadenylation, on the other hand, requires the addition of a string of adenines to the 3' end of the RNA molecule subsequent to the cleavage of the transcript. The processes of capping and polyadenylation enhance the stability of the newly formed transcript.

Cleavage and polyadenylation are referred to as the 3' processing events and take place after the transcription of the poly(A) site at the 3' end of a gene. The cleavage of the transcript coincides with the completion or **termination** of transcription. The termination of transcription involves the dissociation of the Pol II transcription complex from the DNA template. There are two models that have been proposed. They are not entirely mutually exclusive. The first is the allosteric model, which requires a conformational change to take place in the polymerase following transcription of the poly(A) sequence. This, in turn, makes the enzyme accessible to termination. PCF11 is a protein factor that has been found to trigger the break down of the Pol

II transcription complex in a CTD-dependent manner (Zhang and Gilmour, 2006). The second is the torpedo model, in which a 5' to 3' RNA exonuclease degrades any RNA that is still associated with Pol II following the cleavage and release of the RNA transcript. This leads to the termination of Pol II that has already served its purpose. In support of this model, there are certain genes in yeast and humans (Rat1 in yeast and XRN2 in humans) whereby their termination is dependent on a 5' to 3' exonuclease. Figure 4 illustrates the cycle of transcription.

1.4 The evolution of transcriptional control across eukaryotic species

Recently, the comparison of RNA expression levels between closely and distantly related vertebrate species has demonstrated tissue-specific conservation of gene expression. A recent study compared gene expression levels of over 3000 unambiguous gene orthologs from 10 functionally equivalent vertebrate tissues of pufferfish, frog, chicken, mouse, and human (Chan et al., 2009). The profiles were more comparable and alike amongst homologous tissues of different species than amongst tissues of the same species.

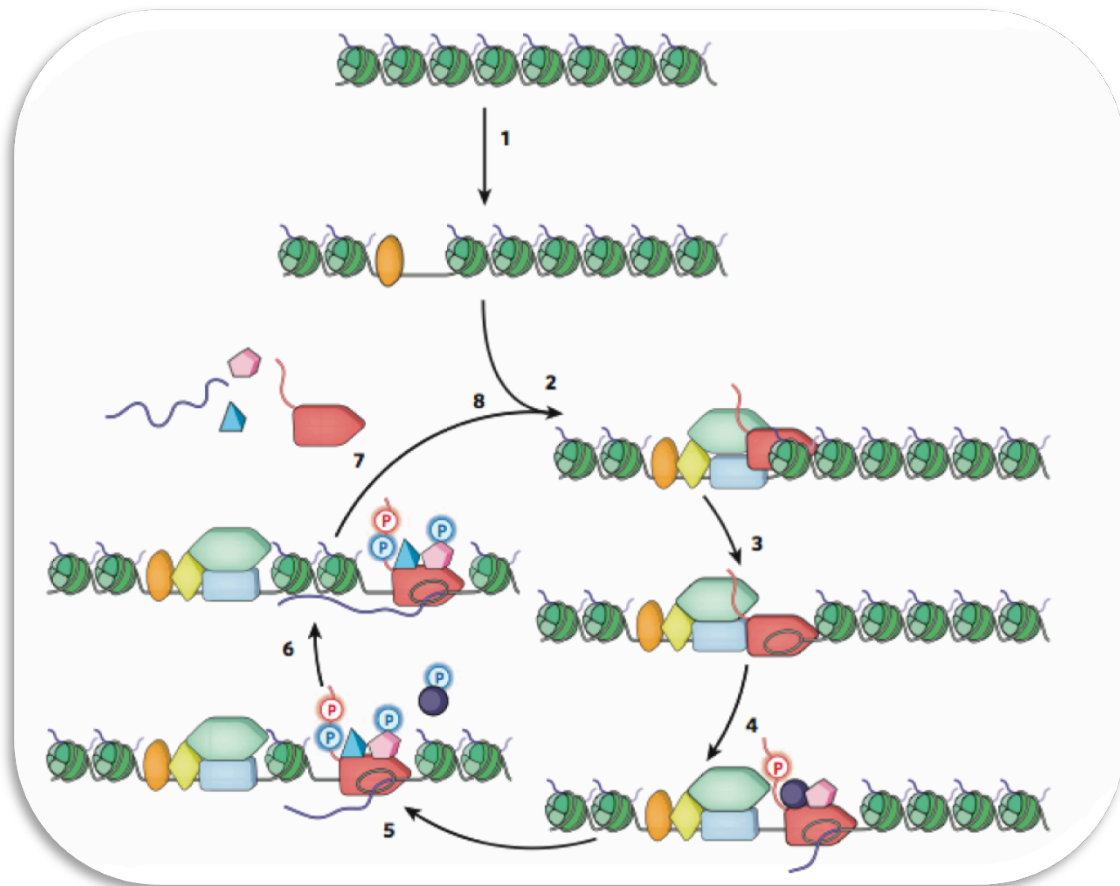


Figure 4: Multistages of the transcription cycle. Step 1: chromatin opening. The repressed gene and regulatory region to be transcribed are entirely packaged as nucleosomes (green). An activator (orange oval) binds and recruits nucleosome remodellers (HATs/Swi/Snf complex) to clear the promoter. Step 2: PIC formation. A second activator (yellow diamond) binds, promotes the binding of GTFs (blue rectangle) and recruits more coactivators such as the mediator (green hexagon), facilitating Pol II (red rocket) entry to the PIC. Step 3: initiation. DNA is unwound (oval inside Pol II) at the TSS, which leads to the formation of an open complex. Step 4: promoter escape/clearance. Pol II breaks contacts with promoter-bound factors, transcribes approximately 25 base pairs downstream of the TSS, produces an RNA (purple line) and pauses, partially mediated by DSIF (Spt4 + Spt5) (pink pentagon) and negative elongation factor (NELF) complex (purple circle). The Ser residues at position 5 (Ser 5) of the Pol II CTD repeats are phosphorylated (red P) during this step. Step 5: escape from pausing. P-TEFb (blue triangle) is recruited directly or indirectly by the activator and phosphorylates Ser 2 of the Pol II CTD repeats, SPT5 and the NELF subunits (blue Ps). NELF dissociates from the rest of the complex. Pol II escapes from the pause, either terminating or entering productive elongation. Step 6: productive elongation. Nucleosomes are disassembled and reassembled as the Pol II elongation complex transcribes through the gene due to the factors present within the elongation complex. Step 7: termination. After the Pol II complex transcribes the gene, it is removed from the DNA, and the RNA is released. Step 8: termination of Pol II by the allosteric or torpedo model or both. Taken from (Fuda et al., 2009).

It is important to note that although there is high conservation of gene expression across different species, this does not necessarily correlate with the amount of nearby conserved non-exonic DNA. This underlines that tissue-specific expression patterns can be maintained in spite of wide sequence divergence in regulatory DNA (Fisher et al., 2006). In contrast to this, many powerful comparative computational approaches have used the fact that sequence constraint can correlate with experimentally validated regulatory regions in DNA in order to predict these regions (Wasserman and Sandelin, 2004). Ultraconserved (perfect sequence identity of at least 200bp) or extremely conserved elements in human vs pufferfish comparisons were shown to be sufficient to identify putative gene regulatory elements (Pennacchio et al., 2006).

The controversy over conservation versus divergence of regulatory sequences could possibly be explained by the relevance that these sequences have to selectivity. Many of the *cis*-regulatory elements detected by ChIP might be functional or biochemically active yet selectively neutral (do not confer a selective advantage or disadvantage to the organism) and thus lack evolutionary conservation. Relevant elements can possibly be identifiable due to their evolutionary conservation (Birney et al., 2007).

2 Skeletal development

2.1 Introduction to skeletal development

The vertebrate skeleton is a complex organ with more than 200 bones. There has been considerable progress in the genetic and molecular understanding of all aspects of skeletal development over the last 10-20 years. One of the main characteristics of the skeleton is that it is an organ made of two distinct types of tissue, cartilage and bone. The main cell types in bone are osteoblasts and osteoclasts whereas in cartilage, it is chondrocytes. Chondrocytes and osteoblasts are both of mesenchymal origin and share a common progenitor while osteoclasts are of haematopoietic origin (Ducy et al., 1997), (Karsenty, 2008).

The formation of bone can be seen as the endpoint in skeletal development. The process is started with mesenchymal cells from three distinct embryonic lineages: The craniofacial skeleton is the product of **cranial neural crest cells**, the axial skeleton is derived from the **paraxial mesoderm (somites)** while **lateral plate mesodermal cells** form the limb or appendicular skeleton. The cells from these separate lineages migrate to their correct locations in the embryo where the appropriate future skeletal elements will develop. They develop due to all these mesenchymal cells aggregating to form mesenchymal condensations of high cell density (Hall and Miyake, 1992).

The cells can then follow either one of the following pathways:

The first, which comprises most of the vertebral bones such as the long bones of the limbs and elements of the axial skeleton, involves most embryonic mesenchymal condensations. In this case, they differentiate into cartilage cells (chondrocytes) that form the template, or anlagen, of

the future bone. Later on, through a process called endochondral ossification (described in Introduction section 2.2), this cartilaginous template is eventually replaced by bone.

In the second case, which comprises fewer bones such as the flat bones of the skull and clavicle, the cells of the mesenchymal condensations bypass the cartilaginous intermediary step and differentiate directly into osteoblasts. This is called intramembranous ossification (Hall, 1988).

2.2 Endochondral ossification- formation of the primary ossification centres (POCs).

The first change indicative of beginning the process of endochondral ossification occurs around the centre of the future bone shaft. The process of endochondral ossification can be divided into 5 stages. The first is the commitment of mesenchymal cells to becoming cartilage cells, which is achieved by the expression of the two transcription factors Pax1 and Scleraxis (Cserjesi et al., 1995), (Sosic et al., 1997)

During the second phase, these committed mesenchymal cells condense into compact nodules (or mesenchymal condensations) that eventually differentiate into chondrocytes. The formation of these nodules is initiated by N-cadherin (Oberlender and Tuan, 1994). N-CAM, however, seems to be critical for maintaining them (Hall and Miyake, 1995). These precartilaginous condensations express the transcription factor *sox9*, a high mobility group (HMG) box-containing DNA binding protein. It seems that mutations in this gene can cause campptomelic dysplasia, a rare disorder of skeletal development (Wright et al., 1995), (Kist et al., 2002). Additional proof of this transcription factor's importance comes from an elegant mouse embryo chimera experiment. In this experiment, both copies of *sox9* were inactivated in embryonic stem (ES) cells. The mutant ES cells were then injected into blastocysts from wild-

type females to generate chimeras. Further analysis of these chimeras demonstrated that *sox9*^{-/-} cells were able to contribute to multiple cell lineages but always failed to differentiate into chondrocytes. This block occurred at the mesenchymal condensation stage, identifying *sox9* as the earliest determinant of the chondrocyte cell fate (Bi et al., 1999), (Bi et al., 2001).

In addition to this, two other HMG-containing proteins, L-*sox5* and *sox6*, are co-expressed with *sox9* in all precartilaginous condensations and continue to be expressed in the cartilages of mouse embryos. They bind to the chondrocyte-specific enhancer of $\alpha 1$ (II) collagen *in vitro* and induce its expression in non-chondrocytic cells. Together with *sox9*, they induce endogenous expression of chondrocyte differentiation marker genes such as $\alpha 1$ (II) collagen and aggrecan *in vivo*. Although L-*sox5* and *sox6* have a high degree of homology to each other, their only similarity to *sox9* is their HMG box (Yang and Karsenty, 2002).

Throughout the third stage of endochondral ossification, chondrocytes proliferate rapidly to form the cartilage model for the bone. During their division, they secrete a cartilage-specific extra-cellular matrix (ECM) and form two distinct subpopulations of cells that can be separated by morphological and biochemical characteristics into round, low-proliferating chondrocytes (distal chondrocytes) and high-proliferating chondrocytes that are organized into columns (columnar chondrocytes). The distal chondrocytes are found at the future distal joint surfaces of the developing bone while the columnar ones are present more towards the center (Long et al., 2001).

In the fourth phase, some of the columnar chondrocytes exit the cell cycle and no longer divide. They differentiate into prehypertrophic and hypertrophic chondrocytes increasing their volume dramatically thus contributing significantly to the elongation of the skeletal elements. The hypertrophic chondrocytes produce collagen X and fibronectin, which they release into the

ECM. This enables the ECM to become mineralized (calcified) by calcium phosphate. Hypertrophic chondrocytes also secrete VEGF, the pro-angiogenesis factor in order to transform mesodermal mesenchymal cells into blood cells and vascularise the region (Gerber et al., 1999).

In the fifth phase, the cells that surround the cartilage model (fibroblast-like cells that form a layer called the perichondrium) differentiate into osteoblasts, forming the periosteum. Blood vessels from the periosteum, which were produced in the previous phase, invade the calcified matrix of hypertrophic chondrocytes, providing osteoclasts to degrade the mineralized cartilage matrix as well as osteoblasts and hematopoietic cells. This results in the formation of primary ossification centres. Within these centres, hypertrophic cartilage matrix is degraded, the hypertrophic chondrocytes undergo apoptosis (Hatori et al., 1995) and the osteoblasts replace chondrocytes with trabecular bone forming bone marrow. At the same time, the osteoblasts in the perichondrium begin to construct a bone collar of compact bone around the middle portion (diaphysis) of the dying cartilage resulting in the primary ossification centre being located inside a tube of bone (Wagner and Karsenty, 2001), (Hall and Miyake, 2000). Eventually, most of the cartilage is replaced by bone, leaving permanent hyaline cartilage only at joints or articular surfaces of bones.

2.3 Synovial joint formation

Synovial joints are necessary for skeletal function. They allow vertebrates to display a great range of motions by articulating bones with one another. They consist of several highly specialized tissues such as bone, articular cartilage, tendons, ligaments, synovium and joint capsule. The articular cartilage coats the surface of bones and enables pain-free and deformation-free movement. The tendons ensure articular motion by transmitting muscle force to bones. Ligaments, stabilize the joints by interconnecting bones. The synovium consists of a

synovial membrane that produces a lubricating synovial fluid rich in nutrients for the avascular articular cartilage and internally lines the joint capsule. This fibrous capsule, in turn, surrounds the joint and seals its cavity. The most complex synovial joint is the knee due to it featuring additional exclusive tissues. These include shock-absorbing menisci, cruciate ligaments as well as space-filler fat pads (Ralphs and Benjamin, 1994).

As described earlier, skeletogenesis starts with the formation of uninterrupted mesenchymal condensations. Mechanisms must therefore exist to dictate the precise location at which these condensations become interrupted and eventually form joints and interzones. Thus, the development of a functional synovial joint is generally divided into two phases. The first is the formation of cartilaginous anlagen and the intervening zones in which the joints will develop, and the second is the formation of the joint cavity described above. During joint cavitation, the peripheral interzone is absorbed into each adjacent cartilaginous zone, evolving into the articular surface; this surface is destined to become a specialized cartilaginous structure that does not normally undergo vascularisation and ossification (Pacifci et al., 2005), (Pitsillides and Ashhurst, 2008). Figure 5 demonstrates the major steps in digital synovial joint formation.

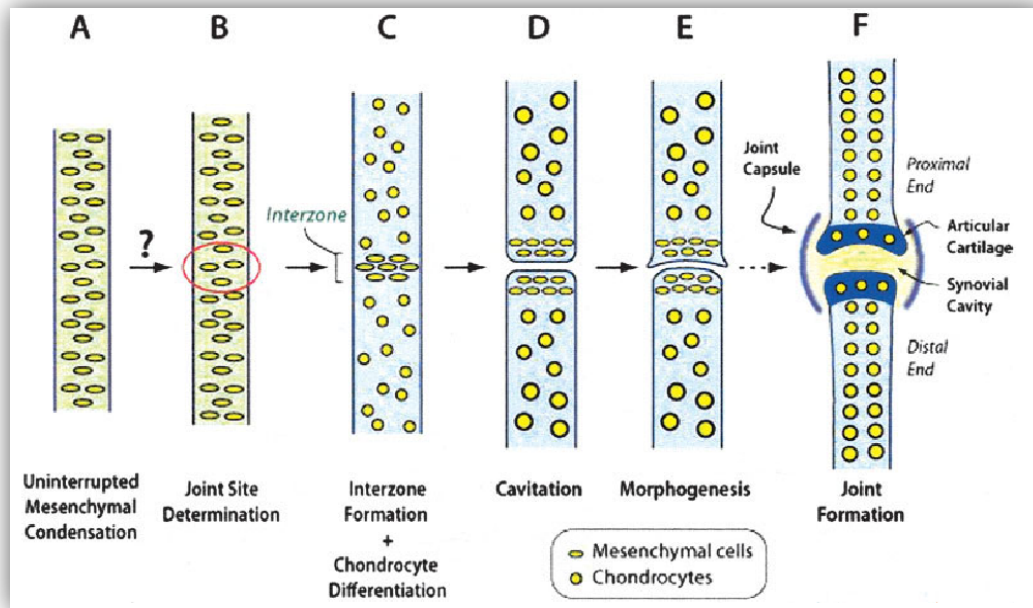


Figure 5: Schematic diagram depicting the major steps in digital synovial joint formation. A: Mesenchymal prechondrogenic condensations are appearing in the early limb bud, exhibit no overt signs of joint formation and are therefore uninterrupted. B: Unknown upstream patterning mechanisms determine the exact location for joint initiation. C: the interzone forming at each future joint site represents the first overt morphological sign of joint formation. There is further gathering and condensation of cells at this site. In the meantime, mesenchymal cells constituting the bulk of the condensations differentiate into chondrocytes and establish the cartilaginous skeleton. D: A cavitation process begins that eventually leads to the physical separation of the adjacent skeletal anlagen and formation of a synovial cavity. E: The opposing sides of the joint begin to mold into reciprocally shaped and interlocking structures. F: Eventually, all the distinct component of a mature joint form and include articular cartilage, capsule and a full-fledged synovial cavity. All these processes are continuous and are presented here as distinct events for descriptive purposes. Adapted from (Pacifici et al., 2005).

2.4 Postnatal development of bones- development of secondary ossification centres (SOCs)

A developing long bone has an epiphysis and metaphysis at each end and a diaphysis (shaft) in between. By the middle of the embryonic stage, these regions are established and go through proportional changes in size until skeletal maturity. The epiphyses are important structures that are responsible for the transverse and spherical growth of the ends of the bone, the shaping of the articular surfaces, and the longitudinal growth of the metaphyses and the

diaphysis. In addition to this, a small amount of longitudinal growth also occurs with interstitial expansion of the epiphyseal cartilage, including the undersurface of the articular cartilage. Each epiphysis, formed initially completely in cartilage, subsequently differentiates into three histologically distinct regions. These are: i) the cartilage at the outermost boundary of the epiphysis adjacent to the joint space that is the articular cartilage, ii) the cartilage adjacent to the metaphysis that forms the growth plate (physeal cartilage) and iii) the cartilage between the articular cartilage and the physeal cartilage referred to as the epiphyseal cartilage, which will form a secondary ossification center after vascular and osteoprogenitor cell invasion (Forriol and Shapiro, 2005), (Forriol and Shapiro, 2005).

The formation of the POC and SOC is quite similar in that they both involve the mineralization and degradation of cartilage, angiogenesis, cartilage resorption associated with apoptosis of habitant cells and finally *de novo* synthesis of bone matrix by osteoblasts. However, the development of the SOC differs in some fundamental aspects from that of the POC in that vascularization occurs prior to the formation of a bony collar, chondrocyte hypertrophy and cartilage matrix mineralization (Alvarez et al., 2005). Furthermore, a primary and critical event that precedes the formation of the SOC is the early generation of vascularized cartilage canals (Blumer et al., 2005).

The growth plate is a specialized structure comprising of columns of resting, proliferating and hypertrophic chondrocytes. The growth plate components go through a sequential process of cell proliferation, synthesis of extracellular matrix, cell hypertrophy, mineralization of the matrix, localized vascular invasion accompanied by osteoprogenitor cells, and apoptosis. These highly coordinated activities lead to longitudinal bone growth and bone formation at the physeal-metaphyseal region by proliferation of chondrocytes as well as the expansion of the hypertrophic region. The mechanisms that govern their differentiation seem to encompass

those controlling differentiation during embryonic development (endochondral ossification) (Pechak et al., 1986), (Olsen et al., 2000). Growth cartilage renews itself through the germinal zone and is constantly replaced with bone at the physeal- metaphyseal junction. Though these developmental mechanisms, the length of the entire bone increases, the epiphyses and metaphyses at either end are displaced progressively further away from the centre of the bone, and the growth plate itself maintains the same height throughout the growth period. Concurrently, there is radial growth of the diaphysis and parts of the metaphysis by intramembranous bone formation with direct apposition of cortical bone by osteoblasts from the inner cambial layer of the periosteum. This is coordinated closely with resorption of bone by osteoclasts on the inner cortical endosteal surfaces and lateral metaphyseal surfaces to maintain the relative proportions of the marrow cavity to the cortices and the overall shape of the bone as it grows. Figure 6 and Table 1 show the complete histologic stages in long bone and epiphyseal development from an early embryonic stage to adulthood.

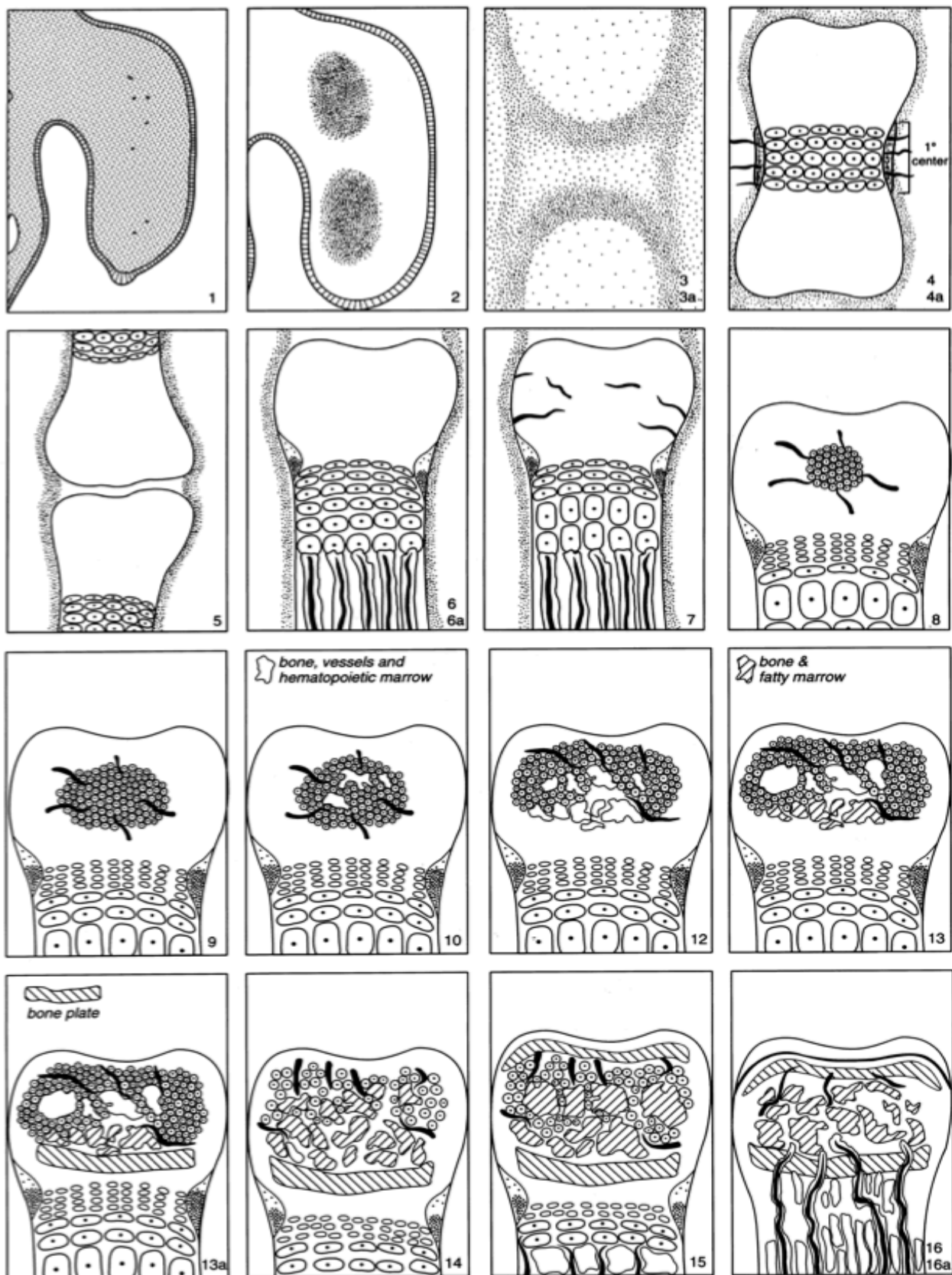


Figure 6 The major stages in the formation and development of long bones and epiphyses (as described in Table 1), from stage 1 to stage 16a. From (Rivas and Shapiro, 2002).

Table 1 Histologic stages in long bone and epiphyseal development. The substages labeled 'a' and 'b' refer to an event that is occurring at the same time but in a different part of the same bone or to a structurally important continuation of the same process at a slightly later time. From (Rivas and Shapiro, 2002)

Stage	Histologic events
1	Limb-bud formation, uniform distribution of mesenchymal cells, and formation of apical ectodermal ridge
2	Mesenchymal condensation
3a	Cartilage differentiation
3b	Interzone formation
3b	Chondrocyte hypertrophy in middle part of long-bone cartilage model
4	Epiphyseal shaping
4a	Formation of intramembranous periosteal bone at mid-diaphysis (primary center of ossification)
5	Resorption of joint interzone and formation of smooth articular cartilage surface
5a	Vascular invasion of hypertrophic chondrocyte area, endochondral bone formation (mid-diaphysis), and completion of formation of primary center of ossification
6	Formation of the physis and of peripheral perichondrial groove tissue
6a	Farthest relative extent of epiphyseal/physeal position
7	Vascularization of epiphyseal cartilage with formation of cartilage canals
8	Central chondrocyte hypertrophy to form spherical mass, development of growth plate completely surrounding secondary ossification center
9	Vascular invasion of developing secondary ossification center into hypertrophic chondrocytes adjacent to mineralized cartilage matrix
10	Bone formation and marrow cavitation in secondary ossification center, formation of hematopoietic marrow
11	Increase in size of secondary ossification center, decrease in size of epiphyseal cartilage
12	Central chondrocyte hypertrophy and secondary-ossification-center growth-plate change from spherical to hemi- spherical orientation
13	Fat in marrow, hematopoietic marrow adjacent to secondary-ossification-center growth plate
13a	Epiphyseal bone-plate formation
14	Fullest relative extent of secondary-ossification-center development in epiphyseal cartilage
15	Thinning of physis
15a	Involution of secondary-ossification-center growth plate
15b	Subchondral bone-plate formation
16	Resorption of physis with linkage of epiphyseal and metaphyseal circulations
16a	Calcification of lowest zone of articular cartilage, tidemark formation, and transformation of all marrow to fat

2.5 Cartilage

The embryology of cartilage can be viewed in the following way: The mesenchymal cells that aggregate to produce cartilage form a blastema at 5 weeks of gestational age in humans. They are called chondroblasts when they begin to secrete cartilage matrix. With increasing development, the ECM that is formed steadily pushes the cells apart. These cells, which are encased in this tough and specialized matrix, are then called chondrocytes. It is the mesenchymal tissue that surrounds the blastema that gives rise to the perichondrium (Bhosale and Richardson, 2008).

Cartilagenous tissues are present throughout the body and according to the composition of the ECM (based on the presence of collagen and elastic fibers) and on tissue morphology, are classified histologically into **hyaline cartilage** (Bhosale and Richardson, 2008), **elastic cartilage** (present in the outer ear, larynx and epiglottis), and finally **fibrocartilage** (found in the pubic symphysis, intervertebral discs, meniscus, TMJ and annulus fibrosus). Hyaline cartilage has unique and diverse characteristics. Like all cartilage, it is a specialized form of connective tissue that consists of cells and extracellular material. Chondrocytes are embedded in an ECM, and it is the matrix that endows cartilage with its unique characteristics. Hyaline cartilage is covered by a tough and fibrous sheath of dense, irregular connective tissue (the perichondrium). The perichondrium contains a graded series of cells caught at different stages of development. The cells at the periphery of the perichondrium are flat fibroblasts that, in time, differentiate to form the egg-shaped chondroblasts, which in turn develop into round, mature chondrocytes. The chondrocytes, which often contain large lipid droplets, are responsible for the synthesis, secretion, and maintenance of the ECM. Should the chondrocytes die or become damaged, the cartilage matrix degenerates and can undergo calcification. Although the perichondrium can act as a source of new chondrocytes to a limited extent, cartilage does not regenerate well unless the animal is quite young. The ECM of **elastic cartilage** is criss-

crossed by numerous branched elastic fibers rich in the protein elastin. These elastic fibers lend a high degree of flexibility and resilience to the matrix. When bent, elastic cartilage will rapidly bounce back to its original form. The outer margins of elastic cartilage, like hyaline cartilage, are defined by a fibrous perichondrium made of dense connective tissue. This perichondrium contains collagen fibrils, fibroblasts as well as chondroblasts. The mature chondrocytes occupy spaces in the matrix called lacunae. Hyaline and elastic cartilage appear very similar when viewed with a light microscope but can be differentiated with an electron microscope due to the detection of elastic fibers in the elastic cartilage. **Fibrocartilage**, on the other hand, is a unique subtype of cartilage in that it lacks a true perichondrium, and often, as in the intervertebral disc, forms a structural and functional junction where cartilage and ligament meet. Also, collagen type I is the prominent structural element in fibrocartilage as opposed to collagen type II in hyaline cartilage. The variation in total proteoglycan and collagen content and localization pattern of minor collagens in the matrix of different subtypes of cartilage is either due to a response to the different physiological requirements and mechanical stresses that the cartilage tissues have at different anatomic sites or as a consequence of developmental events (Naumann et al., 2002), www.visualhistology.com.

Nevertheless, **hyaline cartilage** is the predominant form of cartilage and is the type of cartilage that is most commonly associated with the skeletal system. It is *this* form of cartilage that generates the anlage for many bones in the embryo, the growth plate through which many bones increase their size during juvenile development and finally the bearing surface upon which bones articulate in movable joints (articular cartilage) (Roughley, 2006).

2.5.1 Articular Cartilage

2.5.1.1 Basic overview

In adults, articular cartilage is a relatively acellular tissue, with cell volume averaging approximately 2% of the total cartilage volume (Poole et al., 2001). Articular cartilage is a highly specialized connective tissue that covers the weight-bearing surfaces of diarthrodial joints. Cartilage layers cover bone ends and are principally responsible for permitting low-friction, high velocity movement between bones, absorbing the transmitted forces associated with locomotion, and contributing to joint stability. The chondrocytes of mature articular cartilage are terminally differentiated cells and are mainly responsible for synthesizing and maintaining the highly specialized cartilage matrix macromolecules (see Table 2). This cartilage ECM is composed of an extensive network of collagen fibrils (mainly collagen II), which confers tensile strength, as well as an interlocking mesh of proteoglycans (mainly aggrecan), which themselves provide compressive stiffness in joints through their ability to absorb and extrude water (Harris et al., 2004).

It is not clear whether articular cartilage can be defined as a distinct tissue in the juvenile, as it forms a continuum with the hyaline cartilage of the central epiphysis, which is destined to be resorbed (table 1 stage 16). Indeed, it is still somewhat of a mystery why articular cartilage survives at all and is not also replaced by bone.

2.5.1.2 Structural organization

The components of the articular cartilage ECM are highly ordered from the cartilage surface to the deepest layers. It is divided into four individual zones with separate functions: The superficial, middle or transitional, deep or radial and calcified zones. Unlike the deepest two zones, there is no sharp boundary between the first three zones (superficial, middle and deep).

The superficial zone is the thinnest of the four and consists of many fine collagen fibrils that are oriented in a tangential manner, low levels of proteoglycan as well as elongated chondrocytes that are aligned parallel to the surface (Venn, 1979). This zone is in contact with the synovial fluid of the joint and is accountable for most of the tensile properties of cartilage that enable it to withstand shearing as well as the tensile and compressive forces imposed on it by the movement of the articulation (Martel-Pelletier et al., 2008).

The middle or transitional zone is composed of a higher content of proteoglycans (to resist compression) and thicker but less collagen fibrils that are organized into radial bundles or layers. The chondrocytes in this zone, which represents approximately 40-60% of the total cartilage height, are at low density and have a round shape (Schinagl et al., 1997). In the deep zone, chondrocytes have the same shape as those in the middle zone. They are, however aligned perpendicular to the articular surface but with low density. Compared to the other zones, this zone contains the largest collagen fibrils in a radial disposition, which extend into the bone to reinforce tissue attachment. Aggrecan (proteoglycan) content is at its highest (Venn, 1979), (Martel-Pelletier et al., 2008), (Hu and Athanasiou, 2006).

The deepest calcified cartilage zone, which is divided from the other zones by the tidemark, mechanically and physically separates the hyaline cartilage from the subchondral bone. It mainly functions to anchor the cartilage to the bone since collagen fibrils from the deep zone penetrate into the calcified cartilage. Chondrocytes are very scarce and hypertrophic (Martel-Pelletier et al., 2008). The structure of articular cartilage is shown in Figure 7.

Throughout the four zones, the pericellular area that surrounds the chondrocytes is composed of a thin layer of non-fibrillar material. This most probably represents the products synthesized by the chondrocytes such as proteoglycans and glycoproteins. Immediately adjacent to the

pericellular region lies the territorial cartilage extra-cellular matrix. Here exists the dense meshwork of collagen fibers that form a capsule-like structure around the cells providing mechanical protection to the chondrocytes (Martel-Pelletier et al., 2008), (Guilak et al., 2006).

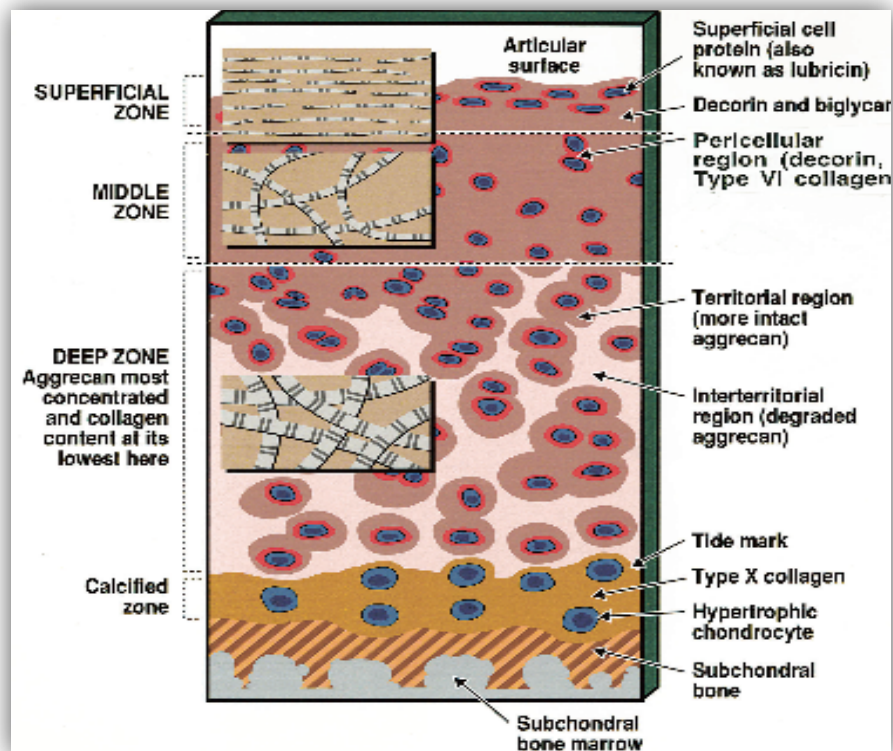


Figure 7: Representation of the general structure of adult human articular cartilage to show the zones, regions and relationship with subchondral bone. Particular important features of molecular content or properties are indicated. Also, shown are the relative diameters and organizations of collagen macrofibrils as well as large aggregating proteoglycans in the different zones. Adapted from (Poole et al., 2001).

2.5.1.3 Composition

Whereas most of the tissue wet weight of cartilage is composed of water and dissolved inorganic salts (including sodium, calcium and potassium chloride), the rest -approximately 20%- is occupied by collagen, proteoglycans, as well as other proteins or glycoproteins. The water content in cartilage is mostly found in the extracellular matrix and is unevenly distributed with its concentration at its highest (80%) close to the articular surface. However, with increasing depth, the concentration of water decreases and reaches approximately 65% of the

total wet weight in the deep zone. The water in the cartilage functions in maintaining the resilience of the tissue and contributes to the nutrition and lubrication system. More accurately, the diffusion of water through the cartilage aids the movement of nutrients from the synovial fluid through the tissue, contributing to the nutrition of the chondrocytes. The proper maintenance and flow of water in the tissue relies on its interaction with the macromolecules found there (Bhosale and Richardson, 2008).

Collagen represents approximately 50-60% of the cartilage dry weight. The collagen is organized in fibrils, which construct a network that dictates the shape and form of the tissue. The fibril diameters vary from 20nm in the superficial zone to 70-120 nm in the deep zone. The most abundantly found collagen is collagen type II, which is specific to articular cartilage and occupies 90-98% of the total tissue collagen. As mentioned earlier, these fibers give cartilage tensile stiffness and strength through their mechanical properties (Buckwalter and Mankin, 1998), (Martel-Pelletier et al., 2008).

Proteoglycans, on the other hand, constitute the second largest portion of the solid phase in articular cartilage occupying approximately 5-10% of the tissue wet weight. In articular cartilage, proteoglycans are complex supramolecular aggregates that consist of a central hyaluronic acid (HA) filament to which multiple monomers are non-covalently attached (Heinegard, 2009). In addition to HA (Holmes et al., 1988), other examples of articular cartilage specific proteoglycans include aggrecan (the most abundant proteoglycan), link protein (Neame and Barry, 1994), small leucine-rich repeat proteoglycans (SLRPs) such as decorin, biglycan, fibromodulin and lumican (Schaefer and Iozzo, 2008) and finally versican (Sztrolovics et al., 2002) and perlecan (Arikawa-Hirasawa et al., 1999).

In the extracellular matrix of cartilage, the proteoglycans are mainly of the large aggregating type (approximately 50-85%) and of the large nonaggregating type (approximately 10-40%). An aggregating proteoglycan consists of a monomer formed by the core protein, which acts as a protein backbone. From this core protein, numerous glycosaminoglycan side chains radiate at a 90° angle, extending stiffly. These side chains are composed of a long chain of polydimeric saccharide repeats including chondroitin-4-sulfate, chondroitin-6-sulfate and keratan sulfate, of which chondroitin sulfate is the most plentiful (approximately 60-90%). The average chondroitin sulfate chain has 25-30 disaccharide repeats while keratan sulfate chains only consist of around 5 or 6 repeating dimeric units (Heinegard, 2009), (Martel-Pelletier et al., 2008).

Table 2 shows a wide range of articular cartilage ECM components.

Table 2: Extracellular matrix components of cartilage. Abbreviations: HA = hyaluronic acid. HS = heparan sulphate. OA = osteoarthritis. PRELP = proline- and arginine-rich repeat protein. TGF- β = transforming growth factor- β (Harris, E et al. 2004).

Molecule	Function and location
Collagens	
Type II	Tensile strength, Major component of collagen fibrils
Type IX	Tensile properties, Interfibrillar connections; cross-links to surface of collagen fibril, NC4 domain projects into matrix
Type XI	Nucleation/control of fibril formation; within collagen fibril
Type VI	Forms microfibrillar network, binds hyaluronan, biglycan, decorin; pericellular
Type X	Support for endochondral ossification; hypertrophic zone and calcified cartilage
Type XII	Associated with collagen I fibrils in perichondrium and articular surface
Type XIV	Associated with collagen fibrils throughout articular cartilage
Proteoglycans	
Aggrecan	Compressive stiffness through hydration of fixed charge density: binding through G1 domain to HA stabilized by link protein
Versican	Low levels in articular cartilage throughout life; calcium –binding and selectin-like properties
Perlecan	Cell-matrix adhesion; pericellular
Biglycan	Binds collagen VI and TGF- β ; pericellular
Decorin	Controls size/shape of collagen fibrils, binds collagen II and TGF- β ; interterritorial
Fibromodulin	Same as decorin
Lumican	Same as decorin
PRELP	Mediates cell binding through HS sulfate in syndecan
Chondroadherin	Binding to cells via $\alpha 2\beta 1$ integrin
Other Molecules	
Hyaluronic acid	Retention of aggrecan within matrix
Link protein	Stabilizes attachment of aggrecan G1 domain to HA
Cartilage oligomeric	Interterritorial in articular cartilage; stabilizes collagen network or promotes collagen fibril assembly; calcium binding

matrix protein (COMP)	
Cartilage matrix protein (CMP)	Tightly bound to aggrecan in immature cartilage
Cartilage intermediate layer protein (CILP)	Restricted to middle/deep zones of cartilage, increase in early and late OA
Glycoprotein (gp)-39, or YKL-40	Marker of cartilage turnover; chondrocyte proliferation; superficial zone of cartilage
Fibronectin	Cell attachment and binding to collagen and proteoglycans; increase in OA cartilage
Tenascin-c	Binds syndecan-3 during chondrogenesis; angiogenesis
Superficial zone protein (SZP) or lubricin	Joint lubrication (superficial zone only)
Membrane Proteins	
CD44	Cell-matrix interactions; binds HA
Syndecan-3	Receptor for tenascin-c during cartilage development; cell-matrix interactions
Anchorin CII	Cell-surface attachment to type II collagen; calcium binding
Integrins ($\alpha 1, 2, 3, 5, 6, 10$; $\beta 1, 3, 5$)	Cell-matrix binding: $\alpha 1\beta 1$ /collagen 1 or VI, $\alpha 2\beta 1$ or $\alpha 3\beta 1$ / collagen II, $\alpha 5\beta 1$ /FN; intracellular signalling

3 Aggrecan

3.1 Overview

Aggrecan, the large aggregating chondroitin sulphate proteoglycan, is one of the major structural components of hyaline cartilage as it is the most abundantly established proteoglycan in the tissue. It is also present in significant concentrations in the extracellular matrices of other tissues such as meniscus, heart, brain and liver (Dudhia, 2005).

It does not exist in isolation within the extracellular matrix of cartilage. Instead, it is present in the form of proteoglycan aggregates (Hascall, 1988). Up to 100 individual monomers of aggrecan are non-covalently attached to a central filament of HA forming macromolecular aggregates of aggrecan (Figure 8) (Hardingham and Fosang, 1992). Each of these interactions is in turn stabilized by the presence of a link protein (Morgelin et al., 1988). The structure of the proteoglycan aggregate is therefore influenced by three parameters: the length of the HA, the proportion of link protein and finally the degree of aggrecan processing. To date, two molecular forms have been characterized for proteoglycan aggregates that have been derived directly from cartilage without the use of dissociative agents (Manicourt et al., 1991). Apart from the apparent differing content of link protein across these two forms, they may also have different functional characteristics (Buckwalter et al., 1994).

The aggregates of aggrecan (molecular mass of 10^8 - 10^9 Da) occupy a large hydrodynamic volume due to their high fixed negative charge resulting from the numerous attached polyanionic glycosaminoglycan chains (Horkay et al., 2008). This brings about a rise in osmotic pressure and subsequent swelling of the cartilage. This swelling pressure is counteracted by

the resistance of intact collagen fibres in cartilage. Collectively, these features enable joints to resist compressive force while maintaining a high tensile strength (Akizuki et al., 1987).

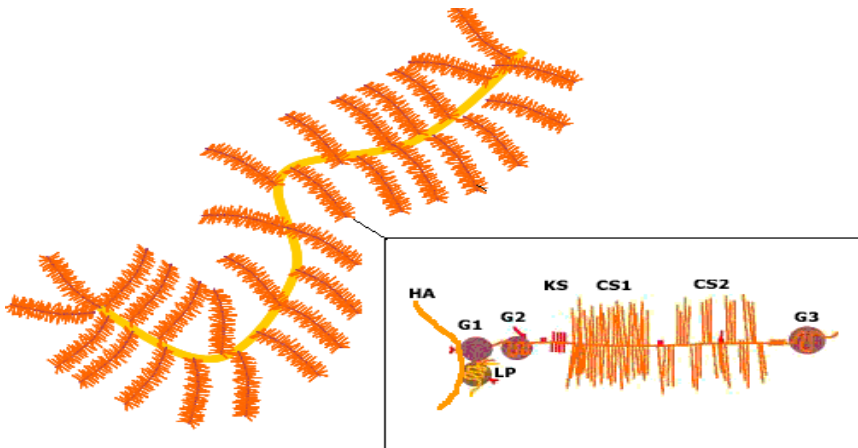


Figure 8: Schematic representation of an aggregate of aggrecan, with detail of monomer in boxed region. Globular domains (G1-G3) and attached GAGs are depicted. Aggrecan binds via G1 domain to hyaluronic acid (HA). This binding is stabilised by link protein (LP). Adapted from www.cmb.lu.se/ctb/photos/Aggrecan.gif.

3.2 Aggrecan protein

The aggrecan monomer consists of an extended multidomain protein core with a predicted size of approx. 220-250 kDa to which several chondroitin sulphate (CS) and keratan sulphate (KS) side chains are attached. There is a globular domain (G1) at the N-terminus of the protein core, which mediates specific interactions with the complex carbohydrate hyaluronan (HA) (Watanabe et al., 1997a) (Figure 8). These interactions are in turn stabilised by simultaneous binding to link protein (also bound with high affinity to HA). The G1 domain has a typical three-looped structure. The first of these is the N-terminal A loop which has amino acid sequence similarity to the immunoglobulin superfamily. The 2 additional loops (B and B') are named the proteoglycan tandem repeats (PTRs) and are homologous to each other. The G1 domain is separated from a second highly homologous domain (G2) by a linear interglobular domain (IGD). G2 is made up of a second pair of PTRs (loop B and B') that do not appear to interact

with HA (Fosang and Hardingham, 1989). Although the exact function of this domain is still unknown, it appears to function by inhibiting aggrecan secretion prior to GAG modification in the Golgi (Luo et al., 1996), (Doege et al., 1991), (Yasumoto et al., 2003).

The GAG-attachment region provides the high anionic charge density that is needed for the unique properties of aggrecan. It is composed of three domains that are responsible for attachment of keratan sulphate and chondroitin sulphate: The KS-attachment domain along with the neighbouring CS-attachment region, which is divided into two subdomains – the CS1 and CS2 domains. The KS-attachment domain resides adjacent to the G2 region and is composed mainly of repeat motifs that vary in number between species (Barry et al., 1994). Human aggrecan contains 30-60 keratan sulphate GAG chains per monomer, which are composed of shorter repeats of acetylgalactosamine and galactose. The 6- hydroxyl group of the N-acetyl-galactosamine is sulphated.

Like the KS-rich domain, the CS1 domain (which lies adjacent to it) is also composed largely of repeat motifs that vary in number across species. Furthermore, the human aggrecan CS1 domain displays size polymorphism across different individuals due to the flexible number of 19 amino acid tandem repeats, each of which contains two serine residues that can act as an attachment site for chondroitin sulphate (Doege et al., 1997). On average, there are 100-150 chondroitin sulphate glycosaminoglycan (GAG) chains per monomer of human aggrecan and each is made up of 40-50 repeating disaccharide units of N-acetylgalactosamine and glucuronic acid. In this case, the 4- and 6- hydroxyl groups of the N-acetyl-galactosamine are sulphated. Linkage to the protein occurs via a specific Ser-Xylose-Galactose-Galactose-Glucuronate sequence. Irrespective of the varying numbers of CS chains that the aggrecan molecules of different individuals are able to bear, the structure of the CS chains can change

throughout life due to changes in length and sulphation pattern (Roughley and White, 1980), (Silbert and Sugumaran, 2002), (Funderburgh, 2000), (Watanabe et al., 1998).

In addition to the KS and CS attachment sites, the GAG attachment region also possesses sites for the attachment of O-linked + N-linked oligosaccharides (Nilsson et al., 1982). With increasing age, these can become substituted with KS (Santer et al., 1982). Therefore, the location of KS may be expanded to the G1, the IGD and the G2 region where O-linked or N-linked oligosaccharide attachment sites can reside (Barry et al., 1995). Like the CS chains, the KS chains show age-related changes in structure (Brown et al., 1998).

At the C-terminus of the protein core, there are two alternatively spliced epidermal growth factor-like regions (EGF1 and EGF2), a lectin (LEC)-like G3 domain, and an alternatively spliced domain which shows sequence homology to complement regulatory proteins (CRPs). The G3 domain appears to contribute to the intracellular trafficking of aggrecan (post-translational processing of the core protein and subsequent aggrecan secretion) (Zheng et al., 1998). Very little is known about the functions of the EGF-like domains although it is likely that they have high affinity to Ca^{2+} ions indicated by their functions in proteins such as human protein S and human coagulation factors IX and X. Thus, it is possible that the EGF-like domains of aggrecan bind Ca^{2+} and interact with the Ca^{2+} -dependent LEC-like carbohydrate-binding domain (Baldwin et al., 1989), (Valhmu et al., 1995).

The LEC-like domain of aggrecan shows significant homology with several other known proteins such as the human leucocyte adhesion molecule 1 (LAM-1), the sea raven antifreeze protein and the human macrophage mannose receptor (Valhmu et al., 1995).

3.3 The Aggrecan Gene

3.3.1 Intron/exon organisation of the human aggrecan gene

The human aggrecan gene (AGC1) has been localized to chromosome 15q26.2 by *in situ* hybridization (Just et al., 1993), (Korenberg et al., 1993). The sequencing of genomic cosmids of the human aggrecan gene has revealed that the gene consists of 19 exons and 18 introns. The entire coding region (exons 2-19) is restricted to a 39.4kb segment of the overall ~70kb gene. The exons range in size from 77 to 4224bp (Table 3) (Valhmu et al., 1995). Although it's not clear whether this is of any functional significance, the G3 region can give rise to different aggrecan transcripts due to alternative splicing (Fulop et al., 1993).

The aggrecan gene has been characterised in several species including *R. norvegicus* (Doege et al., 1990), *H. sapiens* (Valhmu et al., 1995), *M. musculus* (Watanabe et al., 1995), and *G. domesticus* (Li and Schwartz, 1995). However, until the beginning of this work, little was understood about the transcriptional regulatory mechanisms that govern the expression (particularly the chondrocyte specific expression) of the gene in these different species.

Table 3 Some of the characteristics of the exons/introns of the human aggrecan gene. A protein domain is coded for by each exon or a combination of exons. Intron sizes are accurate within ± 0.1 kb. In addition to the 3'UTR, exon 19 also codes for the 25-amino acid C-terminus of aggrecan. Adapted from (Valhmu et al., 1995).

Protein domain	Exon		Intron	
	No	Size (bp)	No	Size (kb)
5'UTR	1	375	1	>13
SP	2	77	2	2.4
G1-A	3	384	3	1.0
G1-B	4	175	4	1.6
	5	128	5	1.5
G1-B'	6	294	6	1.9
IGD	7	378	7	1.5
G2-B	8	175	8	0.5
	9	128	9	1.4
G2-B'	10	294	10	2.1
KS-1	11	240	11	2.9
KS-2/CS	12	4224	12	0.9
EGF-1	13	114	13	8.4
EGF-2	14	114	14	3.2
LEC	15	158	15	0.5
	16	83	16	0.8
	17	145	17	0.8
CRP	18	183	18	0.5
3'UTR	19	212	-	-

3.3.2 The expression of the aggrecan gene

The aggrecan gene is mainly expressed in cartilaginous tissue. Unlike *Col2a1*, aggrecan is not expressed in skeletogenic mesenchymal cells, the perichondrium or presumptive joint cells (Shepard et al., 2007) and is mainly found in differentiated chondrocytes such as the chondrocytes of the growth plate and articular cartilage embryonically and throughout adulthood. In the knee, in addition to its expression in the hyaline cartilage of the growth plate and the articular surfaces, aggrecan is expressed in the fibrocartilage of the menisci. A study performed on ovine cartilage showed that in 2-7-day old menisci, aggrecan is localized in the inner meniscal zone, which displays a hyaline cartilage-like character at this stage in meniscal development. At later age points (> 7 months), aggrecan was also shown to be widely distributed throughout the other meniscal zones and did not decline markedly with age (Melrose et al., 2005).

It is important to note that although the expression of aggrecan is mainly found in differentiated chondrocytes, it has been detected before the appearance of cartilage matrix. In E14.5 mice, although no *Col2a1* mRNA was observed, aggrecan mRNA was detected in the posterior region of the mesenchymal cell aggregation continuous with the ossifying mandibular bone anlage prior to chondrogenesis (Fukada et al., 1999).

Even though high levels of aggrecan protein expression are found predominantly in cartilage, its mRNA has also been observed in the:

- 1- Central nervous system (CNS) in which aggrecan glycoforms are major constituents of perineuronal nets in the brain of the adult rat. Matthews et al (2002) showed by *in situ* hybridization that aggrecan is found in the cerebral cortex, the deep cerebellar nuclei, the colliculi, the spinal core, the dentate gyrus as well as the CA1, CA2, and CA3 fields of the hippocampus. Also, they showed that neurons are the predominant cellular

source of aggrecan in the adult CNS. Finally, using *in situ* hybridization, RNase protection analysis, and antibody staining, they demonstrated that aggrecan mRNA is not expressed in the CNS until late in embryonic development, peaks in expression at 1 month of age, and is maintained at relatively high levels into adulthood (Matthews et al., 2002). Another group showed more recently that aggrecan regulates astrocyte development and is expressed by embryonic glial precursors (Domowicz et al., 2008).

- 2- Notochord, which has its own extracellular matrix composed of several proteins including proteoglycans such as aggrecan (Sandell, 1994). The notochord has major roles in vertebral column formation by inducing sclerotome differentiation and by forming the nucleus pulposus of intervertebral discs. Within intervertebral discs, aggrecan is expressed in the nucleus pulposus as well as the inner annulus fibrosus (that is transitional in its composition between the nucleus pulposus and the outer annulus fibrosus) (Le Maitre et al., 2007).
- 3- Liver, where aggrecan mRNA has been shown by northern analysis and *in situ* hybridization to be expressed differentially during liver fibrogenesis (Krull and Gressner, 1992).
- 4- Heart, where aggrecan is specifically colocalized with several groups of migrating cells that include the endocardial cushion tissue cells, the epicardial cells, a mesenchymal cell population in the outflow tract that may be of neural crest origin, and a mesenchymal cell population in the inflow tract. This was shown by western blotting, immunohistochemical studies as well as RT-PCR of chicken embryo hearts at stages 12-21 of development (Zanin et al., 1999).
- 5- Carotid artery and the aorta. Using a proteomic-based approach, the extracellular proteoglycan core protein composition of intimal hyperplasia from both the internal carotid artery and the internal thoracic artery was shown to include aggrecan (Talusán et al., 2005). Also, by immunohistochemical studies, it was shown that aggrecan is

present at low levels in the cells arranged into lamellar and interlamellar cell layers within the embryonic chick aortic wall during development (Arciniegas et al., 2004).

- 6- Adipose tissue. Voros et al showed that aggrecan mRNA is present in murine intra-abdominal and inguinal subcutaneous fat pads. They suggested that aggrecan might contribute to the regulation of lipid uptake and obesity in mice (Voros et al., 2006).

The aggrecan chondroitin sulphate proteoglycans (CSPG) that are expressed in different tissues such as cartilage, notochord and brain are all products of the same aggrecan gene. However, in each tissue the CSPG product exhibits significantly different chemical, structural, and immunological properties, indicating that although each is initiated on an identical core protein, they are processed differently, yielding distinct proteoglycans (Schwartz and Domowicz, 2004).

3.3.3 The regulation of the aggrecan gene

The basal promoter of the aggrecan gene has been defined by mapping the TSS in different species including *H. sapiens* (Valhmu et al., 1995) and *R. norvegicus* (Doege et al., 1994) and was shown to be active in different cell lines in the latter species. In humans, this preliminary promoter was shown to lack canonical TATA and CAAT boxes and contain several putative binding sites for ubiquitous transcription factors such as SP-1, AP-2, and AP-4 that are located in both the promoter and untranslated exon1. They may possibly mediate the basal activities of this promoter. The following are *cis* and *trans*-regulatory elements that have been identified by *in vitro* studies and have been shown to regulate the expression of aggrecan by directly driving/repressing promoter activity:

1- Exon sequences (*cis*-regulation): Using luciferase reporter constructs driven by the human aggrecan promoter (taken as -701bp - +52bp relative to the human aggrecan TSS, 87,147,670 on chromosome 15, Assembly: March. 2006 NCBI36/hg18), the 5' and 3' untranslated regions of the human aggrecan gene were found to regulate gene expression transcriptionally in a cell-type specific manner. The 5' untranslated region (+1 - +375bp) is also the 1st untranslated exon of the human aggrecan gene. It stimulated the activities of the human aggrecan promoter by about 8 fold in chondrocytes and 40 fold in NIH 3T3 cells. In contrast, the 3' untranslated region inhibited the activities of the human aggrecan promoter by about 40-70% in both cell types (Valhmu et al., 1998a).

Sequential deletions of a large portion of this so-called human aggrecan proximal promoter from -701bp - -52bp had very little effect on promoter activity when the 5'UTR was present. However, when the 5'UTR was removed, leaving a 5' flanking sequence of -701bp - +25bp or -52bp - +25bp, luciferase activity was reduced 85-100%. Thus, it seems that the 5'UTR/exon 1 of the human aggrecan gene can act as the minimal promoter required to drive basal expression of the aggrecan gene. It is possible that sequences within this region bind to general transcription factors required for the assembly of the transcription initiation complex. Also, as mentioned above, this region contains several putative SP-1 binding sites characteristic of most TATA-less and CAAT-less promoters, and may be involved in recruiting the TFIID protein to the transcription initiation complex (Valhmu et al., 1998a).

2- Silencing region (*cis*-regulation): A 1.8kb genomic fragment, containing 5' non-coding chick aggrecan DNA as well as exon 1, was able to drive the expression of the pGL2-Basic reporter gene in a tissue-specific manner. Transient transfections of chick sternal chondrocytes and fibroblasts with pGL2-Basic plasmids bearing progressively reduced portions of this 1.8kb 'promoter fragment' allowed mapping of chondrocyte-specific transcription enhancer and

silencer elements that are consistent with sequence analysis data. Positions -873 and -721, relative to the chick aggrecan TSS and in this 1.8kb fragment, are at the 5' ends of two copies of the sequence CACCTCC (CIIS2) that has been suggested to be a silencer motif in the *COL2A1* promoter (Savagner et al., 1990). Deletion of these two motifs reduced the cell type specificity of luciferase expression while the overall promoter activities increased (Pirok et al., 1997).

3- Sox9 (*trans*-regulation): The mouse aggrecan promoter is responsive to the *sox9* transcription factor in the cartilage derived cell line, TC6. Sox9 overexpression in TC6 cells enhanced the transcriptional activity of the AgCAT-8 construct (containing 8kb of promoter/ 1st exon/intron fragments of the aggrecan gene) by up to ~3 fold. This enhancement was lost when a 4.5kb fragment from the 3'end of the 8kb construct (corresponding to intron 1) was deleted (Sekiya et al., 2000).

Although Sekiya et al showed that a segment of *Agc1* was responsive to Sox9, they did not show whether this gene segment binds Sox9 and is sufficient to drive cartilage expression *in vivo*. Han and Lefebvre (2008), however, showed a highly conserved enhancer within *Agc1* that is a direct target of the Sox trio (Sox9, 5 and 6). The binding of Sox9 to this enhancer occurs only in the presence of L-Sox5/Sox6. A similar mechanism applies to other cartilage enhancers (Han and Lefebvre, 2008). Furthermore, the ChIP hybridization experiment performed by Oh et al (2010) showed that rat aggrecan contained two separate Sox9 interaction sites which possibly correspond to the findings of the two studies mentioned above (Oh et al.).

Other than the regulatory elements/factor described above, there are regulators that are associated with altered expression of aggrecan that may or may not act directly on the

promoter. Examples are:

- 1- Biomechanical stimuli:** The levels of aggrecan mRNA are affected by short and long term load-controlled compression. There is a transient upregulation of aggrecan mRNA synthesis in bovine articular cartilage explants when a compressive stress of 0.1 MPa is applied for 1,4,12 and 24 hours. For example, at 1 hour, aggrecan mRNA levels in loaded explants were increased 3.2 fold compared to control explants (Valhmu et al., 1998b). Also, more recently, aggrecan expression in the annulus fibrosus cells of the human intervertebral disc was shown to be significantly increased by mechanical load (Neidlinger-Wilke et al., 2009).
- 2- Hyperosmotic loading:** Osmotic loading, which is the process of incubating cells with solutions of different osmolarity for various time points, influenced the expression of the aggrecan gene (driven by the sequence -2368bp to +375bp relative to the TSS) particularly through the 1st exon/5'UTR. More specifically, aggrecan promoter activity and mRNA levels were suppressed approximately two-fold when bovine monolayer chondrocytes were subjected to 24 hours of hyperosmotic loading. There was no suppression, however, with the deletion of the 1st exon/5'UTR of the gene in the reporter construct. This osmotic regulation was time-dependent and found to correlate with cell shrinking and swelling (Palmer et al., 2001).
- 3- Hif1 α and Hif2 α :** The nucleus pulposus is an aggrecan-rich, avascular tissue that permits the intervertebral disk to resist compressive loads. Owing to this, the relationship between HIF-1 α (Hypoxia-inducible-factor 1 α - a transcription factor that responds to oxygen tension and regulates glycolysis) expression and aggrecan promoter activity was measured. It was shown that when HIF-1 α was suppressed, irrespective of the oxemic state, there was a partial loss of aggrecan expression and

promoter activity. Moreover, when cells were treated with 2-DOG, a glycolytic inhibitor, there was inhibition in aggrecan promoter activity. Results of this study indicated that oxygen-independent stabilization of HIF-1 α in nucleus pulposus cells is a metabolic adaptation that drives glycolysis and aggrecan expression (Agrawal et al., 2007). Furthermore, it has been shown that human chondrocytes from knee articular cartilage enhance the production of aggrecan in response to hypoxia specifically through HIF-2 α (Hypoxia-inducible-factor 2 α) (Thoms and Murphy). At least in articular chondrocytes, aggrecan is not a direct HIF-2 α target, but is upregulated by hypoxia through the cartilage-specific transcription factor SOX9, which itself is upregulated in response to HIF-2 α (Murphy et al., 2009).

4- IL-1: The cytokine interleukin 1 (IL-1), which is considered to be the principal inducer of mediators of cartilage degradation in osteoarthritis (see Introduction section 4), can activate numerous signalling pathways involved in cartilage destruction and dedifferentiation of chondrocytes. With respect to aggrecan, IL-1 was shown to induce the reduction in its expression. IL-6 has also been found to act synergistically in the IL-1 induced suppression of aggrecan in chondrocytes. It seems that IL-1 induced IL-6 secretion by chondrocytes can eventually downregulate the expression of aggrecan through the p38MAPK and/or the PI3K/JNK signalling pathways (Radons et al., 2006).

5- TGF- β 1: TGF- β 1 has been observed to have profound effects on the initiation of mesenchymal cell condensation, which leads to chondrogenesis, and eventual production of ECM components. James et al (2009) showed that addition of TGF- β 1 to chondrogenic medium of postero-frontal cranial suture-derived mesenchymal cells *in vitro* resulted in significant induction of aggrecan by 24 fold over baseline expression. The expression of *SOX9* was increased by 30 fold as well. They did not, however,

confirm whether the regulation of aggrecan gene expression occurred indirectly through the upregulation of *SOX9* or by some other mechanism (James et al., 2009).

6- MicroRNA-140 (miR-140): MicroRNA-140 has been shown to be a marker for chondrogenesis. In a study performed by Miyaki et al (2009), pellets of MSCs, which were strongly stained by Safranin O after chondrogenesis induction for 14 days, revealed a gradual increase in miR-140 expression in parallel with expression of *sox9*, aggrecan, and *col2a1*. This indicates that the expression of miR-140 increases during chondrocytic differentiation of MSCs, and this is consistent with its high expression in chondrocytes. Furthermore, the group performed experiments that potentially reveal that miR-140 may regulate the expression of aggrecan. Chondrocytes were stimulated with IL-1 β with and without transfection of double stranded (ds)-miR-140. Conversely, aggrecan expression with IL-1 β stimulation was significantly increased by ds-miR-140. In the absence of IL-1 β , the nonspecific dsRNA (dsNeg) did not change the basal levels of aggrecan. These results demonstrated that miR-140 regulates aggrecan. However, the mechanism of this regulation is unknown. It is possible that IL-1 β negatively regulates the expression of aggrecan through the downregulation of the positive aggrecan regulator miR-140 since *in vitro* treatment of chondrocytes with IL-1 β suppresses miR-140 expression (Miyaki et al., 2009).

Despite all the preliminary data on the various *cis*-regulatory factors that affect the activities of the aggrecan promoter, regulatory elements directing specific expression in cartilage has yet to be described *in vivo*. The first *in vivo* study to search for these regulatory elements involved the rat aggrecan gene. Initially, this search had focused primarily on the proximal promoter, a 922-bp DNA fragment with 640-bp 5'-flanking DNA and 282-bp exon 1 sequence. This sequence had showed higher promoter activity in transfected chondrocytes than in fibroblasts (Doege et

al., 1994). However, this was not reproducible in subsequent experiments which indicated that there may be additional regulatory sequences other than the proximal promoter that are responsible for expression of the rat aggrecan gene in chondrocytes. DNA fragments upstream of the TSS and throughout the first intron (30kb) along with flanking sequences at the 3' end of the gene were analysed in cell transfection assays as well as transgenic mice. The results of the cell transfection experiments showed an upstream ~5kb tissue specific enhancer/repressor at about -12kb – -7kb relative to the TSS (called P3) and a far downstream enhancer without tissue specificity. When fused to an elongated proximal promoter sequence ranging from -3400 - +280bp (called P0), this upstream 'enhancer' also displayed its functionality *in vivo*. However, although both elements were active in cell culture experiments, the downstream 'enhancer' was not effective *in vivo*. The function of the downstream enhancer thus remains unresolved (Doege et al., 2002).

The combination of the upstream sequence P3 fused to P0 seems to direct expression of aggrecan in the developing cartilages of the long bones of the fore and hind limbs, the ribs, the scapula and some craniofacial cartilages. However, it is not completely sufficient for all aspects of aggrecan regulation; there was no expression found in vertebral bodies at embryonic day 14.5 of development, nor in the cartilage of the digits. Also, the expression was present only in hypertrophic chondrocytes. It is possible that other controlling sequences may remain outside these two described elements and/or silencer sequences may be present within them (Doege et al., 2002).

The second and more recent *in vivo* study, which was published during the course of my work, and that was performed in search of aggrecan *cis*-regulatory elements involved the mouse aggrecan gene. A highly conserved and regulatory 359-bp sequence located approximately 10Kb upstream of the mouse *Agc1* TSS (corresponding to a sequence present within the rat

P3 sequence) was identified. When multimerised (x4), this sequence was able to direct expression of the *Col2a1* minimal promoter (-309 bp - +308 bp relative to the TSS) in both embryonic and adult cartilage in transgenic mice, in a manner that matches aggrecan expression. When tested in transient transfection, this sequence was almost as active in RCS (rat chondrosarcoma) cells as a similar construct containing 4 copies of the *Col2a1* enhancer fused to the *Col2a1* minimal promoter (Han and Lefebvre, 2008).

From these experiments, it has been concluded that aggrecan expression appears to be at least partially regulated by a combination of proximal promoter sequences and distant elements at the 5' end of the promoter.

3.4 The proteolysis of aggrecan

The core aggrecan protein is subject to extracellular proteolytic processing and is rarely present in an intact form within the proteoglycan aggregates of the cartilage matrix. This leads to an accumulation of fragments that bear the G1 regions coupled with a loss of those that do not by diffusion from the tissue. As a result of this, proteoglycan aggregates become enriched with aggrecan G1 regions as opposed to the more intact molecules. These G1 regions can accumulate in the cartilage ECM with age, thereby competing for binding sites on HA with newly synthesised proteoglycan during turnover and remodelling of the cartilage ECM (Maroudas et al., 1998).

Alterations in cartilage with normal aging are independent of the changes that take place in degenerative joint disease or injury, in which pathology-related changes in the cartilage ECM and cells can accumulate over 2-3 decades. A prominent feature of joint disease is the increased proteolytic cleavage of aggrecan, which considerably reduces the levels of the

protein. The proteolysis of aggrecan is mediated by aggrecanases and matrix metalloproteinases (MMPs) (Sztrolovics et al., 1997). It is now well recognized that the principal proteinases responsible for aggrecan degradation *in situ* in articular cartilage are the aggrecanases (Sandy, 2006). Two aggrecanase isoforms have been identified recently and are members of the 'A Disintegrin and Metalloproteinase with Thrombospondin motifs' (ADAMTS) gene family – ADAMTS-4 and ADAMTS-5. The five aggrecanase cleavage sites that have been identified in aggrecan include one residing in the IGD and four in the CS2 domain (Tortorella et al., 2000). Within the IGD, in addition to the known aggrecanase cleavage site at TEGE³⁷³ ↓ ³⁷⁴ARGS and the MMP cleavage site at DIPEN³⁴¹ ↓ ³⁴²FFGV, a new cleavage site VQTV³⁵⁶ ↓ ³⁵⁷TWPD (located between the aggrecanase and MMP sites) has been identified recently and is cleaved by the protease HtrA1 (Chamberland et al., 2009).

The IGD has been the focus of intense research due to its susceptibility to proteolytic cleavage in experimental systems that study proteoglycan catabolism and cartilage degenerative disease. Altered glycosylation patterns within this domain during aging or in degenerative changes may further influence this susceptibility (Pratta et al., 2000).

The remodelling and turnover of articular cartilage components such as aggrecan has to be carefully regulated so as to maintain the biomechanical properties of the tissue. Normal cartilage function depends on a high aggrecan content, high GAG substitution and large aggregate size. The loss of cartilage integrity in arthritis is associated with the impaired function of the aggrecan core protein due either to its proteolytic cleavage (decreases the overall negative charge) or the proteolytic cleavage of HA (decreases the size of the aggregates). Osteoarthritis is a common condition of cartilage failure in which the careful regulation to maintain normal cartilage function has been disrupted.

3.5 The aggrecan knockout mouse

The first example of a mutated proteoglycan gene identified in mammals is the *cmd* aggrecan gene, which gives rise to cartilage matrix deficiency (*cmd*) in mice. The *cmd* mouse is thus a natural knockout mouse model of the aggrecan gene. The *cmd* aggrecan gene has a single 7-bp deletion in exon 5, which encodes the B loop of the G1 domain. This deletion gives rise to a frame shift resulting in the creation of a termination codon in exon 6. The potentially truncated 32-kDa polypeptide that is created by this nonsense mutation contains an incomplete proteoglycan tandem repeat structure that is incapable of HA-binding. However, biochemical and immunological studies demonstrate a complete absence of aggrecan in the cartilage matrix of *cmd* mice. Homozygotes *cmd* (-/-) are characterized by dwarfism, a short snout, and a cleft palate and die shortly after birth due to respiratory failure. Unlike the cartilage of wild-type mice in which extensive matrix can be seen, the cartilage of homozygous mice appears as tightly packed chondrocytes with little matrix. Heterozygotes *cmd* (+/-), on the other hand, are born normal (Watanabe and Yamada, 2002).

Although normal at birth, *cmd* heterozygotes exhibit two abnormal phenotypes as some differences in the metabolism of aggrecan may occur and create abnormal phenotypic changes after birth (due to the presence of only one normal allele of the gene). These abnormal phenotypes include slight dwarfism as well as late onset spinal misalignment. These heterozygotes usually die at 12-15 months of age compared to the average age of 2-2.5 years for wild-type mice. They die because they are unable to eat and starve to death within a month following acquisition of gait disturbance and decreased movement, which the mouse acquires due to its spinal distortions. Histology reveals that the degenerative changes that occur in heterozygotes are present within the cervico-thoracic spine where the vertebral bodies are deformed with the disappearance of the apophysis, and the herniated intervertebral disc likely

causes the spastic gait. Despite this, the knee joint and other cartilaginous tissues are apparently normal (Watanabe and Yamada, 2002).

The mRNA levels of aggrecan are reduced in *cmd*-mutant mice. RT-PCR studies reveal that these levels are reduced by 59% in homozygotes and 19% in heterozygotes compared to wild-type mice. Because of the relatively rapid turnover of aggrecan within cartilage, decreased synthesis may cause significant reduction of its deposition in tissue, which eventually leads to abnormal phenotypes in the heterozygotes (Watanabe and Yamada, 2002).

4 Osteoarthritis

4.1 Overview

The frequency and continual development of osteoarthritis (OA) represents a widespread cause of physical morbidity and impaired quality of life throughout the industrialized world. Short of surgical joint replacement, treatment of OA is limited to taking care of patients symptomatically and supportively.

Being the most common joint disorder, OA affects 10-25% of all people aged over 65 (data taken from <http://www.arc.org.uk>). It is a degenerative joint disease resulting from a combination of risk factors that contribute to the pathogenesis of the disease. These include both mechanical stresses - such as obesity, misalignment, joint trauma and surgery - as well as biomechanical abnormalities - such as genetic predisposition and other metabolic disorders. Age and gender are also significant risk factors. Eventually, these risk factors lead to injury in all of the different components of the joint but primarily involving cartilage, bone and synovium (Figure 9). The effects on these three joint compartments manifest as articular cartilage breakdown, osteophyte formation, subchondral sclerosis, and alterations of the synovium on both morphologic and biochemical levels (Krasnokutsky et al., 2007).

Cartilage has received the most attention in the study of OA due to articular chondrocytes acting as potential instigators of disease. These cells are responsible for the anabolic-catabolic balance required for matrix maintenance and tissue function due to their ability to synthesise the structural components of the extra-cellular matrix (ECM) along with matrix-degrading proteases (Aigner et al., 2007).

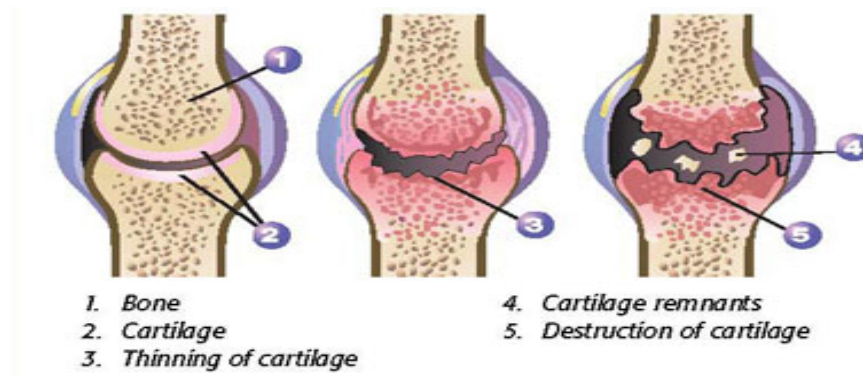


Figure 9: Diagram illustrating the evolution of osteoarthritis over time. Left: joint is normal. Middle: initial stages of disease in which you start to see thinning cartilage as indicated by the arrow. Right: a joint in which the disease has progressed so far, the cartilage has been destroyed leaving behind remnants as indicated by the arrows. Adapted from www.cvtechnologies.com/images/osteoarthritis.jpg

4.2 The development of osteoarthritis

During OA, osteoarthritic chondrocytes are exposed to abnormal extracellular stimuli. This is due to several factors: firstly, the ECM surrounding the cells is damaged by mechanical stress or enzymatic degradation through synovial proteases, making it more likely to exhibit altered biomechanical properties. These alterations include the production of matrix fragments such as fragments of fibronectin, collagen, hylauronic acid and other cartilage matrix proteins, all of which may be important contributors to disease progression. These fragments can stimulate chondrocytes to produce inflammatory mediators and proteolytic enzymes that are present in OA cartilage (Homandberg, 1999), (Knudson et al., 2000), (Tchetina et al., 2007)

Secondly, there are differences in the concentrations of growth factors, cytokines, chemokines and other inflammatory mediators diffusing into the cartilage from the synovial joint space. These factors are potent regulators of the chondrocyte phenotype and their concentrations are also modified in diseased cartilage by their increased expression in chondrocytes themselves as mentioned above. The inflammatory mediators produced by chondrocytes include: cytokines and chemokines such as IL-1, IL-6, IL-7, IL-8, IL-17, IL-18, MCP-1, LIF, GRO and

oncostatin M; reactive oxygen species such as nitric oxide, superoxide, hydrogen peroxide and peroxynitrate; and lipid-derived inflammatory mediators such as prostaglandins and leukotrienes (Goldring and Goldring, 2007), (Middleton and Tyler, 1992), (van der Kraan et al., 2002), (Middleton et al., 1996). These mediators can work in an autocrine and paracrine manner to alter gene expression in chondrocytes. More specifically, they activate and even upregulate the expression of several matrix-degrading proteases, for example the matrix metalloproteinases (MMPs), which have an important role in the breakdown of collagenous and noncollagenous matrix components (Aigner et al., 2007). Type II collagen is cleaved mainly by collagenase 3 (MMP-13) and is further degraded by gelatinases (MMP-2 and MMP-9) (Billinghurst et al., 1997). As previously mentioned, it is thought that aggrecan is degraded mainly by the enzyme ADAMTS-5 (Sandy, 2006).

As well as stimulating the production of catabolic factors in chondrocytes, the inflammatory mediators produced by these cells also contribute to cartilage loss through the inhibition of matrix synthesis. As the disease develops, chondrocytes typically respond to the damage and loss of ECM in several ways. The first is by proliferating, which results in the formation of chondrocyte clusters or 'clones' and the second is by attempting to produce new matrix. This new matrix, however, includes matrix proteins more commonly found during development (such as type IIA procollagen as opposed to Type II collagen). Regardless of the novel matrix being produced, there is still an imbalance between synthesis and degradation, which results in a net loss of matrix. Osteoarthritic chondrocytes in the articular cartilage undergo a phenotypic switch to chondrocytes present in the hypertrophic zone of the growth plate (which produce type X collagen and collagenase MMP-13) and this may at least be partly responsible for the changes seen in OA (Loeser, 2008).

So, not only do chondrocytes have a direct role in this degradation process but they also fail to compensate for matrix damage induced externally. Normally, when a tissue is damaged or under stress, the surrounding cells sense a change in the matrix and react with an appropriate repair response. One of the significant repair responses is the removal of the damaged matrix proteins, which may explain the initial production of proteolytic enzymes by chondrocytes as OA progresses. As soon as the damaged proteins are cleaned up, the proteolytic process needs to be switched off so that the cells can synthesize new matrix components. In order to shut down this catabolic process and stimulate the production of new matrix, a host of growth factors (such as BMP-2, BMP-7, CDMPs, IGF-1 and TGF- β) that are produced by chondrocytes and stored in cartilage bound to matrix proteins are released from the matrix as it is degraded. In the OA joint, however, this is not the case and it seems that this anabolic phase of matrix remodelling is insufficient or defective (Loeser, 2008).

As mentioned above, although alterations in cartilage with normal aging are independent of the changes that take place in OA, aging may be an important factor that contributes to the imbalance in cartilage remodelling. With increasing age, cells are less responsive to growth factor stimulation (Loeser and Shakoor, 2003). This is most probably the cause of the insufficient growth factor activity in aged cartilage. As a result of this poor activity, cartilage would continue to degrade under mechanical forces due to a continued cycle of unchecked matrix degradation.

4.3 Understanding Osteoarthritis pathogenesis and testing therapeutic ideas

There is a crucial need for animal models of osteoarthritis that can be used as research tools to uncover potential molecular markers for the onset of the disease and understand the role of gender and environmental factors in regulating the occurrence of the disease. There are now several lines of transgenic mice that acquire premature osteoarthritis. These have mutations or deletions of genes encoding several matrix and associated proteins, including type II, IX and XI collagen, membrane type 1 matrix metalloproteinase (MT-1-MMP), the $\alpha 1$ integrin subunit as well as ADAMTS5 (Young, 2005). Taken together, mouse models that either resist osteoarthritis or prematurely acquire osteoarthritis can be useful research tools to understand the molecular basis of osteoarthritis and for designing new treatments for the disease. Unfortunately, within these mouse models, embryonic lethality is often encountered due to the essential role of various genes in early embryonic development. Moreover, if the animal does survive, the outcomes of the gene deletions are difficult to interpret because of the severe skeletal malformations that may occur during development. An inducible and tissue specific (cartilage) gene deletion/overexpression approach would therefore be a better method to determine the function of a gene, its involvement in the progression of OA coupled with any potential therapeutic value it may have. This would increase the versatility and power of the murine system.

As mentioned earlier, aggrecan is a chondrocyte-specific protein and its expression is detected in growth plate and articular chondrocytes of long bones and other cartilage tissues in the body (Murphy et al., 1999). Unlike collagen type II, aggrecan is expressed continuously in cartilage (Chambers et al., 2002). To generate a chondrocyte-specific transgenic mouse model, the enzyme Cre recombinase can be fused to a mutated ligand-binding domain of the human

estrogen receptor (ER) and driven by the aggrecan promoter. This fusion protein will be expressed in chondrocytes and translocated into the nuclei in response to the estrogen antagonist tamoxifen or 4-hydroxy tamoxifen (Feil et al., 1996). Therefore, there is a need to increase our understanding of the transcription mechanisms that govern the expression of aggrecan.

The advent of the Cre-*loxP* recombination system has enabled the conditional inactivation of genes in the adult mouse. Bacteriophage P1 encodes the 38kDa Cre recombinase protein that catalyzes site-specific DNA recombination between 34 bp repeats named *loxP*. In Cre-mediated recombination, the resultant DNA structures depend on the orientation of these *loxP* sites. Two directly orientated repeats of *loxP* lead to the excision of intervening sequences leaving behind a single site within the genome whereas inverted repeats lead to the inversion of the involved sequences (Lakso et al., 1992), (Orban et al., 1992).

The Cre-*loxP* system in mice relies on two key components - the conditional allele and the Cre driver. To produce a transgenic mouse with the conditional allele, a targeted mouse mutation in which mouse embryonic stem cells are targeted by homologous recombination is generated. This conditional allele harbors *loxP* recognition sites flanking critical DNA sequences such as exons and/or regulatory elements of the gene of interest. The Cre driver, on the other hand, is a transgenic mouse in which the expression of the enzyme Cre recombinase is controlled by tissue-specific regulatory elements. In order for this system to be inducible through which the precise timing of gene activation can be controlled, a fusion polypeptide between Cre recombinase and a mutant ligand binding domain of the estrogen receptor (called *CreER^{T2}*) was also generated (Indra et al., 1999). The G400V/M543A/L544A triple mutation of the human estrogen receptor prevents binding to the natural ligand beta estradiol and retains the fusion protein in the cytosol. However, upon association with the synthetic ligand 4-hydroxy

(OH)-tamoxifen - an antagonist of the estrogen receptor – the protein translocates into the nucleus, where it mediates the excision of the *loxP*-flanked DNA sequence. Tamoxifen can therefore be administered to mice at specific times either during embryonic development or in adult animals to allow for temporal control of gene expression (Leone et al., 2003).

Eventually, the production of an inducible chondrocyte-specific transgenic mouse model will enable us to activate/inactivate different genes specifically in cartilage following or preceding the induction of OA through surgery. For example, the mouse model can be mated to floxed transgenic mice in which particular genes (such as FGF-2, cytokines or TIMP-3) or a transcriptional stop sequence (that inhibits transcription) can be excised following activation of Cre recombinase. This can be done at any point in the progeny with the injection of tamoxifen.

Increasing our understanding of the regulation of aggrecan would also be of direct importance towards our understanding of several cartilage diseases, such as those that are characterized by malformations of skeletal elements (chondrodysplasias) as well as those that are characterized by progressive and irreversible loss of adult articular cartilage (OA) since several disease-causing mutations have been found within the gene of humans and mice. For instance, as described in the 'Introduction section 3.5', there is a nonsense mutation in *Agc1* that leads to mouse cartilage matrix deficiency (Watanabe et al., 1994), (Kimata et al., 1981). Heterozygotes develop mild dwarfism along with spinal degeneration (Watanabe et al., 1997b). Homozygotes, on the other hand, die at birth and display severe dwarfism (Rittenhouse et al., 1978). In humans, mutations of the aggrecan gene have been associated with spondyloepiphyseal dysplasia, OA and intervertebral disc degeneration (Gleghorn et al., 2005), (Horton et al., 1998), (Roughley et al., 2006). Furthermore, different forms of arthritis feature early with severe loss of aggrecan protein due to increased degradation and/or decreased production (see section 3.4).

5 Aims of PhD project

Since little is known about the human aggrecan promoter, I have set out to try to identify the transcriptional regulatory mechanisms that direct chondrocyte specific expression of the **Human** aggrecan gene, which will in turn give us greater insight into osteoarthritis pathogenesis and therapies. This will be achieved by:

- 1- Identifying and characterising the key regions in the human gene involved in regulating chondrocyte specific expression.
- 2- Isolating these regions from genomic human DNA library clones and cloning them into expression cassettes for generating transgenic mice.
- 3- Generating an inducible cartilage specific expression system using an aggrecan promoter and testing the functionality and specificity of this system under normal conditions as well as injury.
- 4- Analysing the response of an identified human aggrecan promoter to injury.

Materials and Methods

1 Isolation of nucleic acids:

1.1 Isolation of plasmid DNA

1.1.1 Mini-prep of plasmid DNA

A single colony of transformed DH5 α *E.coli* was used to inoculate 2 ml of Luria broth (LB) with appropriate antibiotic selection at the appropriate concentration (100 μ g/ml ampicillin and 20 μ g/ml chloramphenicol). This was then incubated overnight at 37°C with shaking. The following day, 1.4 ml of overnight culture was transferred into a sterile 1.5 ml microcentrifuge tube and centrifuged at 10,000 x g for 5 minutes. The supernatant was discarded and the pellet of cells resuspended in 100 μ l of buffer P1 (re-suspension buffer, Qiagen) by vortexing. This was followed by the addition of 100 μ l of buffer P2 (lysis buffer, Qiagen). The solution was inverted 6-8 times and incubated for 5 minutes. Next, 100 μ l of buffer P3 (neutralisation buffer, Qiagen) was added. The solution was again inverted 6-8 times and incubated on ice for 15 minutes. Following this, the chilled solution was centrifuged at 12,000 x g for 5 minutes. The supernatant was transferred into a clean microcentrifuge tube containing 750 μ l of ice-cold 100% ethanol. This was centrifuged again at 12,000 x g for 10 minutes at 4°C. The supernatant was removed and the pelleted DNA washed in 100 μ l 70% ethanol (at room temperature). Once the ethanol was removed, the DNA pellet was allowed to air dry. Lastly, the DNA pellet was resuspended in 40 μ l of TE buffer. The purified plasmid was checked by digestion with appropriate enzymes and electrophoresed in a 1% (weight: volume) agarose gel, supplemented with ethidium bromide (0.2 ng/ μ l). The DNA was stored at -20°C.

1.1.2 Maxi-prep of plasmid DNA

50 μ l of the starter culture described above was used to inoculate 200 ml of LB (with appropriate antibiotic) and incubated overnight at 37°C with shaking. The 'Qiagen Plasmid Maxi-prep Kit Cat No. 12163' was used according to the manufacturer's instructions to prepare the plasmid DNA from this culture.

1.2 Isolation of genomic DNA from embryonic and adult tails

The embryonic tail or adult tail biopsies were placed in sterile tubes with 200 μ l of tail buffer (see Methods section 9.4) and 20 μ l of proteinase K (10mg/ml, Sigma). After incubation at 55°C overnight, 200 μ l of 3.0M NH_4OAc pH 8.0 was added. The tube was then vortexed and placed on ice for 5-10 minutes. Following this, the mix was spun down at 13,000 x g for 5 minutes. The supernatant was transferred to a new tube and 200 μ l of chloroform was then added. After vortex mixing, the solution was spun down at 13,000 x g for a further 5 minutes. Approximately 350 μ l of the upper aqueous phase was removed into fresh tubes, and 1ml of absolute ethanol was added to precipitate the DNA. The tubes were centrifuged at 13,000 x g for 10 minutes at 4°C, the ethanol decanted and the resultant pellet washed in 100 μ l of 70% ethanol. This ethanol was then removed and the resultant pellet left to air dry for 1-2 hours. 50 μ l and 200 μ l of buffer TE were then added to the embryonic and adult tails respectively. The DNA was allowed to fully dissolve overnight.

1.3 Oligonucleotides

The web-based program primer 3 (www.primer3.com) was used to design all oligonucleotides. These were then obtained from Sigma Aldrich and the stock re-suspended in TE buffer to a concentration of 100 pmol/ μ l, and stored at -20°C.

2 Standard manipulations of DNA

In order to ligate DNA fragments into vectors and unless stated otherwise, inserts were generated by PCR, both were digested with appropriate restriction enzymes (vector and insert), isolated by gel electrophoresis, purified and then ligated. Vectors were always dephosphorylated.

2.1 Polymerase chain reaction

Proofreading Pfx polymerase (Invitrogen) was used for the amplification of DNA fragments to be cloned into constructs directly used for microinjections. Taq polymerase (GoTaq®Flexi DNA Polymerase Promega Cat No. M8291) was used for the amplification of DNA fragments to be cloned into constructs used for recombineering. For Pfx polymerase, the reaction was carried out with the final concentrations as follows: 50 ng of plasmid DNA or 100 ng of BAC DNA (template), 0.3 μ M primer (for each forward and reverse), 1X Pfx amplification buffer, 1 mM MgSO_4 , 0.3 mM each dNTP, 1.0-2.5 Units of Pfx polymerase depending on amplicon size of DNA product and finally, distilled water made up to 50 μ l. For Taq polymerase, the reaction was carried out with the final concentrations as follows: 50 ng of plasmid DNA or 100 ng of BAC DNA (template DNA), 0.1-1.0 μ M primer (for each forward and reverse), 1X reaction buffer, 1-4 mM MgCl_2 , 0.2 mM each dNTP, 1.25 U Taq polymerase and distilled water made up to 50 μ l. The DNA was denatured at 95°C and an annealing temperature range of 55-60°C was used in the form of a 'touchdown PCR' to account for different % of GC content in primers. The final part of the cycle was an extension at 72°C. The number of cycles used was 20-30 depending on the nature of the experiment.

2.2 Digestion of DNA by restriction enzymes

All restriction enzymes were obtained from New England Biolabs (NEB). The reactions were performed under conditions described in the NEB catalogue. The final reaction volumes varied depending on the volume of DNA to be digested with incubation periods of 1-8 hours.

2.3 Agarose gel electrophoresis

Gels were made at 1% (w:v) using agarose (Sigma) and 1 x TAE buffer (see Methods section 9.1). The agarose was added to the TAE buffer and dissolved by boiling in a microwave for approximately 2 minutes per 100 ml. Ethidium bromide was then added to the cooled molten gel at a concentration of 0.2 $\mu\text{g}/\mu\text{l}$ to enable visualisation of DNA under UV exposure. A loading buffer was always added to the DNA sample to be run at a final concentration of 1X prior to loading. Electrophoresis (Flowgen Horizontal 10 X 10 cm unit for restriction enzyme site analysis) was carried out at a constant field of 5-10 V/cm. DNA was visualised by Transillumination with ultraviolet light at 312 nm (for cloning) and 254 nm (for analysis).

2.4 Isolation and purification of DNA from agarose gels

For DNA samples ranging in size from 100bp-10kb, the 'Qiagen QIAquick Gel Extraction Kit Cat. No. 28704' was used for isolation and purification of DNA from agarose gels. This was done as described in the manufacturer's protocol, with the resultant DNA suspended in a final volume of 30-50 μl of provided 'elution buffer'. For DNA samples >10kb, the 'Qiagen QIAEX II Gel Extraction Kit Cat No. 20021' was used instead. This was done as described in the manufacturer's protocol, with the resultant DNA suspended in a final volume of 20-40 μl of provided 'elution buffer'. Electrophoresis of a sample of the eluted DNA was carried out to determine the quality and concentration. A nano-drop spectrophotometer was then used to measure the final concentration of the DNA.

2.5 Recombinant DNA procedures

2.5.1 Blunt ending of 5' and 3' overhangs

T4 DNA polymerase (NEB 3000 Units/ml) was used to blunt end 5' and 3' overhangs. Reactions were performed under conditions described in the NEB catalogue.

2.5.2 Dephosphorylation

Dephosphorylation of 5' terminal phosphate groups was performed using Antarctic Phosphatase (NEB 5000 Units/ml) in an appropriate volume supplemented with 1 x Antarctic phosphatase buffer at 37°C for 30 minutes. Reactions were performed under conditions described in the NEB catalogue.

2.5.3 DNA ligation

T4 DNA ligase (NEB) was used to ligate insert DNA fragments into vector DNA. The ligations were performed with 20-50 ng of vector DNA and insert:vector molar ratio was at 3:1. Both sticky and blunt-ended ligations reactions were at a final volume of 20 µl for 4 hours at 25°C, and for blunt- were performed at a final volume of 20 µl for 16 hours at 15°C.

2.6 Transformation of DH5α E.coli

DH5α cells were transformed with newly cloned vectors. The competent cells were taken out of -80°C storage and thawed on ice for 20 minutes. Next, 10 µl of the ligation solution was added to 120 µl of cells, mixed and incubated on ice for 30 minutes. The transformation reaction was heat-shocked at 37°C for 5 minutes and then incubated on ice for a further 5 minutes. Following this, 500 µl of LB was added to the cells and incubated in a 250 rpm bacterial shaker for 1 hour at 37°C. The cells were spun down, re-suspended in a smaller volume of LB and plated on LB-agar plates with appropriate antibiotic selection. These were incubated at 37°C

overnight. Depending on the number of colonies that grew on the plate, 1-20 colonies were picked, grown overnight in 2 ml of medium and miniprepmed. Restriction enzyme analysis was performed to identify successful cloning.

3 Cloning vectors

3.1 pBluescript

DNA fragments were cloned into pBluescript®II KS (+) phagemid (Figure 10) (Stratagene) as an intermediate vector for carrying out recombineering (see Methods section 4). pBluescript was either used as the targeting vector or the gap-repaired vector. All desired DNA fragments were eventually transferred into the appropriate expression vector, p β gal-Basic. pBluescript was grown with ampicillin selection.

3.2 p β gal-Basic

The p β gal-Basic vector (Clontech) (Figure 11) was used in all *in vivo* transgenic work in order to identify potential human aggrecan regulatory sequences. p β gal-Basic is a mammalian reporter vector that lacks promoter or enhancer elements and so can be used to analyse the potential regulatory activity of cis-regulatory elements on gene expression in mammalian cells by assaying for β -galactosidase. The vector has a polyadenylation signal upstream of the multiple cloning site (MCS), which reduces the background expression of β -galactosidase. It has a pUC origin of replication and an ampicillin resistance gene both allowing for the propagation and selection of varying forms of the vector in *E.coli*. All DNA fragments to be tested for regulatory activity were cloned into p β gal-Basic.

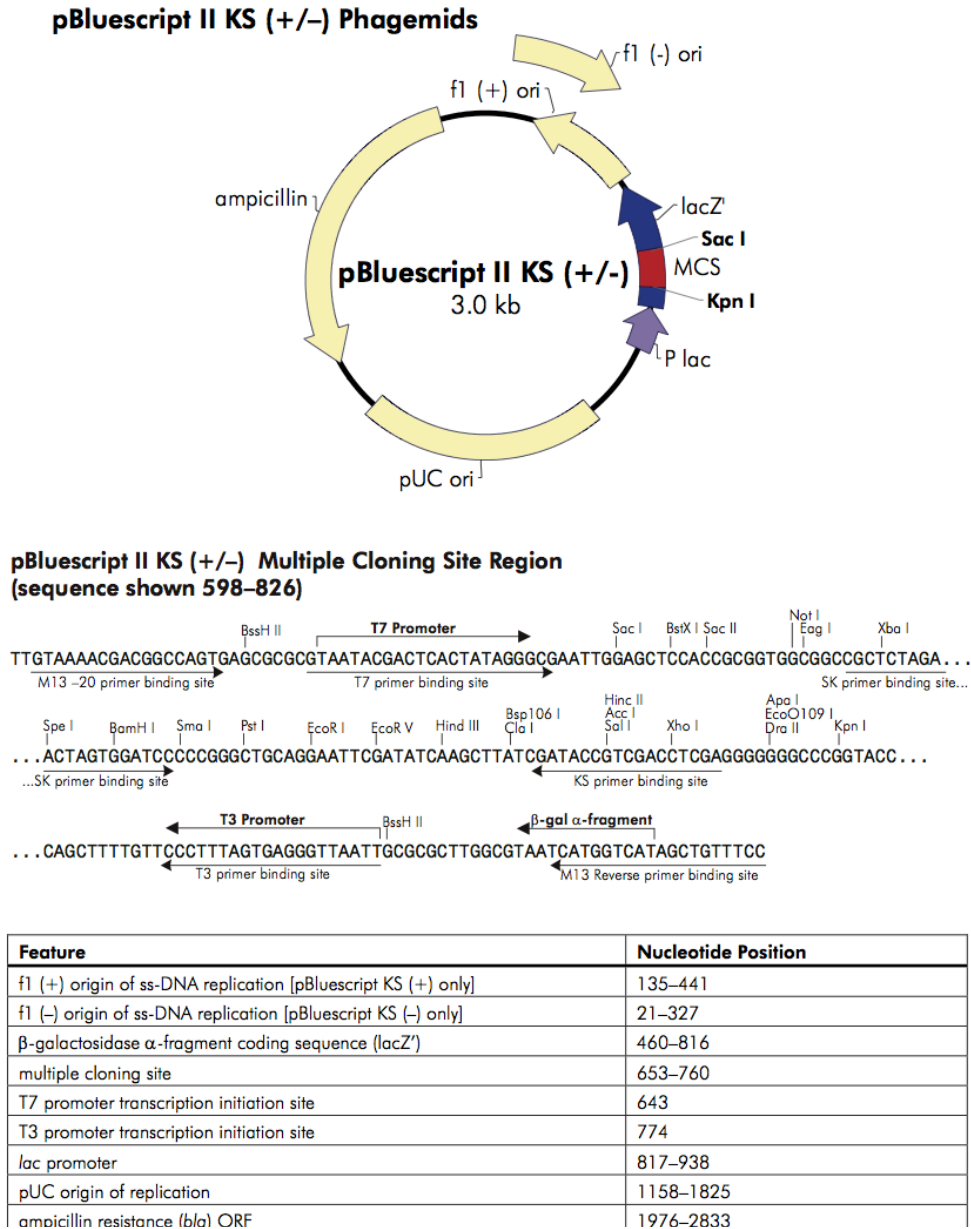
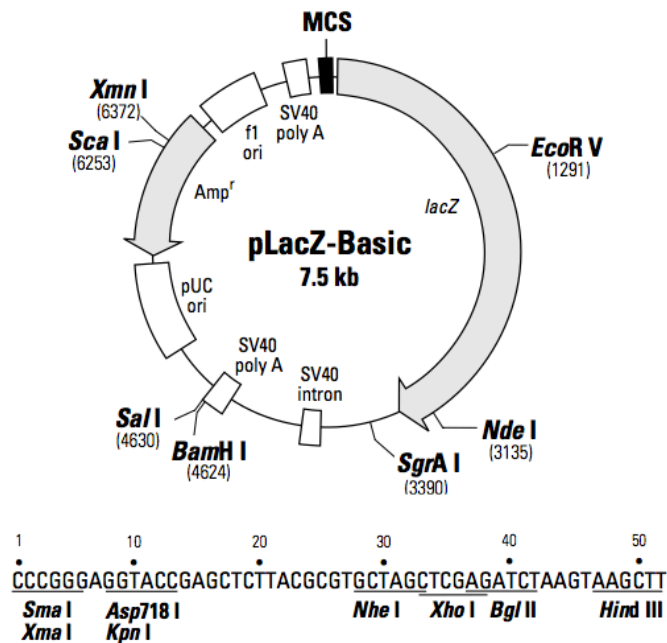


Figure 10: pBluescript II KS (+/-) vector. Taken from http://www.genomics.agilent.com/files/Manual/212205_A01.pdf

pLacZ-Basic Vector Information
GenBank Accession No. U13184

PT2109-5
Cat. No. 631707



Location of features

- Multiple cloning site (MCS): 1–52
- β -galactosidase fusion protein
 - Start codon (ATG): 94–96; stop codon: 3235–3237
 - Amino acids from *D. melanogaster* alcohol dehydrogenase: 94–186
 - Amino acids from *E. coli* β -galactosidase (1): 190–3234
- SV40 small t antigen intron: 3853–3918
- SV40 early mRNA polyadenylation signals: 4524–4529 & 4553–4558
mRNA 3' ends: 4562 & 4574
- pUC plasmid origin of replication: 4908–5551
- Ampicillin resistance (β -lactamase) gene:
 - Promoter: –35 region: 6629–6624; –10 region: 6606–6601
 - Transcription start point: 6594
 - Ribosome binding site: 6571–6567
 - β -lactamase coding sequences: start codon (ATG): 6559–6557; stop codon: 5701–5699
 - β -lactamase signal peptide: 6559–6491; β -lactamase mature protein: 6490–5702
- f1 single-strand DNA origin (packages the coding strand of the *lacZ* gene): 6691–7146
- SV40 early mRNA polyadenylation signals: 7367–7372 & 7396–7401

Figure 11: p β gal-Basic vector. Taken from <http://www.clontech.com/images/pt/PT2109-5.pdf>

3.3 RP11-343B18

RP11-343B18 is a bacterial artificial chromosome (BAC) cloned from *Homo sapiens* chromosome 15 and is 197370bp in size. It contains the entire human aggrecan gene along with >100kb of non-coding sequences upstream of the 5' end of the gene. Sequencing of the BAC revealed a deleted fragment of DNA 92bp in size located between -68bp to +23bp

relative to the human aggrecan TSS. Rp11-343B18 was grown with chloramphenicol selection. This BAC was obtained from the Sanger Institute (Cambridge, UK).

3.4 RP11-1123E8

RP11-1123E8 is a bacterial artificial chromosome (BAC) cloned from *Homo sapiens* chromosome 15 and is 142971bp in size. It contains the entire human aggrecan gene along with >80kb of non-coding sequences upstream of the 5' end of the gene. Sequencing of the BAC revealed that the 92bp sequence missing from RP11-343B18 was present in RP11-1123E8. RP11-11238 was grown with chloramphenicol selection. This BAC was obtained from CHORI (Children's Hospital Oakland Research Institute, CA).

4 Recombineering

The recombineering strain EL350 *E.coli* was used throughout the first part of the work described here. This was later substituted with the updated strain SW106. SW106 cells are derived from EL350 cells but can also be used for *galK* positive/negative selection. Like EL350, they contain an *ara*-inducible Cre gene. All recombineering strains were obtained from the National Institute of Health (<http://recombineering.ncifcrf.gov/>) in the USA.

4.1 Making electro-competent EL350 or SW106 *E.coli* (for introducing BAC DNA, prior to recombineering):

On day 1, the recombineering cells were streaked on an antibiotic free LB/agar plate. This was then incubated at 32°C overnight. On day 2, a colony was used to inoculate 2 ml of LB without antibiotic (grown overnight at 32°C and at 220 rpm). On day 3, 2 ml of the overnight culture was used to inoculate 100 ml LB without antibiotic (grown at 32 °C until OD₆₀₀=0.4-0.6). Next, the cells were pelleted at 3500 rpm for 15 minutes at 4°C. They were then washed 3 x in 30 ml 10% glycerol. Finally, they were resuspended in a few mls of 10% glycerol. Cells were used immediately or stored in 40 µl aliquots at -80°C.

4.2 Transforming a target BAC (RP11-343B18/RP11-1123E8) into electro-competent cells

An aliquot of cells was taken out of -80°C storage and thawed on ice. Approximately 100 ng of 'RP11-343B18' or 'RP11-1123E8' DNA was added to these cells and incubated on ice for 10 minutes. The DNA/cells mixture was then transferred to chilled 0.1 cm cuvettes (BIORAD 165-2089) and pulsed using an electroporator (Resistance: 200 Ω, Volts: 1.75 V, Capacitance 2.5 µF). The time constant should be approximately 4-5 ms. One ml of antibiotic free LB was then

quickly added to the cuvette and the solution transferred to a clean 15 ml tube. This was then shaken at 32°C for 1 hour. An LB/agar plate with chloramphenicol was then spread with 300 µl of the electroporated cells and incubated at 32°C overnight. A glycerol stock was always made from the transformed recombineering cells.

4.3 Recombineering

Once the appropriate BAC (RP11-343B18 or RP11-1123E8) had been introduced into the recombineering competent cells, they had to be made electro-competent again, in order to allow for transformation of the recipient (retreival) plasmid. On day 1, cells were streaked from a stock of EL350 or SW106 cells containing the desired BAC on an LB/agar plate with chloramphenicol selection and incubated overnight at 32°C. On day 2, a colony was used to inoculate 2 ml of LB with appropriate antibiotic resistance (grown overnight at 32°C and at 220 rpm). On day 3, 2 ml of the overnight culture was used to inoculate 100 ml LB containing chloramphenicol (grown at 32 °C until OD₆₀₀=0.4-0.6). The 100 ml culture was then aliquoted into 5 x 10 ml samples on ice. Next, the samples were subject to heat shock by shaking at 42°C for 15 minutes. These were then cooled in ice-cold water and pooled together. Next, the cells were pelleted at 3500 rpm for 15 minutes at 4°C. They were then washed 3 x in 30 ml 10% glycerol. Finally, they were resuspended in a few mls of 10% glycerol. 1-3 µl of 20-50 ng recipient plasmid DNA was immediately electroporated into 40 µl aliquots of the cells as described in Methods section 4.1. The electroporation of the targetting DNA into these prepared E-competent cells to be recombined with RP11343B18 or RP11-1123E8 was carried out using the protocol in Methods section 4.2 with one exception; BAC DNA is substituted with the targetting DNA.

5 Generation of Constructs

Unless stated otherwise, all the locations of the sequences below are relative to the Human AggreCAN TSS based on a TSS at position 87,147,670 on chromosome 15 (Assembly: March. 2006 (NCBI36/hg18)).

AB: A short sequence representing the core/basal promoter of the human aggreCAN gene and located -452bp to +375bp. This sequence, however, contains a deletion corresponding to 92bp between -68 to +23bp.

AB⁺⁹²: A short sequence representing the core/basal promoter of the human aggreCAN gene and located -452bp to +375bp. This sequence includes the missing 92 bp.

EF: A short DNA sequence located -20,054bp to -19,227bp.

cis 3: A short DNA sequence located -10,703bp to -10,343bp.

IJ: A short DNA sequence located -13,519bp to -13,019bp.

cis 7: A short sequence located within the 14th intron of the human aggreCAN gene at +64,630bp to +65,005bp.

pBS/*cis* 6 5'arm: A short sequence of DNA encompassing 416bp of AB which corresponds to the location -41bp to +375bp with an additional 31 base pairs at its 3'end corresponding to the sequence in the pβgal-Basic vector from the *Kpn*/site to the *Xho*/site in the MCS.

pBS/*cis* 6 3'arm: A short sequence of DNA encompassing the first 292bp of the *LacZ* gene within the the pβgal-Basic vector with an additional 150bp at the 5'end corresponding to the sequence found in the pβgal-Basic vector between the *Xho*I site in the MCS and the *LacZ* gene.

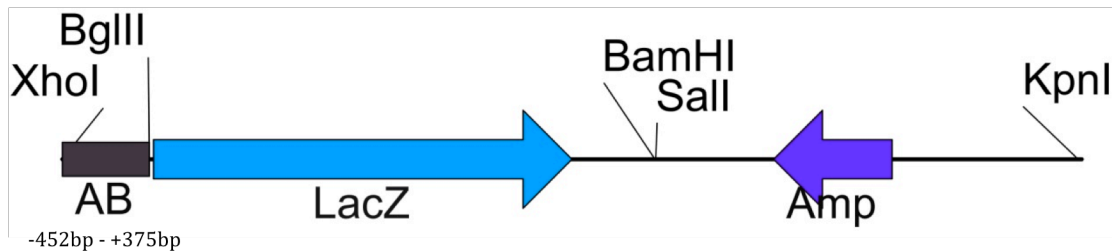
cis 6: An approximately 3kb sequence located +5,000bp to +7,813bp.

A1: A short sequence located -9,477bp to -9,118bp relative to the mouse *Agc* TSS. A1 is the mouse equivalent sequence to the human *cis* 3 fragment.

Table 4: List of constructs generated. The names of the constructs are shown in the middle column while the methods in which they were cloned are shown in the column on the right. The constructs are illustrated below with further details on their generation.

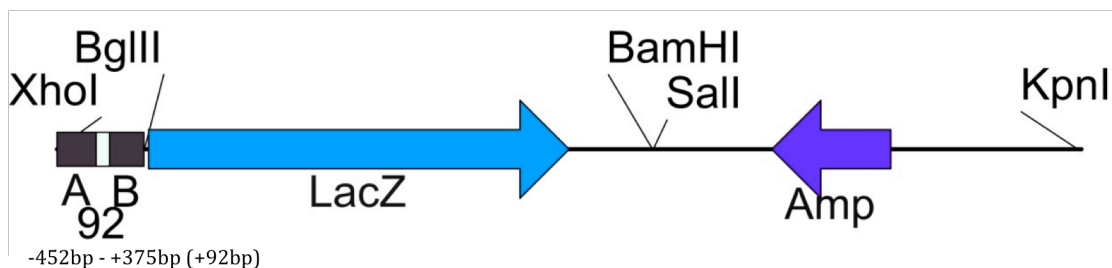
	Name of Construct	Method of Cloning
1	BgalAB	restriction site cloning
2	BgalAB ⁺⁹²	restriction site cloning
3	BgalAB+EF	restriction site cloning
4	βgalAB ⁺⁹² +EF	restriction site cloning
5	βgal AB + <i>cis</i> 3	restriction site cloning
6	βgalAB ⁺⁹² + <i>cis</i> 3	restriction site cloning
7	βgalAB + IJ + EF	restriction site cloning
8	βgalAB + A1 + A4	restriction site cloning
9	pBS/ <i>cis</i> 7 recomb	restriction site cloning
10	pBSAB + EF	restriction site cloning
11	pBSAGG20	recombineering
12	βAGG20	restriction site cloning
13	βAGG10	restriction site cloning
14	βAGG13	restriction digestion
15	βgalAB + 7	recombineering
16	βAGG20 + int3	recombineering
17	ColII4xA1Cre	restriction site cloning using sequence from Veronique Lefebvre's group, Lerner Research Institute, Ohio

1- β galAB



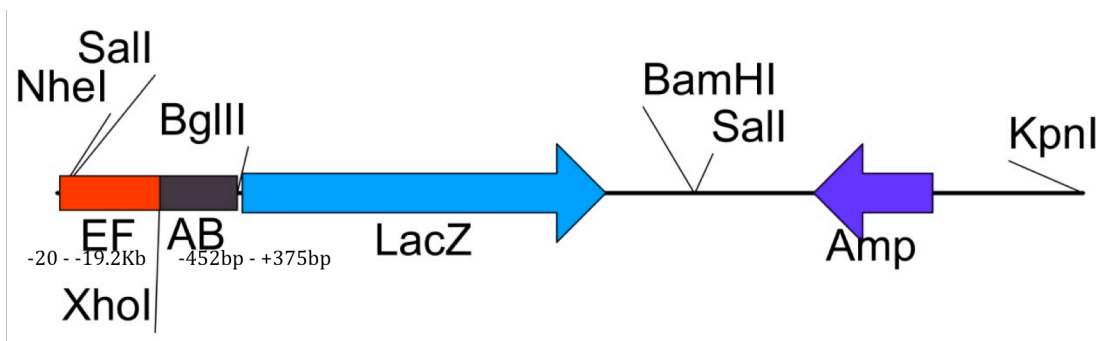
p β gal-Basic vector was digested with *BglII* and *XhoI* restriction enzymes. AB was amplified by PCR performed on the BAC RP11343-B18. The primers used were 'forward A (*XhoI*)' and 'reverse B (*BglII*)'. The bracketed restriction sites *XhoI* and *BglII* were designed at the 5'ends of these primers. The subsequent PCR product AB was verified on an agarose gel and digested overnight with *XhoI* and *BglII* and finally ligated to the digested p β gal-Basic vector. The transgene was excised from the plasmid backbone by digestion with *KpnI* and *BamHI*.

2- β galAB⁺⁹²



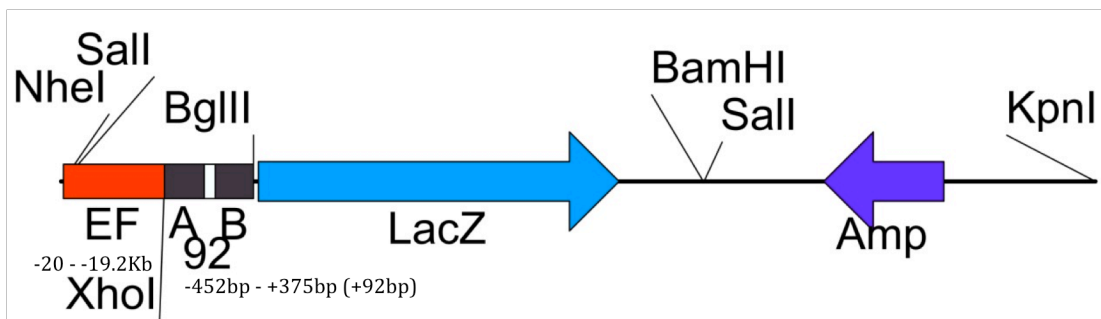
p β gal-Basic vector was digested with *BglII* and *XhoI* restriction enzymes. AB⁺⁹² was amplified by PCR performed on the BAC RP11-1123E8. The primers used were 'forward A (*XhoI*)' and 'reverse B (*BglII*)'. The bracketed restriction sites *XhoI* and *BglII* were designed at the 5'ends of these primers. The subsequent PCR product AB⁺⁹² was verified on an agarose gel and digested overnight with *XhoI* and *BglII* and finally ligated to the digested p β gal-Basic vector.

3- β galAB+EF



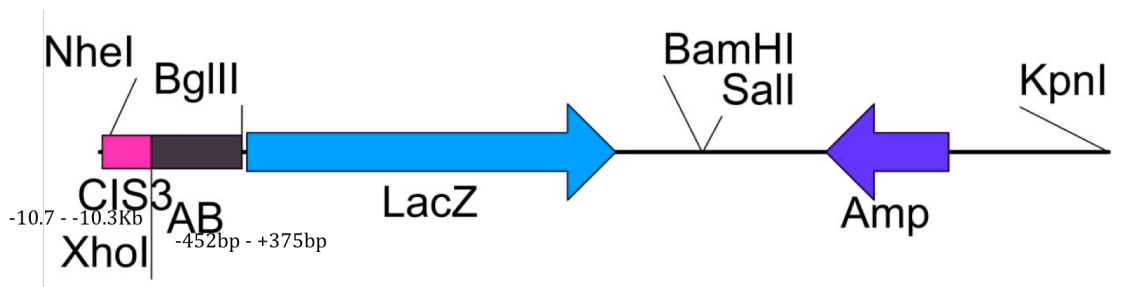
p β gal-Basic vector was digested with *Bgl*III and *Nhe*I restriction enzymes. AB was amplified by PCR performed on the BAC RP11343-B18. The primers used were 'forward A (*Xho*I)' and 'reverse B (*Bgl*III)'. The bracketed restriction sites *Xho*I and *Bgl*III were designed at the 5'ends of these primers. EF was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'forward E (*Nhe*I/*Sall*)' and 'reverse F (*Xho*I)'. The bracketed restriction sites 5'*Nhe*I+*Sall*3' and *Xho*I were designed at the 5'ends of these primers. The subsequent PCR product AB was verified on an agarose gel and digested overnight with *Xho*I and *Bgl*III. EF however was prepared in the same way but digested with *Nhe*I and *Xho*I instead. Finally, they were ligated via a three-way ligation to the digested p β gal-Basic vector. β galAB+EF was used to clone β galAB+IJ+EF.

4- β galAB⁺⁹²+EF



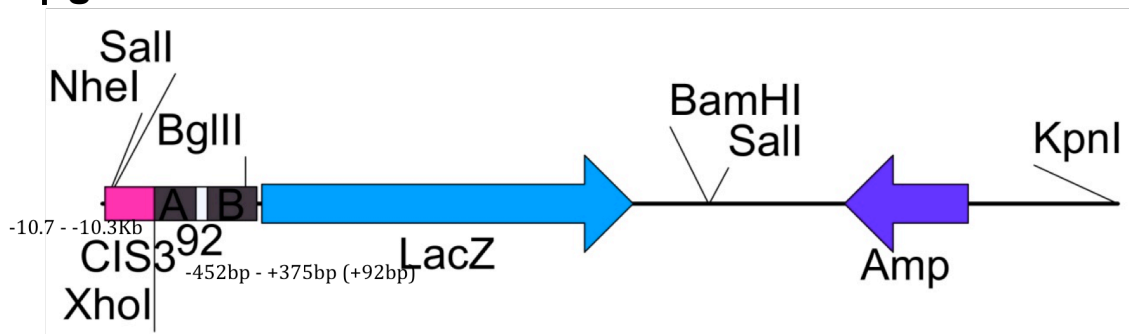
p β gal-Basic vector was digested with *Bgl*III and *Nhe*I restriction enzymes. AB⁺⁹² was amplified by PCR performed on the BAC RP11-1123E8. The primers used were 'forward A (*Xho*I)' and 'reverse B (*Bgl*III)'. The bracketed restriction sites *Xho*I and *Bgl*III were designed at the 5'ends of these primers. EF was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'forward E (*Nhe*I)' and 'reverse F (*Xho*I)'. The bracketed restriction sites *Nhe*I and *Xho*I were designed at the 5'ends of these primers. The subsequent PCR product AB⁺⁹² was verified on an agarose gel and digested overnight with *Xho*I and *Bgl*III. EF however was prepared in the same way but digested with *Nhe*I and *Xho*I instead. Finally, they were ligated via a three-way ligation to the digested p β gal-Basic vector. β galAB⁺⁹²+EF will be used for recombineering of β AGG20⁺⁹².

5- β galAB+ *cis* 3



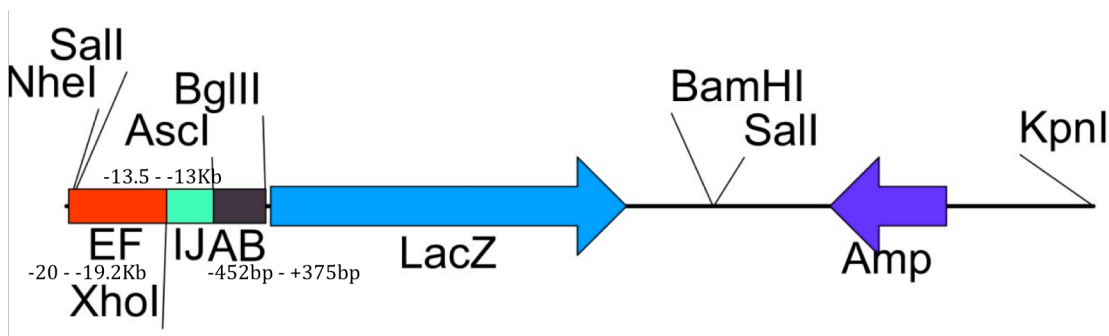
p β gal-Basic vector was digested with *BgIII* and *NheI* restriction enzymes. AB was amplified by PCR performed on the BAC RP11343-B18. The primers used were 'forward A (*XhoI*)' and 'reverse B (*BgIII*)'. The bracketed restriction sites *XhoI* and *BgIII* were designed at the 5'ends of these primers. *cis* 3 was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'forward *cis* 3 (*NheI*)' and 'reverse *cis* 3 (*XhoI*)'. The bracketed restriction sites *NheI* and *XhoI* were designed at the 5'ends of these primers. The subsequent PCR product AB was verified on an agarose gel and digested overnight with *XhoI* and *BgIII*. *cis* 3 however was prepared in the same way but digested with *NheI* and *XhoI* instead. Finally, they were ligated via a three-way ligation to the digested p β gal-Basic vector. The transgene was excised from the plasmid backbone by digestion with *KpnI* and *BamHI*.

6- β galAB⁺⁹²+ *cis* 3



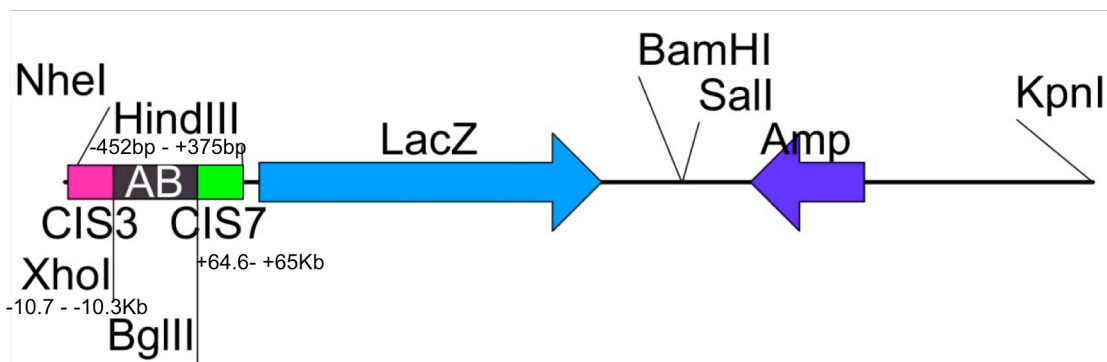
p β gal-Basic vector was digested with *BgIII* and *NheI* restriction enzymes. AB⁺⁹² was amplified by PCR performed on the BAC RP11-1123E8. The primers used were 'forward A (*XhoI*)' and 'reverse B (*BgIII*)'. The bracketed restriction sites *XhoI* and *BgIII* were designed at the 5'ends of these primers. *cis* 3 was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'forward *cis* 3 (*NheI*)' and 'reverse *cis* 3 (*XhoI*)'. The bracketed restriction sites *NheI* and *XhoI* were designed at the 5'ends of these primers. The subsequent PCR product AB⁺⁹² was verified on an agarose gel and digested overnight with *XhoI* and *BgIII*. *Cis* 3 however was prepared in the same way but digested with *NheI* and *XhoI* instead. Finally, they were ligated via a three-way ligation to the digested p β gal-Basic vector. The transgene was excised from the plasmid backbone by digestion with *KpnI* and *BamHI*.

7- β galAB+IJ+EF



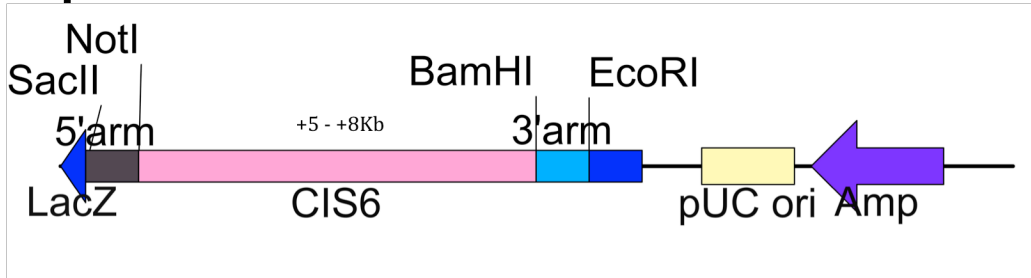
β galAB+EF was digested with *XhoI* and *Ascl* restriction enzymes. *Ascl* restriction site in β galAB+EF is located 264bp downstream of the *XhoI* site within AB. IJ was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'forward I (*XhoI*)' and 'reverse F (*Ascl*)'. The bracketed restriction sites *XhoI* and *Ascl* were designed at the 5'ends of these primers. The subsequent PCR product IJ was verified on an agarose gel and digested overnight with *XhoI* and *Ascl*. Finally, it was ligated to the digested β galAB+EF vector. β galAB+IJ+EF was used for recombineering of β galAB+7.

8- β galAB+ *cis* 3+ *cis* 7



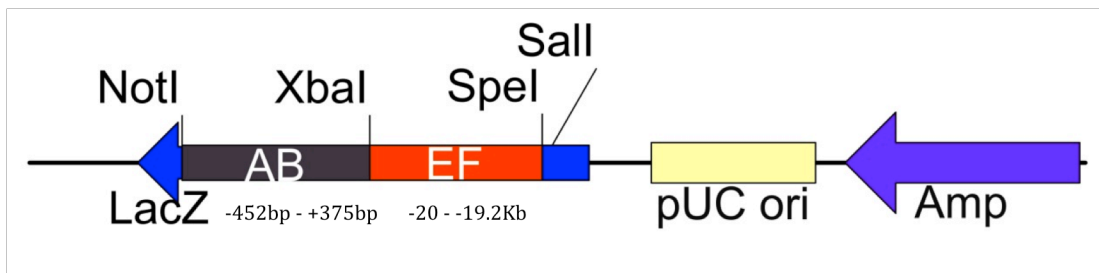
β galAB+*cis* 3 was digested with *BglII* and *HindIII* restriction enzymes. *HindIII* restriction site is found within the MCS of the p β gal-Basic vector. *Cis* 7 was amplified by PCR performed on the BAC RP11343-B18. The primers used were 'forward *cis* 7 (*BglII*)' and 'reverse *cis* 7 (*HindIII*)'. The bracketed restriction sites *BglII* and *HindIII* were designed at the 5'ends of these primers. The subsequent PCR product *cis* 7 was verified on an agarose gel and digested overnight with *BglII* and *HindIII*. Finally, it was ligated to the digested β galAB+*cis* 3 vector. The transgene was excised from the plasmid backbone by digestion with *KpnI* and *BamHI*.

9- pBS/*cis* 6 recomb



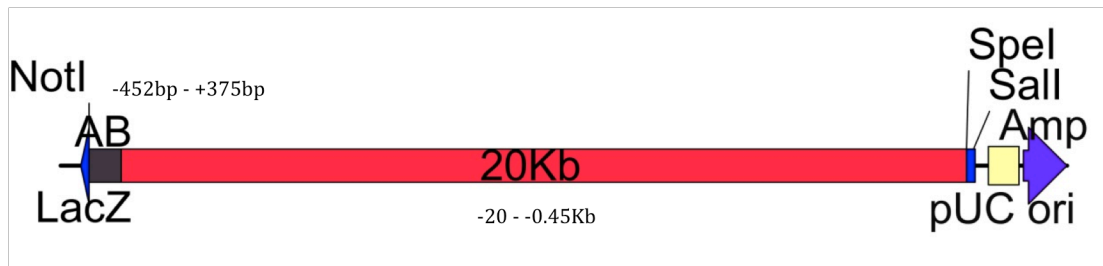
pBluescript®II KS (+) vector was digested with *SacII* and *EcoRI* restriction enzymes. The pBS/*cis* 6 5'arm was amplified by PCR performed on the vector βgalAB. The primers used were 'Forward 5'arm for pBS/*cis* 6 (*SacII*)' and 'Reverse 5'arm for pBS/*cis* 6 (*NotI*)'. The bracketed restriction sites *SacII* and *NotI* were designed at the 5'ends of these primers. The pBS/*cis* 6 3'arm was amplified by PCR performed on βgalAB. The primers used were 'Forward 3'arm for pBS/*cis* 6 (*BamHI*)' and 'Reverse 3'arm for pBS/*cis* 6 (*EcoRI*)'. The bracketed restriction sites *BamHI* and *EcoRI* were designed at the 5'ends of these primers. *cis* 6 was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'Forward *cis* 6 (*NotI*)' and 'Reverse *cis* 6 (*BamHI*)'. The bracketed restriction sites *NotI* and *BamHI* were designed at the 5'ends of these primers. The subsequent 5'arm PCR product was verified on an agarose gel and digested overnight with *SacII* and *NotI*. The 3'arm and *cis* 6 were verified in the same way but digested with *BamHI*+*EcoRI* and *NotI*+*BamHI* respectively. Finally, all three fragments were ligated via a four-way ligation to the digested pBluescript®II KS (+) vector. pBS/*cis* 6 recomb was used as the targeting vector for recombineering of βAGG20+ *cis* 6.

10- pBSAB + EF



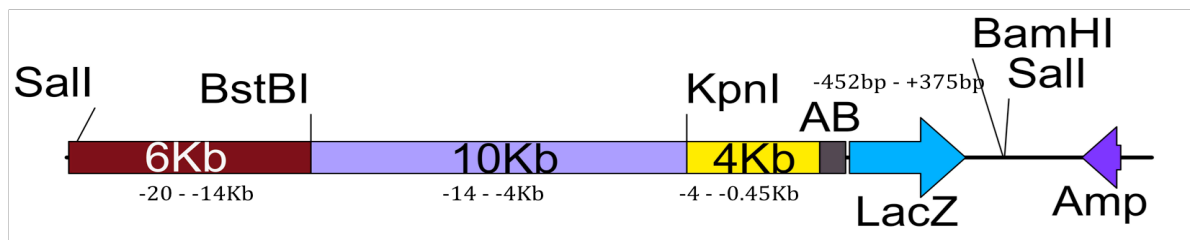
pBluescript®II KS (+) vector was digested with *SpeI* and *NotI* restriction enzymes. AB was amplified by PCR performed on the BAC RP11343-B18. The primers used were 'forward A (*XbaI*)' and 'reverse B (*NotI*)'. The bracketed restriction sites *XbaI* and *NotI* were designed at the 5'ends of these primers. EF was amplified by PCR performed on the BAC RP11-343B18. The primers used were 'forward E (*SpeI*)' and 'reverse F (*XbaI*)'. The bracketed restriction sites *SpeI* and *XbaI* were designed at the 5'ends of these primers. The subsequent PCR product AB was verified on an agarose gel and digested overnight with *XbaI* and *NotI*. EF however was prepared in the same way but digested with *SpeI* and *XbaI* instead. Finally, they were ligated via a three-way ligation to the digested pBluescript®II KS (+) vector. pBSAB+EF was used for recombineering of pBSAGG20.

11- pBSAGG20

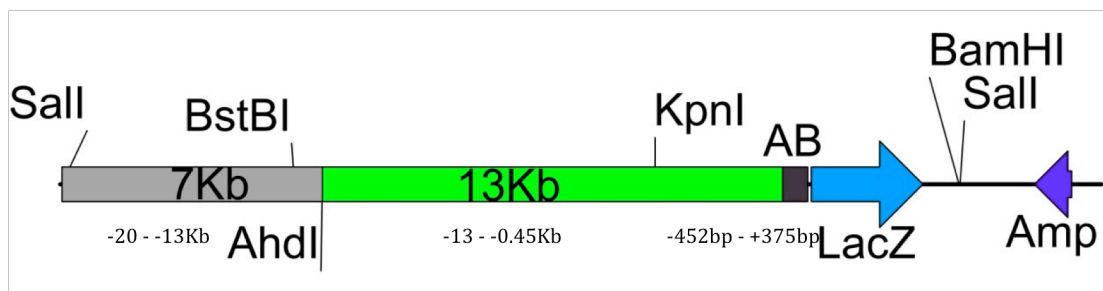


pBSAGG20, which contains approximately 20kb of human aggrecan non-coding sequence (-20,054bp to +375bp relative to the TSS), was generated by recombineering using pBSAB+EF as the gap-repaired vector. For more details, see Results Chapter 1 section 1.2.4.1.

12- β AGG20 (version 1)

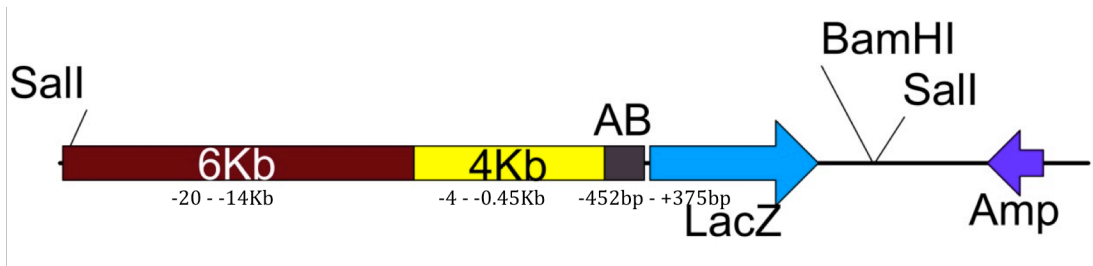


β AGG20 (version 2)



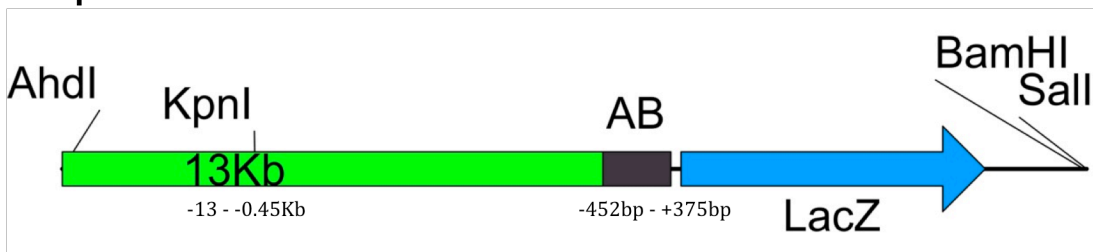
To generate β AGG20, the p β gal-Basic vector was digested with *KpnI* and blunt ended. Next, pBSAGG20 was digested with the restriction enzymes *SalI* and *NotI* to retrieve the non-coding 20kb human aggrecan sequence (-20,054bp to +375bp relative to the human Agc TSS), which was then blunt-ended. The two DNA fragments were then ligated to each other and a construct containing the 20kb sequence in the correct orientation was identified. The fusion of the blunt ended *SalI* site on the 20kb sequence to the blunt ended *KpnI* site on the p β gal-Basic vector regenerated a *SalI* site. The transgene was excised from the plasmid backbone by digestion with *SalI*. β AGG20 version 1 and 2 are the exact same construct with different representations to further clarify the construction of some of the subsequent constructs below.

13- β AGG10



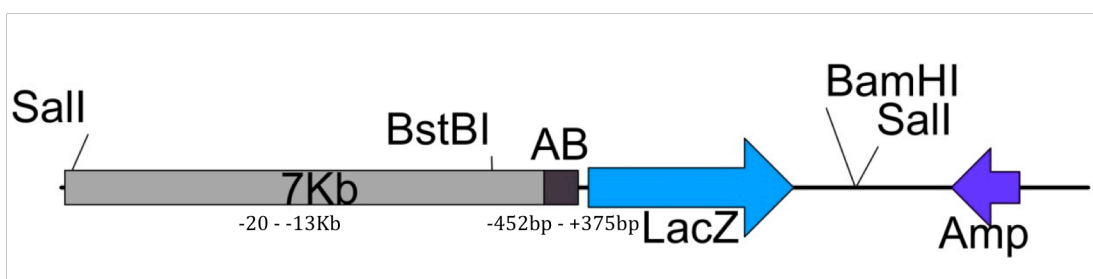
To generate β AGG10, the vector β AGG20 was digested with *BstBI* (-13,838bp relative to the human Agc TSS) and *KpnI* (-4,182bp relative to the human Agc TSS) to release an approximately 10kb fragment within the 20kb sequence in β AGG20 (See β AGG20 version one to note relative locations of restriction sites used). The vector was then blunt-ended and ligated to recircularize. It was excised from the plasmid backbone by digestion with *Sall*.

14- β AGG13



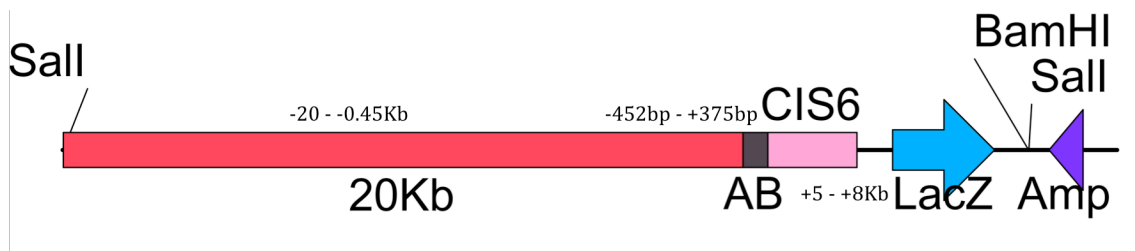
Unlike all the other vectors described, β AGG13 isn't a self-propagating vector and is merely a digested product of the vector β AGG20. To generate β AGG13, the vector β AGG20 was digested with *AhdI* (-13,030bp relative to the human Agc TSS) and *Sall* deleting the ampicillin resistance gene along with approximately 7kb of the 20kb sequence in β AGG20 (see β AGG20 version 2). The *AhdI* and *Sall* sites in the figure represent digested sites and the fragment is linear. β AGG13 was injected into pronuclei in this format.

15- β galAB+7



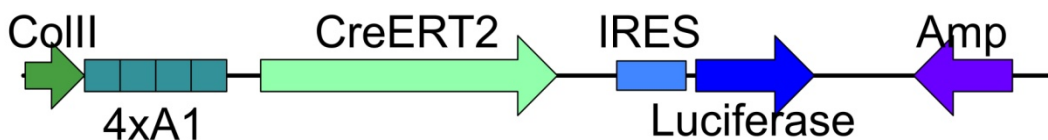
β galAB+7, which contains approximately 7kb of the human aggrecan non-coding sequence (-20,054bp to -13,030bp relative to the TSS) fused to AB, was generated by recombineering using β galAB+IJ+EF as the gap-repaired vector. It was excised from the plasmid backbone by digestion with *Sall*. For more details, see Results Chapter 1 section 1.2.5.2 .

16- β AGG20+ *cis* 6



β AGG20+*cis* 6, which contains approximately 20kb of human aggrecan non-coding sequence (-20,054bp to +375bp relative to the TSS) fused to *cis* 6 (located downstream of the 20kb sequence), was generated by recombineering using pBS/*cis* 6 recomb as the targeting vector and linearized β AGG20 as the gap-repaired vector. It was excised from the plasmid backbone by digestion with *Sall*. For more details, see Results Chapter 1 section 1.2.6.1.

17- ColII4xA1Cre



The original vector containing ColII4xA1 sequence was obtained from Veronique Lefebvre's group (Han and Lefebvre, 2008). The A1 sequence was amplified by PCR performed on 129 mouse genomic DNA. The Forward and Reverse primers used were designed with *Bam*HI and *Bgl*II restriction sites respectively at the 5' ends. The A1 PCR product was verified on an agarose gel and digested with *Bam*HI and *Bgl*II. It was then ligated into pCR4-TOPO (invitrogen) and a clone consisting of 4 tandem copies of the A1 sequence was selected. Cloning the four copies of A1 into a *LacZ* reporter containing the -309 to +308 Col2a1 sequence (Zhou et al., 1995) generated the colII4xA1 transgene. ColII4xA1Cre was generated by cloning colII4xA1 upstream of the CreER^{T2} fusion gene, an IRES sequence and the firefly luciferase gene.

List of primers used are shown below:

Table 5: Table representing the names of the primers used, their sequences and the generated products of each primer pair. Restriction cloning sites are underlined.

Name of Primer	Sequence of Primer	PCR product description
Forward (XbaI)	A 5' ATT <u>CTA GAC</u> CAG CCT CAG CTG CAG AAC C 3'	Basal promoter called AB to be cloned into pBS
Reverse B (NotI)	5' ATA AGA ATG <u>CGG CCG CCC</u> TTG AAG AGT TGG ACT CC 3'	
Forward E (SpeI)	5' TGA <u>CTA GTG</u> CTG AGC TGG GTT AAC ACC 3'	5' arm of ~20kb aggrecan sequence to be cloned into pBS
Reverse (XbaI)	F 5' AGT <u>CTA GAG</u> TTA AGC CCG CAA ACT TGT 3'	
Forward A (XhoI)	5' ATA <u>ACT CGA GCC</u> AGC CTC AGC TGC AGA ACC 3'	Basal promoter called AB to be cloned into pbgal basic
Reverse B (BglII)	5'ATA <u>AGA TCT</u> CCT TGA AGA GTT GGA CTC C 3'	
Forward E (NheI)	5' TAA <u>GCT AGC</u> GTC GAC GCT GAG CTG GGT TAA CAC C 3'	5' arm of ~20kb aggrecan sequence to be cloned into pbgal basic
Reverse F (XhoI)	5'ATA <u>ACT CGA GGT</u> TAA GCC CGC AAA CTT GT 3'	
Forward I (XhoI)	5' ATA <u>ACT TAA GCA</u> ACT GAG GCT GTT CAC AA 3'	3'arm of deleted 7kb sequence to be cloned into pβgal
Reverse J (AscI)	5' ATA <u>ACT CGA GCT</u> TGG ACC CAA AGT CAG AA 3'	
Forward cis 3 (NheI)	5' ATA <u>GCT AGC</u> TCC AGA ATG GAA GGG CCA AG 3'	cis 3 sequence to be cloned into pβgal basic
Reverse cis 3 (XhoI)	5' AAT <u>ACT CGA GCT</u> TTT CCA GAG TCA AAA GTC A 3'	
Forward cis 7 (BglII)	5' ATA <u>AGA TCT</u> GCA GCA GCA GCA ACA GTT C 3'	cis 7 sequence to be cloned into pβgal basic
Reverse cis 7 (HindIII)	5' ATA <u>AAG CTT</u> GGC CCC TGA AAT GGC TCT 3'	
Forward 5'arm for pBS/cis 6 (SacII)	5' ATA <u>CCG CGG</u> CCC GTC CCT ATG TAT GTG T 3'	5'arm sequence fused to cis 6 to be recombined into pβgal
Reverse 5'arm for pBS/cis 6 (NotI)	5' ATA AGA ATG <u>CGG CCG</u> CTC GAG CTA GCA CGC GTA AG 3'	

Forward 3'arm for pBS/ <i>cis</i> 6 (<i>Bam</i> HI)	5' AAG <u>GAT CCA</u> TCT AAG TAA GCT TGG GGG ATC G 3'	3'arm sequence fused to <i>cis</i> 6 to be recombined into pβgal
Reverse 3'arm for pBS/ <i>cis</i> 6 (<i>Eco</i> RI)	5' AAG <u>AAT TCA</u> TAG GTT ACG TTG GTG TAG ATG G 3'	
Forward <i>cis</i> 6 (<i>Not</i> I)	5' ATA AGA ATG <u>CGG CCG CGG</u> GAA GAC GGA CTG TAA CCA 3'	<i>cis</i> 6 sequence to be recombined into βAGG20.
Reverse <i>cis</i> 6 (<i>Bam</i> HI)	5' AAG <u>GAT CCA</u> AGG TCC ATC AGC ATT TT 3'	

6 Generation of transgenic mouse embryos and lines

Transgenic mice were generated in collaboration with Professor Dominic Wells and Ke Liu from the Kennedy Institute of Rheumatology (KIR). For all procedures, C57BL10 x CBA/J F1 mice were used. All techniques performed on animals were licensed under the Animals (Scientific Procedures) Act 1986, personal license number (PIL) 70/20608 and project license number (PPL) 70/6483.

6.1 Transgenic media

M2 and M16 media was purchased from Speciality Media (New Jersey, USA) and made up according to manufacturers' instructions. M16 was covered in mineral oil and incubated at 37°C in 5% CO₂, 95% air to simulate *in vivo* conditions during use. M2 was kept at room temperature.

6.2 Preparation of DNA for microinjection

Fragments of linear DNA for microinjection were isolated from their plasmid vectors with the appropriate restriction enzymes, blunt ended and gel-purified in Ethidium bromide (EtBr) free gels without UV exposure. This was done by running two identically digested samples alongside each other on an EtBr free gel, cutting the gel vertically between the two samples, staining one of the halves in an EtBr solution and exposing this half to UV to mark by gel excision the location of the desired band. Finally, the desired band was cut from the other non-stained half using the stained gel for guidance. The DNA was then purified from the gel using spin columns (see Methods section 2.4). The concentration of the DNA was established, diluted to 2 ng/μl with embryo water (Sigma), and stored at -20°C.

6.3 Preparation of egg donors by superovulation

Superovulation was performed on F1 females between 4-6 weeks of age. On day one, 5 IU follicle stimulating hormone (Intervet UK Lt; folligon) was administered to each mouse via intraperitoneal injection. On day 3, approximately 46 hours later, 5 IU lutanising hormone (Intervet UK Lt; chorulon) was administered to each mouse via intraperitoneal injection. Following this injection, the mice were mated with F1 males overnight.

On the morning of day 4, copulation plugs were checked to identify mated females. From these, oviducts were removed and dissected in M2 media giving an average of 10-15 eggs per oviduct. These were removed by forceps and transferred into clean media via mouth pipetting. Cumulus cells were removed by treatment with a few drops of hyaluronidase (300 $\mu\text{g/ml}$) in M2. The eggs were gently shaken until the cumulus cells had visibly detached from the egg surfaces (>4 minutes). They were thoroughly rinsed, and transferred to M16 media to be incubated for 1-2 hours.

6.4 Injection of fertilised eggs.

Injection of fertilised eggs was done approximately 72 hours after superovulation. This is described in Hogan et al (1994). Undamaged eggs were transferred to clean M16 and incubated until their transfer into pseudopregnant females.

6.5 Transfer to pseudopregnant recipients

Mating with vasectomised F1 males overnight on day 3 generated pseudopregnant mice. Females with copulation plugs were selected and used 1 day post-coitum. The mice were anaesthetised with 50 μl hypnorm (Janssen; fentanyl citrate, final concentration 0.79 mg/ml; fluanisone, final concentration 2.5 mg/ml) and 50 μl hypnoval (Roche; midazolam, final concentration 1.25 mg/ml) via intraperitoneal injection. Microinjected eggs, at 1 cell stage, were

transferred through the infundibulum to the oviducts. 20-30 eggs were transferred per pseudopregnant female.

7 Transgenic mouse analysis

7.1 Analysis of transgenic mouse embryos

7.1.1 Collection of transient transgenic embryos.

Unless otherwise stated, mice were collected at the E14.5 stage of gestation (14.5 dpc). The embryonic tails were removed into separate sterile tubes and the yolk sac discarded. This was done in ice cold PBS. DNA was extracted from the embryonic tails using the protocol 1.2.

7.1.2 Genotyping by PCR.

For detection of transgenics harbouring DNA from the human aggrecan lacZ constructs, the polymerase chain reaction was used with Taq polymerase (Methods section 2.1).

The primers used were:

Reverse primer: 5' TCA CGT TCT TGT TGG TCA AAG 3'

Forward primer: 5' CGG CAG GTA CCG AGC TCT T 3'

7.1.3 X-gal staining of embryos

To assay for the expression of the β -galactosidase gene, the embryos were stained with X-gal (5-bromo-4-chloro-3-indolyl- β -d-galactopyranoside). To do this, the embryos were put in 5 ml fixative solution (see Methods section 9.4) for 15 minutes, and then rinsed in 5 ml of rinse solution (see Methods section 9.4) for 3X30 minutes at room temperature. The rinse was finally replaced with 5 ml of staining X-gal solution (see Methods section 9.4) and the embryos rotated overnight in the dark at room temperature.

7.1.4 Histological analysis

Embryos that expressed the lacZ gene were embedded in paraffin wax and sectioned for histological analysis. The embryos were removed from the staining solution, washed in rinse

solution and put in 5 ml fixative for two hours. This was then replaced sequentially with 5 ml of 30% ethanol, 50% ethanol, and 70% ethanol, each for 30 minutes. Once in 70% ethanol, the embryos could be stored at -20°C. Before embedding in wax, they were further dehydrated sequentially in 80% ethanol for 30 minutes, 90% ethanol for 30 minutes, and 100% ethanol 3 times for 30 minutes. The embryos were then put in 5 ml of 50% ethanol/ 50% xylene for 15 minutes, then 100% xylene until the embryos sunk. Finally, they were removed and placed in molten paraffin wax at 65 degrees Celcius for 4-8 hours. They were then transferred to a plastic mould containing paraffin wax and left to set. Unless otherwise stated, all steps were performed at room temperature. The wax blocks were then cut by David Essex (Charing Cross Hospital) at 6 µm, counterstained with eosin, Haematoxinilin or both and examined under a microscope.

7.2 Analysis of adult transgenic mouse lines

Throughout my PhD, I helped generate and was responsible for two mouse lines which contained the following transgenes: β AGG20 and ColII4xA1Cre (see constructs 12 + 17, Methods section 5).

7.2.1 Analysis of β AGG20 line

7.2.1.1 Genotyping using Real-time PCR

‘JumpStart™ Taq ReadyMix™ for Quantitative PCR’ (Sigma D7440) was used for genotyping the β AGG20 line. Before setting up the PCR reaction, a mix of the two sets of primers and fluorescent probes to be used was prepared (each at 10 µM) and stored at -20°C. Each reaction was carried out at a final volume of 10 µl as follows: 5 µl of 2x PCR mix, 1.6 µl of 25 mM MgCl₂ (final concentration of 5.5 mM), 100-200 ng of template DNA extracted from adult tails, 0.12 µl of Primer/probe mix (final concentration of 20 nM for each) and 2.28 µl of water. The PCR reaction was carried out as follows: Initial 94°C step for 10 minutes. Then, 40 repeat

cycles of 94°C for 30 seconds followed by 60°C for 60 seconds. The primers used were for *LacZ* as well as 18S (for normalization). The *LacZ* probe was designed with the fluorescein 6-*FAM* (green) at its 5' end while the 18S probe was designed with *HEX* (pink) at its 5' end. Both *LacZ* and 18S fluorescent probes were designed with the universal fluorescence quencher *BHQ1* at their 3' ends.

Sequences of Primers and probes:

LacZ Probe: 5'*FAM*TCAGCCGGAAAACCTACCGGATTGA 3'*BHQ1*

LacZ forward primer:

GTGCACGGCAGATACACTTG

LacZ Reverse:

AACGGTAATCGCCATTTGACCAC

18S Probe:

5'*HEX*CTTCCGGGAAACCAAAGTCT 3'*BHQ1*

18S Forward:

GACCATAAACGATGCCGACTG

18S Reverse:

CCCTTCCGTCAATTCCTTTAAG

7.2.1.2 X-gal staining of β AGG20 adult mice

To assay for the expression of the β -galactosidase gene in adult mice from the β AGG20 line, mice of different ages were culled and dissected to isolate different tissues to be stained with X-gal. To do this, the dissected tissue such as knees, hips, digits or spine were put in 50 ml fixative solution for 2 hours, and then rinsed in 50 ml of rinse solution for 3X30 minutes at room temperature. The rinse was replaced with 50 ml of staining X-gal solution and the tissue rotated overnight in the dark at room temperature. See methods section 9.4 for preparation of fixative, rinse and X-gal solutions. The same was done for the knees of β AGG20 mice that had undergone DMM or axial loading (see methods section 8),

7.2.1.3 Histological analysis of X-gal stained tissue

Dissected tissue that expressed the *LacZ* gene was removed from the staining solution, washed in rinse solution and fixed in 10% formalin. They were then decalcified in dilute formic

acid, and embedded in paraffin wax to be sectioned for histological analysis. The decalcification, embedding and cutting was done by David Essex (Charing Cross Hospital). The wax blocks were cut at 6 μm , counterstained with eosin, Haematoxylin or both and examined under a microscope.

7.2.1.4 Immunohistochemical analysis

7.2.1.4.1 BR1 antibody staining

BR1, generated by Professor Tim Hardingham and donated by Alan Murdoch (University of Manchester), is a polyclonal rabbit anti-pig antibody raised against the G1 domain of aggrecan. It was used to determine endogenous levels of aggrecan in the transgenic line 'BAGG20'. Dako EnVision+ System-HRP (DAB) for use with rabbit primary antibodies was used to visualise the binding of BR1.

Slides with wildtype embryonic paraffin sections at different ages or with paraffin sections of wildtype/ β AGG20 transgenic tail tips were dehydrated overnight in a 37°C oven, deparafinised in xylene baths for 2x5 minutes, and rehydrated in 2x absolute ethanol baths followed by 1x 90% and 1x 70% ethanol baths (changing baths every 5 minutes and tapping off any excess liquid after each step). The slides were then placed in distilled water for a minimum of 30 seconds and washed 3x in PBS for 5 minutes each. Bovine testicular hyaluronidase (1000 U/ml made up in PBS) was added to the slides for 1 hour at 37°C for antigen retrieval. They were then rinsed 3x in PBS for 5 minutes each and incubated for 10 minutes with a peroxidase block to quench any endogenous peroxidase activity the sample may have. After rinsing the slides again 3x in PBS, they were incubated in 1% non-fat milk made up in PBS for 1 hour at room temperature. The milk was tapped off and the slides incubated with BR1 antibody (1:1000 made up in Dako Antibody Diluent) at 4°C overnight. The next day, the slides were rinsed 3x in PBS for 5 minutes each and incubated for 30 minutes with the secondary antibody

(peroxidase labelled polymer conjugated to goat anti-rabbit immunoglobulins). The secondary antibody was washed off 3x in PBS for 5 minutes each and the peroxidase substrate added (DAB+ chromogen made up in DAB+ substrate buffer) for 30 seconds- 5 minutes. Slides were then rinsed in running tap water for 10 minutes and counterstained with Harris haematoxylin for 3 minutes at room temperature. They were then rinsed again in running tap water for 10 minutes and placed in 70% ethanol for 2 minutes, 95% ethanol for 2 minutes, 2x100% ethanol for 2 minutes each and finally 2x Xylene for 5 minutes each. They were mounted by coverslipping with histomount solution (National Diagnostics, HS103).

7.2.1.4.2 β -galactosidase antibody staining

The β -galactosidase antibody staining was performed by Dr Phuong Vo at the Kennedy Institute of Rheumatology. The antibody used was a polyclonal rabbit anti- β -galactosidase antibody (Abcam # Ab616). It was used as a more sensitive method of detection of β -galactosidase protein (indicating *LacZ* reporter gene expression). Dako EnVision+ System-HRP (DAB) for use with rabbit primary antibodies was used to visualise the binding of the anti- β -galactosidase antibody.

Slides with coronal paraffin sections of axial loaded knees/contralateral knees from β AGG20 transgenic mice were dehydrated overnight in a 37°C oven, deparafinised in xylene baths for 2x5 minutes, and rehydrated in 2x absolute ethanol baths followed by 1x 90% and 1x 70% ethanol baths (changing baths every 5 minutes and tapping off any excess liquid after each step). The slides were then placed in 2 consecutive baths of distilled water each for 5 minutes. For antigen retrieval, the slides were boiled within Sodium citrate buffer (pH 6.0) in a pressure cooker for 10 minutes. They were then allowed to cool on a bench for 30 minutes. Following this, they were washed 2x in PBS for 5 minutes each and incubated for 10 minutes with a peroxidase block to quench any endogenous peroxidase activity the sample may have. After

rinsing the slides again 2x in PBS for 5 minutes each, they were incubated in 1% non-fat milk made up in PBS for 1 hour at room temperature. The milk was tapped off and the slides incubated with the anti- β -galactosidase antibody (1:1000 made up in 1% non-fat milk) for 1 hour at room temperature. The slides were then rinsed 2x PBS for 5 minutes each with agitation (0.01% Tween-20 optional) and incubated for 30 minutes with the secondary antibody (peroxidase labelled polymer conjugated to goat anti-rabbit immunoglobulins). The secondary antibody was washed off 2x in PBS for 5 minutes each with agitation and the peroxidase substrate added (DAB+ chromogen made up in DAB+ substrate buffer) for 30 seconds- 5 minutes. Slides were then rinsed in running tap water for 10 minutes, placed in 70% ethanol for 2 minutes, 95% ethanol for 2 minutes, 2x100% ethanol for 2 minutes each and finally 2x Xylene for 5 minutes each. They were mounted by coverslipping with histomount solution (National Diagnostics, HS103).

7.2.2 Analysis of Col14xA1Cre line

7.2.2.1 Genotyping using Real-time PCR

'JumpStart™ Taq ReadyMix™ for Quantitative PCR' (Sigma D7440) was used for genotyping the Col14xA1Cre line. Before setting up the PCR reaction, a mix of the two sets of primers and fluorescent probes to be used was prepared (each at 10 μ M) and stored at -20°C. Each reaction was carried out at a final volume of 10 μ l as follows: 5 μ l of 2xPCR mix, 1.6 μ l of 25 mM MgCl₂ (final concentration of 5.5 mM), 100-200 ng of template DNA extracted from adult tails, 0.12 μ l of Primer/probe mix (final concentration of 20 nM for each) and 2.28 μ l of water. The PCR reaction was carried out as follows: Initial 94°C step for 10 minutes. Then, 40 repeat cycles of 94°C for 30 seconds followed by 60°C for 60 seconds. The primers used were Cre as well as 18S primers (for normalization). The Cre probe was designed with the fluorescein 6-*FAM* (green) at its 5' end while the 18S probe was designed with *HEX* (pink) at its 5' end. Both Cre and 18S fluorescent probes were designed with the universal fluorescence quencher

BHQ1 at their 3' ends.

Sequences of Primers and probes:

Cre Probe: 5'*6FAM*AAACATGCTTCATCGTCGGTCCGG3'*BHQ1*

Cre forward primer:

GCGGTCTGGCAGTAAAACTATC

Cre Reverse primer:

GTGAAACAGCATTGCTGTCACTT

18S Probe:

5'*HEX*CTTCCGGGAAACCAAAGTCT 3'*BHQ1*

18S Forward primer:

GACCATAAACGATGCCGACTG

18S Reverse primer:

CCCTTCCGTCAATTCCTTTAAG

7.2.2.2 Luciferase detection

Han Lim Ngee from the Kennedy Institute of Rheumatology assisted with all the *in vivo* bioluminescent imaging carried out. To assay for the expression of the Luciferase gene in adult mice from the Coll4xA1Cre line, *In vivo* bioluminescent imaging was performed in a Kodak imaging system that includes an integrated isoflurane apparatus and induction chamber for temporary anaesthesia of the mice. First, the mice are placed in the induction chamber and anesthetized with isoflurane/oxygen. After a few minutes, the anesthetized mice were removed from the induction chamber and placed on a stage with continuous isoflurane delivery. Because fur naturally autofluoresces and can also block the emission of light, the legs and knees of the mice were shaved with an electric shaver. The mice were then injected with the substrate D-luciferin (Sigma) into the intraperitoneal cavity at a dose of 150 mg/kg body weight (30 mg/ml D-luciferin) approximately 5 min before imaging. To carry out the bioluminescence imaging, the mice were placed dorsal side up in the imaging chamber with isoflurane. Mice were exposed for 10 minutes after which dorsal images were collected. Mice were then removed from the imaging chambers and allowed to wake up in their cages.

7.2.2.3 Assaying for Cre recombinase activity

To characterize the spatial and temporal pattern of Cre expression and activity in the ColII4xA1Cre line, the double-fluorescent Cre Reporter mouse (described in more detail in Results Chapter 2 section 2.1) was used. A heterozygous male ColII4xA1Cre mouse was crossed with 2 homozygous female double-fluorescent reporter mice. The pregnant mothers were injected intraperitoneally with 4x5 mg of tamoxifen (sigma) dissolved in corn oil on the last 4 days of gestation. Pups from one of the pregnant mothers were sacrificed at 1 week of age while those from the 2nd pregnant mother were sacrificed at 2 weeks of age. The hips and knees of the 1 and 2-week-old mice pups were snap-frozen by embedding them into the bottom of a cork mould pre-filled with OCT (Tissue-Tek) and placing them using forceps on an isopentane-filled beaker that has been put in liquid nitrogen. The cryo-preserved tissues were then sectioned (6 µm sections) using a CV1900 cryostat (Leica Microsystems) equipped with a tungsten blade. The sectioned tissue was fixed for 10 minutes with 4% paraformaldehyde, washed 3x with PBS, then mounted by coverslipping with 80% glycerol and viewed using Ultraview confocal microscopy (60 oil immersion lens).

8 Osteoarthritis models

8.1 Surgical induction of OA through the destabilisation of the medial meniscus (DMM) model

Mice were anaesthetized by intraperitoneal injection of Hypnorm® (VetaPharma Ltd., Leeds, UK) and Hypnovel® (Roche). After shaving of the ventral portion of the right knee, a midline incision was made over the ventral aspect and the medial compartment of the knee. The medial meniscus was identified, and using an ophthalmic scalpel, its anterior horn was released by sharp dissection of the attachments to the tibial plateau. Mobility of the anterior half of the medial meniscus was confirmed by displacement with forceps. The parapatellar window was closed with an absorbable suture (Ethicon, USA), and the skin closed with fine non-absorbable sutures (Ethicon, USA). A technician at the Kennedy Institute of Rheumatology performed the DMM surgeries.

8.1.1 DMM of line 12 (β AGG20 transgenics)

DMM was performed on the right knees of twelve 'line 12' heterozygote mice of both sexes at 8 weeks of age. The left contralateral knees were used as controls. 6 mice were sacrificed at 4 weeks and a further 6 at 8 weeks post-operatively and the operated + control unoperated knees dissected, fixed, stained with X-gal as described in results section 7.2.1.2. Dissected knees specimens were removed from the staining solution, washed in rinse solution and fixed in 10% formalin. They were then decalcified for 3 weeks in EDTA, and embedded in paraffin wax to be sectioned for histological analysis. The decalcification, embedding and cutting was done by David Essex (Charing Cross Hospital). The wax blocks were cut into coronal sections (8 μ m thick) and stained with Safranin O and/or haematoxylin and Eosin. Sections were evaluated across 10 levels at 80 μ m intervals. Each knee yielded approximately 10 slides for scoring qualitatively based on The OARSI Osteoarthritis Cartilage Histopathology scoring

system which itself is based on both the severity (grade) and extent (stage) of OA in the articular cartilage (Pritzker et al., 2006). The qualitative score given indicates the most advanced OA grade/stage observed in the section, e.g., grade 2, stage 3.

8.1.2 DMM of line 39 (Col14x1Cre transgenics)

DMM was performed on the right knees of 5 heterozygote transgenic mice from line 39 as well as 1 wildtype littermate (control) at 8 weeks of age. Both sexes were used. The left contralateral knees were also used as controls. Mice were analysed before surgery (0 weeks) and 1,2,3,4 and 8 weeks post-operatively by *in vivo* bioluminescence imaging as described in Results section 7.2.2.2.

8.2 Axial Loading of murine tibias

The loading experiments were performed in Dr Andrew A. Pitsillides' lab at the Royal Veterinary College (De Souza et al., 2005).

The loading apparatus used was constructed from a single pair of custom-built padded cups; the top cup demonstrates a concavity that is equivalent to an imprint of the flexed knee, acting to restrict forward slippage of the knee from the cup. The bottom cup, however, holds the ankle and is flexed at approximately 45° (Figure 12A and B). The cups are positioned within a servo-hydraulic materials testing machine (Model HC10, Dartec, Ltd., Stourbridge, UK) and are aligned vertically. The upper cup into which the knee is positioned was connected to the actuator, the lower cup found on the 'load cell', and the actuator operated in 'load control' to administer dynamic compression to the tibia (Figure. 12B).

The right tibia of each loaded mouse (left acting as contralateral control) was held in the padded cups, and dynamic (2 Hz, with 10 s rest periods between cycles) axial loads were

applied (peak load 2–13 N, strain rate at the gauge site 0.0015–0.002 $\mu\epsilon/s$, 28–44 N/s, 40 cycles/day) through the knee joint three times a week for 2 weeks (a total of 6 times).

8.2.1 Axial loading of line 12 (β AGG20 transgenics)

Axial loading (Figure 12) was performed on the right knees of 12 mice of both sexes at 8 weeks of age. 2 weeks after a total of 6 axial loads, the mice were sacrificed and the knees dissected, fixed, stained with X-gal as described in results section 7.2.1.2. Dissected knees specimens were removed from the staining solution, washed in rinse solution and fixed in 10% formalin. They were then decalcified for 3 weeks in EDTA, and embedded in paraffin wax to be sectioned for histological analysis. The decalcification, embedding and cutting was done by David Essex (Charing Cross Hospital). The wax blocks were cut into coronal sections (8 μm thick) and stained with Eosin or a β -galactosidase antibody. Sections were evaluated across 10 levels at 80 μm intervals. Each knee yielded approximately 10 slides for scoring qualitatively based on The OARSI Osteoarthritis Cartilage Histopathology scoring system which itself is based on both the severity (grade) and extent (stage) of OA in the articular cartilage (Pritzker et al., 2006). The qualitative score given indicates the most advanced OA grade/stage observed in the section, e.g., grade 2, stage 3.

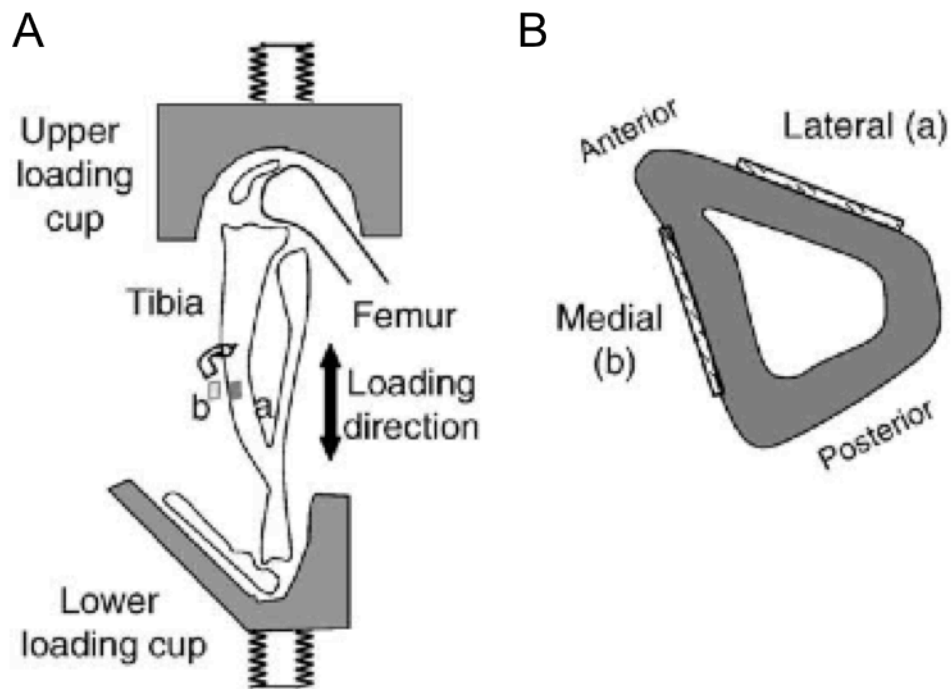


Figure 12: Mechanism of axial loading. A) Diagrammatic representation of the flexed mouse right hind limb in position in the loading apparatus, showing the relative position of the bones, their relationship with the upper and lower loading cups, the direction of loading and approximate location of the strain gauges (a, lateral and b, medial) on the tibial mid-shaft. B) Diagrammatic representation of cross-section of the tibial mid-shaft showing the location of the attached strain gauges on medial and lateral surfaces of the tibia. From (De Souza et al., 2005).

9 Stock solutions

All chemicals were obtained from Sigma unless otherwise stated.

9.1 Standard solutions

TE buffer	10mM Tris/HCL (pH 7.4), 1mM EDTA (pH 8)
PBS	Per litre: 10g NaCl, 0.25g KCl, 1.44g Na ₂ HPO ₄ , 0.25g KH ₂ PO ₄
50x TAE buffer	Per litre: 242g Tris, 100ml 0.5M EDTA (pH 8.0), 57.1ml Glacial Acetic Acid, in water (final pH 7.6-7.8)

9.2 Bacterial culture and transformations

LB-broth with ampicillin	Per litre: 10g Tryptone (pancreatic digest of casein), 5g yeast extract, 5g NaCl, 100µg/ml ampicillin
LB-agar with ampicillin	Per litre: 10g Tryptone (pancreatic digest of casein), 5g yeast extract, 5g NaCl, 15g Agar, 100µg/ml ampicillin

9.3 DNA preparation

Resuspension buffer P1	15mM Tris (pH 8.0), 10mM EDTA (pH 8.0), 100µg/ml RNase A
Lysis buffer P2	0.2M NaOH, 1% (w:v) SDS
Neutralisation buffer P3	3M KOAc (pH 5.5)
DNA loading buffer (6x)	2.5% Ficoll 400, 11mM EDTA, 3.3mM Tris-HCL, 0.017% SDS, 0.015% Bromophenol Blue (pH 8.0)

9.4 Analysis of transgenic embryos

Tail buffer	0.1 M EDTA, 0.1 M NaCl, 0.05 M tris/HCl (pH 8.0).
FiX	2% formaldehyde, 0.2% glutaraldehyde, 0.1 M NaH ₂ PO ₄ , 0.1 M Na ₂ HPO ₄ , 5 mM EGTA (pH 8.0), 2 mM MgCl ₂ , in water
Rinse	2mM MgCl ₂ , 0.1% (w/v) Sodium deoxycholate, 0.2% (v/v) Igepal CA-630 in PBS
Stain	Per 100ml: 4ml of X-gal (dissolved at 25mg/ml in dimethylformamide), mg K ₄ Fe(CN) ₆ , mg K ₃ Fe(CN) ₆ in rinse solution.

Results

I have divided the 'Results' section into three chapters. Each chapter contains a brief overview of the experimental work (background), a report of the data itself (results) and finally a discussion of the data evaluating the findings in context of the published literature (discussion).

Chapter 1: Locating *cis*-acting sequences that regulate the cartilage-specific expression of the human aggrecan gene:

1.1 Background

The main aims of this result chapter are both to identify the human DNA sequence(s) necessary to confer full endogenous aggrecan expression, and to characterize the minimal sequence necessary for this level of expression. This was investigated using transgenic mice. Several combinations of linear DNA fragments were fused to the human aggrecan minimal promoter and the *LacZ* reporter gene (methods section 5). For some of these constructs, I adopted a method based on gap repair homologous recombination in *E.coli* called "Recombineering".

Recombineering uses recombination proteins provided from λ phage. The artificially engineered recombineering bacterial strain (EL350) contains a defective λ prophage inserted into the bacterial genome. The phage genes of interest, *exo*, *bet*, and *gam*, which are involved in bacterial recombination, are transcribed from the λ PL promoter. This promoter is repressed by the temperature-sensitive repressor *cI857* at 32°C and derepressed at 42°C. When bacteria containing this prophage are kept at 32°C no recombination proteins are produced. However, after a brief (15 minutes) heat-shock at 42°C a sufficient amount of recombination proteins are

produced. *exo* is a 5'-3' exonuclease that creates single-stranded overhangs on introduced linear DNA. *bet* protects these overhangs and assists in the subsequent recombination process. *gam* prevents degradation of linear DNA by inhibiting *E.coli* RecBCD protein. Linear DNA (PCR product, oligo, etc.) with sufficient homology in the 5' and 3' ends to a target DNA molecule already present in the bacteria (plasmid, BAC, or the bacterial genome itself) can be introduced into recombination competent EL350 bacteria using electroporation. The introduced DNA will now be modified by *exo* and *bet* and undergo homologous recombination with the target molecule. The method has also been shown to work by co-electroporation of a supercoiled plasmid and a linear piece of DNA into heat-shocked, electrocompetent bacteria (Liu et al., 2003).

The linearised constructs were injected into fertilized mouse oocytes to generate mice in which reporter gene expression indicated the presence of *cis*-regulatory element(s) within the DNA upstream of the reporter. The pattern of reporter gene expression in these transgenic mice, as demonstrated by X-gal staining, was used to elucidate the function of the construct injected. Since a minimal promoter of human *AGC1* has been identified (Valhmu et al., 1998a), transgenic mice were also generated with a construct consisting of the minimal promoter alone fused to the *LacZ* gene. The level and pattern of reporter gene expression seen in these transgenic mice provided a baseline, to which the pattern and extent of expression driven by other combinations of fragments fused to the minimal promoter was compared.

The *LacZ* gene is a bacterial reporter gene that encodes the enzyme β -galactosidase (β -gal). β -galactosidase catalyses the cleavage of galactose rings from a number of different substrates including X-gal. X-gal changes from a colourless compound to a deep blue one following this cleavage. In the transgenic mice that were generated here, this dark blue staining

was indicative of the regulatory activity and expression pattern exerted by the DNA sequences driving this reporter gene.

As the main purpose of this study was to screen for functional *cis*-regulatory elements, and since the insertion site of the transgene and the number of copies inserted could affect the expression pattern, founder embryos were used as they are more practical than establishing transgenic lines. After injection, the oocytes were transplanted into a surrogate mother, and the developing embryos removed before birth. In this way, a number of embryos - each containing different copies of the transgene integrated at different sites in the genome – were examined and the variation of expression due to these factors could be taken into account. Doing this for established transgenic lines would have been a much more lengthy and costly approach and beyond the scope of the work presented here. However, the advantage of this would be the ability to look at the same transgene integration and copy number at different stages of development. This was done for one of the constructed transgenes.

Founder embryos were removed at E14.5, unless otherwise stated. The embryos weren't normally removed after E16.5 due to the keratinisation of the skin, which occurs around then. This results in the inability of the X-gal solution to penetrate, thus making consistent staining of the embryos impossible.

Within embryos of the same litter, and between E14.5 embryos from different litters and different experiments, it is important to note that they may have slight developmental differences as they normally develop at different rates (most notable in the extent of bone growth and ossification). These developmental differences as well as the variations in transgene copy number and integration sites can account for some of the differences seen between embryos expressing the same construct. It is important to note, however, that

embryonic staining is due to expression of the *LacZ* transgene rather than an artefact due to bacterial *LacZ* expression and/or metallic contact with the embryo if the staining is bilateral and symmetrical within the embryo

The differentiating chondrocytes responsible for the synthesis of aggrecan during development are described in the main Introduction section 2.2. Since aggrecan is mainly synthesised in developing cartilage, the expression of the transgene was particularly investigated in the developing axial (skull, ossicles of the inner ear, the hyoid bone of the throat, the rib cage and the vertebral column) and appendicular (upper and lower limbs) skeleton. Ossification in the mouse embryos commences at around E13, when both mesenchymal and pre-cartilagenous condensations are visible. At this time, chondrification of the vertebral column begins, proceeding in rostral to caudal direction. As discussed, the condensation and chondrification of skeletal elements culminate in the differentiation of chondrocytes and in this way form a cartilage model of the future axial skeleton. However, by E14-15.5, ossification also becomes visible in the clavicle, ribs, femur and humerus, ulna and radius, maxilla, mandible, frontal and parietal bones of the head, and scapula, hip bones, tibia and fibula of the body (Kaufmann, 1995). It is important to note that the half-life of the β -galactosidase protein is 24-48 hours and that any staining seen within a particular tissue could be due to the expression of this reporter gene within 24-48 hours prior to assaying for regulatory activity of the transgene.

1.2 Results

1.2.1 The use of the BR1 antibody to monitor endogenous expression of the aggrecan gene in mouse embryos

Before attempting to pursue the *cis*-regulatory regions of the human aggrecan gene that govern its endogenous expression in developing embryos, I needed to establish a reliable pattern of endogenous aggrecan gene expression in embryos at different stages of development. I optimised the use of the BR1 antibody, which was raised against the G1 domain of aggrecan (see Methods 7.2.1.4.1), on paraffin embedded sections of mouse embryos at E11.5, E13.5, E15.5 and E17.5. At E11.5, antibody staining is localised to the somites, which will eventually become rib and vertebra cartilage (Figure 13A). At E13.5, E15.5 and E17.5, the BR1 staining is localised to the developing axial and appendicular embryonic skeleton. At E13.5, there is antibody staining in several developing cartilaginous structures such as the ear, the nasal cartilage, the Meckel's cartilage, the humerus, the vertebrae, the ribs, the digits of the forelimbs and the femur (Figure 13B). There was also staining in the digits of the hindlimb and the ankle in other serial sections (not shown). At E15.5, the staining of the antibody was found in all the positively stained anatomical regions at E13.5 with the addition of the chondrocranium (including the frontal bone) as well as the developing scapular cartilage (Figure 13C). At E17.5, in addition to the developing axial and appendicular skeleton (Figure 13D+F+G), the tracheal cartilage showed positive staining (Figure 13E).

1.2.2 *In silico* analysis of aggrecan gene identity between rats and humans

In order to begin my search for the key areas in the human aggrecan gene that are responsible for chondrocyte specific expression, I used a web-based program, vista (<http://pipeline.lbl.gov/cgi-bin/gateway2>), to look at sequences that have high identity between species indicating evolutionary conserved and therefore important regions within the gene. My

search included ~20kb of non-coding sequences upstream of the aggrecan gene in order to take into account any shifting of enhancer elements that may have occurred during evolution.

Figure 14 demonstrates the homologous sequences found between humans and rats throughout the non-coding sequences ranging between -20kb relative to the transcriptional start site (TSS) up to and including the 3'UTR of the gene. Taking into account this homology search, the information that had already been published on the rat upstream enhancer (Doege et al., 2002) as well as the knowledge that *sox9* plays an important role in the development of cartilage and is therefore important for the regulation of aggrecan ((Bi et al., 2001), (Kist et al., 2002), (Sekiya et al., 2000), (Han and Lefebvre, 2008)), I hypothesised that the regions circled 1-7 (named *cis* 1-7) in Figure 14 could potentially be *cis*-regulatory sequences. *cis* 1-4 lie upstream of the start site and are in the approximate positions '-20 to -19.7kb', '-14.5 to -14.4kb', '-10.9 to -10.2kb', and '-3.8 to -3.7kb' respectively. *Cis* 5 lies around/includes the 1st exon/5'UTR/previously determined basal promoter of the human aggrecan gene (-390 to +375bp). Unlike *cis* 1-5, which correspond to red coloured peaks (Figure 14) indicating highly conserved regions, *cis* 6 is not coloured but was selected due to the presence of 3 putative *sox9* binding sites within it. It is located at the 5' end of the large 1st intron of the gene (between +5 and +8kb relative to the TSS). The *sox9* putative binding sites were identified using the program genomatix (<http://www.genomatix.de/>) and are located at +5.3kb, +6.8kb and +7.4kb relative to the TSS. Finally, *cis* 7 is positioned within the 14th intron of the human aggrecan gene between +64.6kb and +65.0kb relative to the TSS. My interest in *cis* 7 lies in the fact that it's the only sequence within the investigated region in which the homology even to fugu remains relatively high. Figure 15 demonstrates the identity between humans, rats, chickens and fugu in this region.

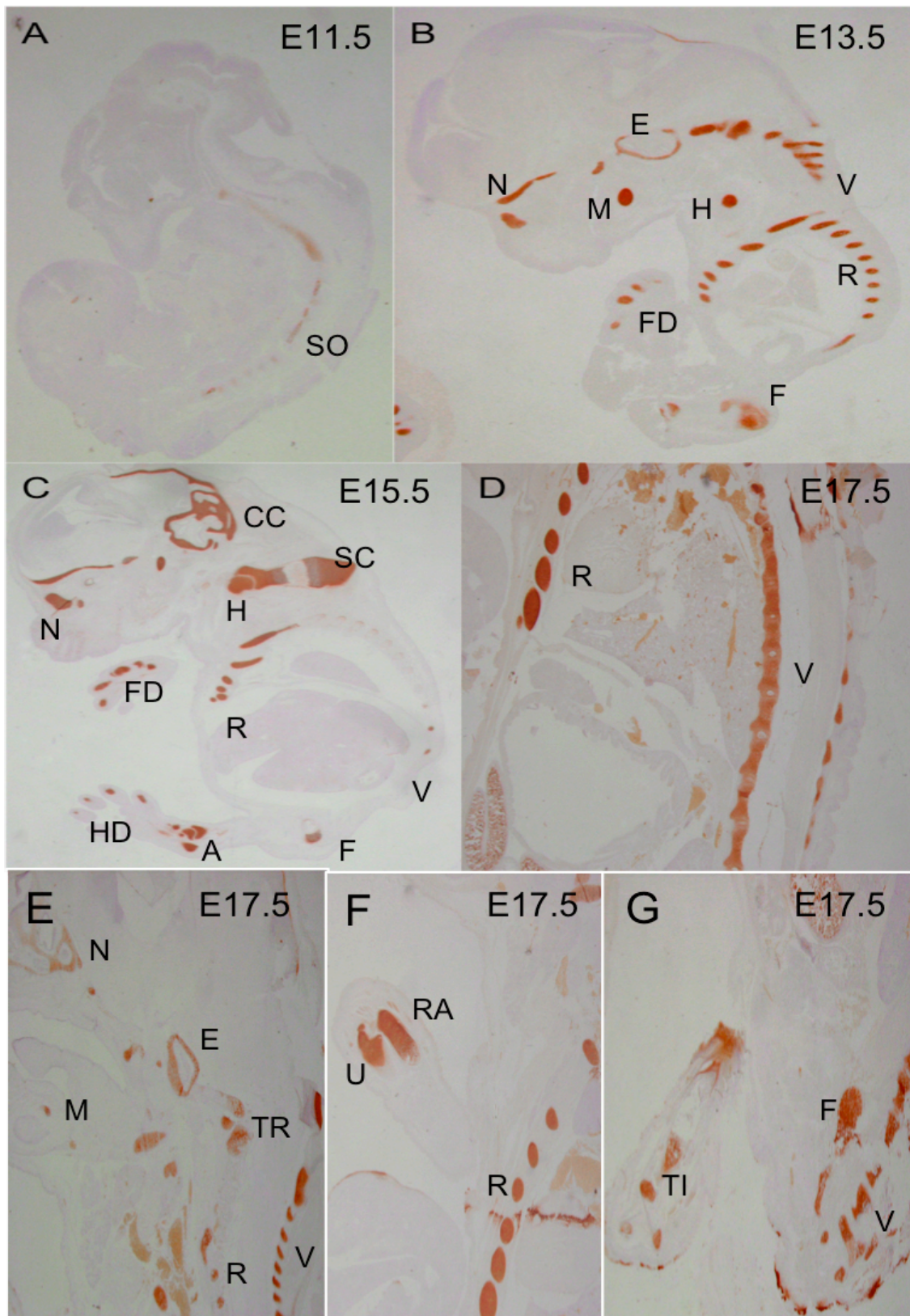


Figure 13: Paraffin sections of mouse embryos at different stages of development stained with the BR1 antibody. A) E11.5 embryo B) E13.5 embryo C) E15.5 embryo D) E17.5 embryo. A (ankle), CC (chondrocranium), E (ear cartilage), F (femur), FD (forelimb digits), H (humerus), HD (hindlimb digits), M (Meckel's cartilage), N (nasal cartilage), RA (radius), R (rib cartilage), SC (scapular cartilage), SO (somites), TI (tibia), TR (trachea), V (vertebral cartilage).

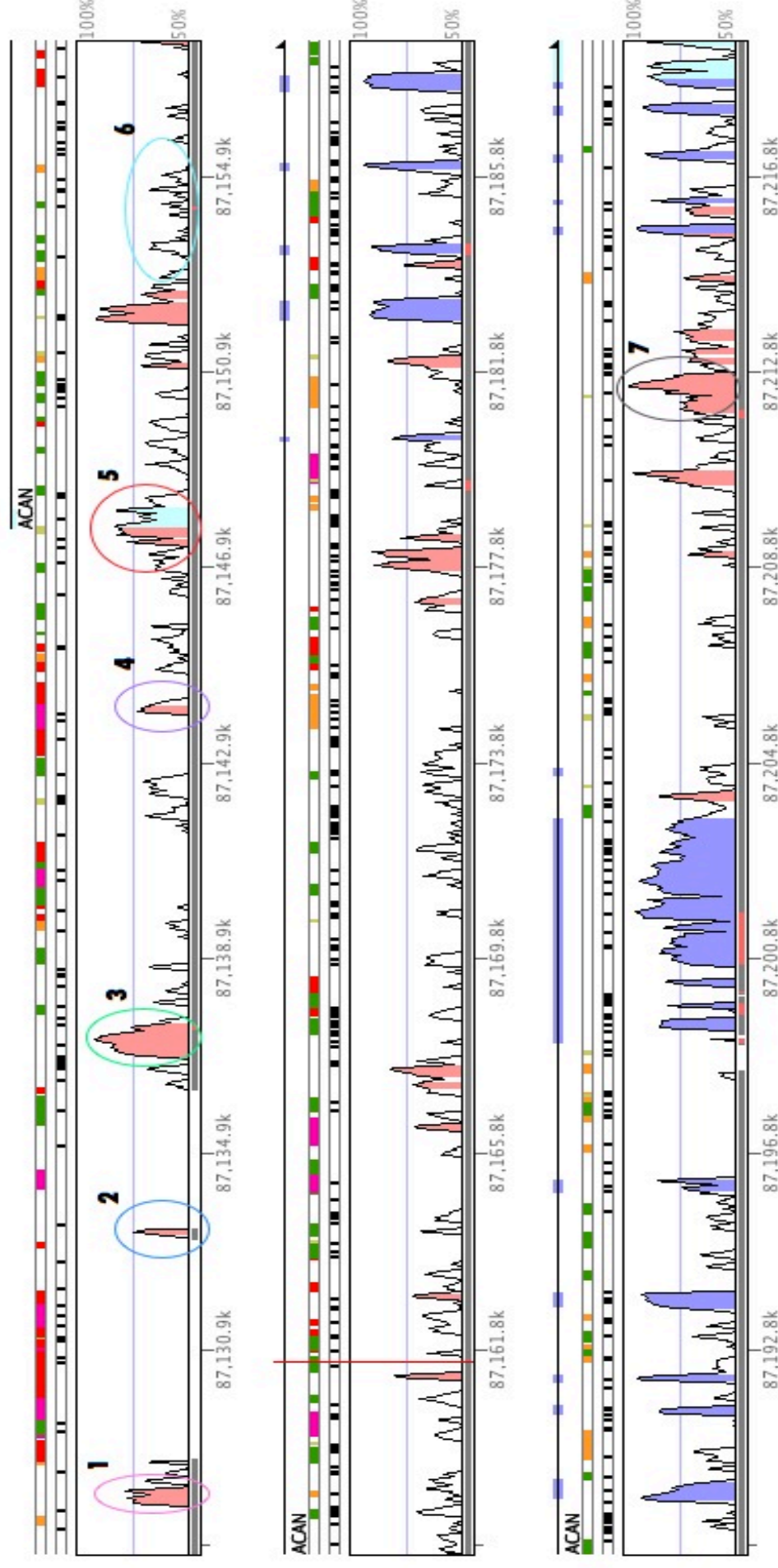


Figure 14: Schematic vista analysis of the aggrecan gene including ~20kb of non-coding DNA upstream of the TSS (-20kb - +72kb relative to the TSS). The base DNA here is human DNA, which is aligned with rat DNA. Conserved regions, that is those greater than 70% identity and 100 or more base pairs long using default settings are shown in pink if they are non-coding regions, blue if they are conserved exons and turquoise if they are untranslated regions. Therefore, the 5'UTR, corresponding to exon1/the basal promoter and the 3'UTR are coloured turquoise. The circled regions 1-7 correspond to DNA sequences called *cis* 1-7. Generated from (<http://pipeline.lbl.gov/cgi-bin/gateway2>)

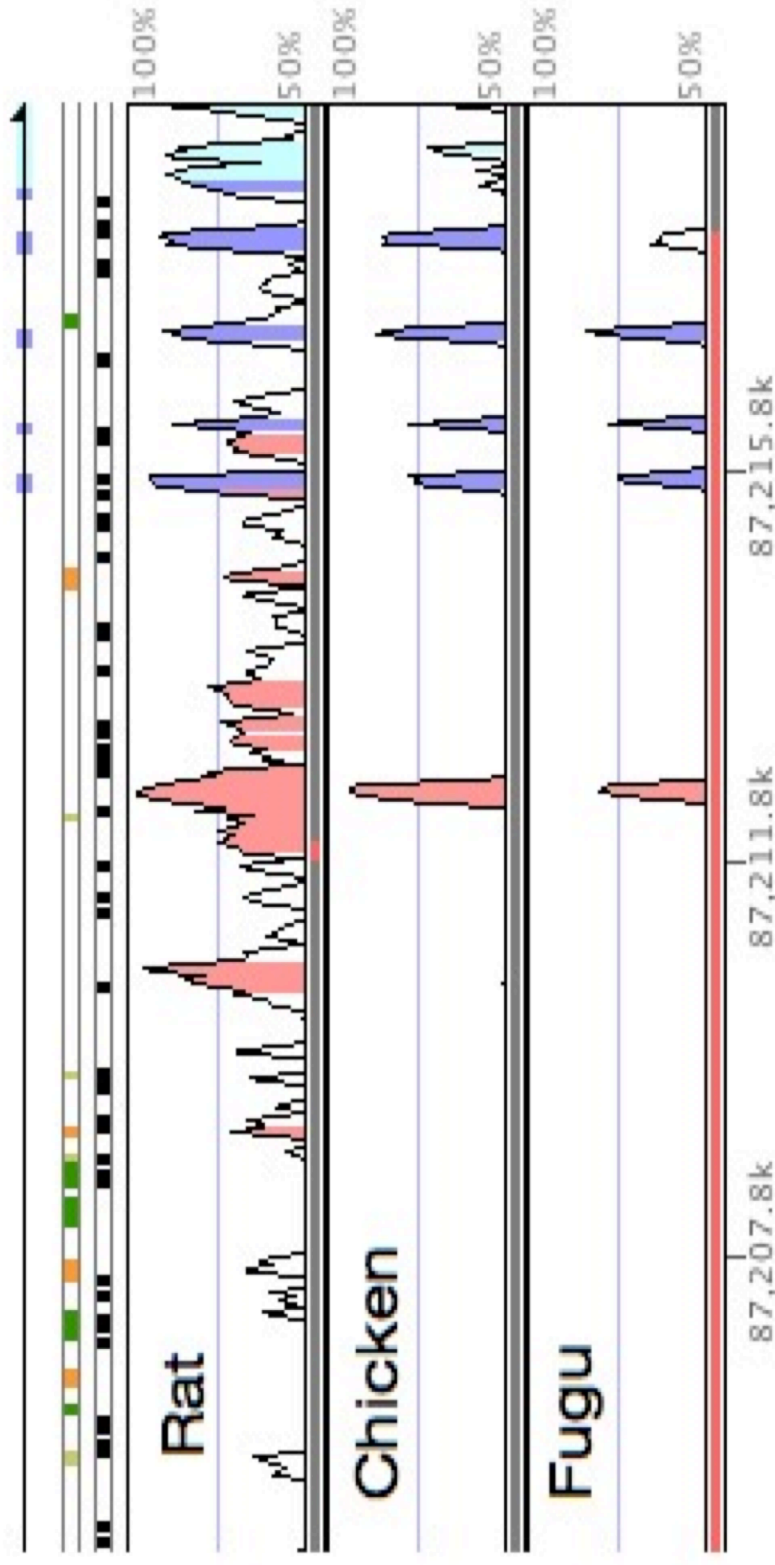


Figure 15: Schematic vista analysis of highly homologous sequence within intron 14 of the human aggrecan gene (*c/s* 7). The base DNA here is human DNA, which is aligned with rat, chicken and fugu DNA. Conserved regions, that is those greater than 70% identity and 100 or more base pairs long using default settings are shown in pink if they are non-coding regions and blue if they are conserved exons. The 3'UTR is coloured turquoise. The red peaks in the human – rat, chicken and fugu alignments correspond to the region *c/s* 7. Generated from (<http://pipeline.lbl.gov/cgi-bin/gateway2>)

1.2.3 The human aggrecan minimal promoter (-452bp to +375bp)

As discussed in the background of this result chapter, the pattern of *LacZ* gene expression in mouse transgenics driven only by the human aggrecan minimal promoter defines the baseline expression that all future staining patterns can be compared to. All other constructs used to create transgenic embryos for the purpose of characterizing the human aggrecan regulatory sequences contain the minimal promoter and *LacZ* gene along with different combinations of DNA sequences containing the circled fragments in Figure 14. Therefore, it is necessary to see the expression driven by the minimal promoter alone.

The human aggrecan minimal promoter used was based on the data published by Valhmu et al (see Introduction section 3.3.3) along with the information from my *in silico* analysis in which *cis* 5 corresponds to this minimal promoter. Although the published data indicates the 5'UTR (0bp to +375bp relative to the TSS) of the gene as the primary regulatory sequence able to act as the minimal promoter required to drive basal expression of the aggrecan gene, the vista analysis shown in Figure 14 showed high sequence identity between humans and rats from -390 to +375bp. Thus, as the human aggrecan minimal promoter, I used the sequence corresponding to -452 to +375bp which was named AB. This ensured the inclusion of the 5'UTR as well as the upstream conserved sequence. AB was cloned in the p β gal-basic vector and the final vector named β galAB (see Methods section 5 construct 1). Transgenic mice were generated harbouring this construct.

Although the 5'UTR was sufficient to drive reporter gene expression *in vitro* (Valhmu et al., 1998a), this wasn't the case for the *in vivo* transgenic embryos that we generated harbouring this sequence. There was no blue staining in any of the positive transgenics. It was thus assumed that any staining seen from now on due to the injection of other constructs was due to sequences that lie outside this minimal promoter interacting with it to drive transcription of

the gene. Table 6 details the number of transgenic and expressing embryos produced with this construct.

Table 6: Summary of mice transgenic and expressing β galAB (minimal promoter/ -452bp to +375bp). 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Positive embryos by PCR	Expressing embryos	Percentage of embryos expressing the transgene (%)
β galAB	10	0	0

1.2.4 The role of the DNA sequences upstream of the human aggrecan minimal promoter

1.2.4.1 *cis* 1-4

In order to examine the functionality of the potential *cis*-regulatory fragments *cis* 1-4 (Figure 14), I cloned a DNA sequence around 20 kilobases in size that contained *cis* 1-4 along with the minimal promoter AB within *cis* 5 into the reporter vector p β gal-basic. To do this, I used 'Recombineering'. This method was used rather than standard DNA restriction cloning because of the size of the DNA fragment to be cloned. Large DNA fragments have few unique restriction sites and are difficult to manipulate.

According to the recombineering method, I cloned a 5' fragment of DNA corresponding to the 5' end of the ~20kb fragment of DNA (named EF) along with a 3' fragment corresponding to the minimal promoter of the aggrecan gene (AB) into the pBluescript KS+ vector to produce pBSAB + EF (Methods section 5 construct 10). This vector was then linearized and

electroporated into recombination competent EL350 cells containing the RP11-343B18 BAC (Figure 16). The resultant vector containing the ~20kb fragment of non-coding aggrecan DNA was called pBSAGG20 (Methods section 5 construct 11).

Once the construction of this new plasmid was confirmed via several restriction digestions, the ~20kb fragment was sub-cloned into the β gal-basic vector upstream of the *LacZ* gene by blunt-ended ligation. This construct was called β AGG20 (Methods section 5 construct 12).

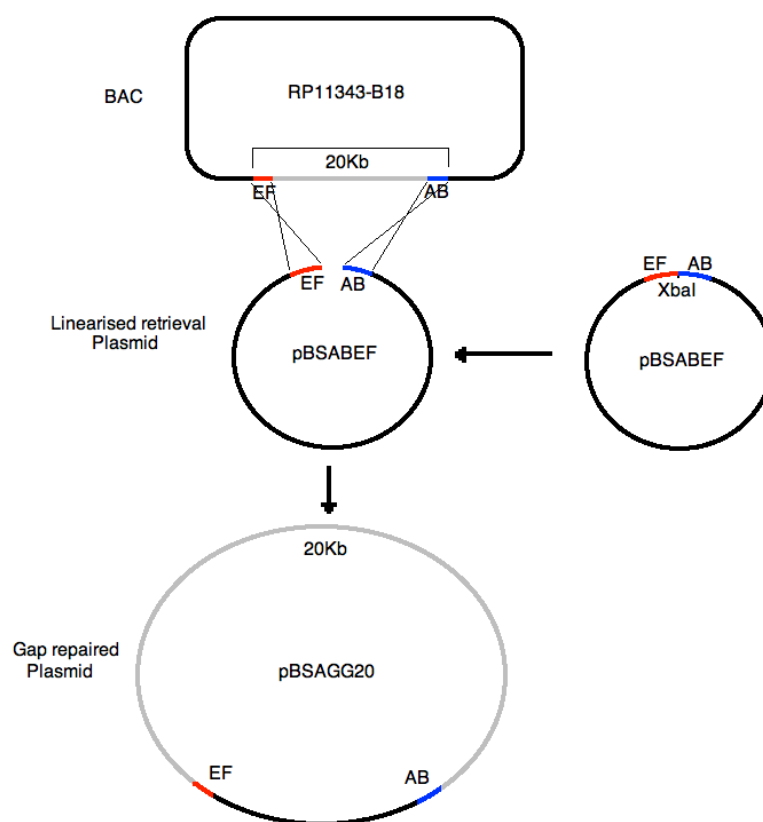


Figure 16: Diagram representing the recombineering strategy used to clone the large ~20kb sequence (-20,054kb to +375bp) from the human aggrecan gene into the pBS KS+ vector. pBSAB + EF, containing the the 5' arm (EF) and 3'arm (AB – corresponding to the minimal promoter) was linearised with *XbaI*.The 5' and 3' arms are coloured in red and blue respectively. The target vector, from which the aggrecan DNA is derived is the BAC RP11-343B18, shown here as a rectangle. The resultant vector containing the ~20kb sequence was called pBSAGG20.

β AGG20 was injected into fertilised oocytes that were in turn transferred into pseudopregnant mothers (Methods section 6). These mothers were sacrificed at E14.5 and the embryos assayed for the expression of the *LacZ* gene. The embryos were evaluated for LacZ expression (blue staining) by whole mount X-gal staining overnight (Methods section 7.1.3). Three out of 7 positive embryos showed identical and symmetrical patterns of reporter gene expression in cartilage (Table 7). Blue staining was observed in the developing cartilage of the long bones of the fore and hind limbs, the ribs, the scapulae, the digits and some craniofacial cartilages (Meckel's cartilage). There was also some staining in the tongue. However, no staining was seen in the vertebrae (Table 8). Figure 17 shows two of the transgenic embryos that expressed the reporter gene exclusively in cartilage. Histological sections of the X-gal blue stained embryos showed that the *LacZ* expression did in fact occur exclusively in the developing cartilage e.g the developing cartilage of the ribs (Figure 18A), the hindlimb digits (Figure 18B), the ulna and radius (Figure 18C) and the femur, tibia and fibula (Figure 18D). The staining was relatively strong in all of these regions and was present in most of the proliferating (smaller cells) and hypertrophic chondrocytes (enlarged cells).

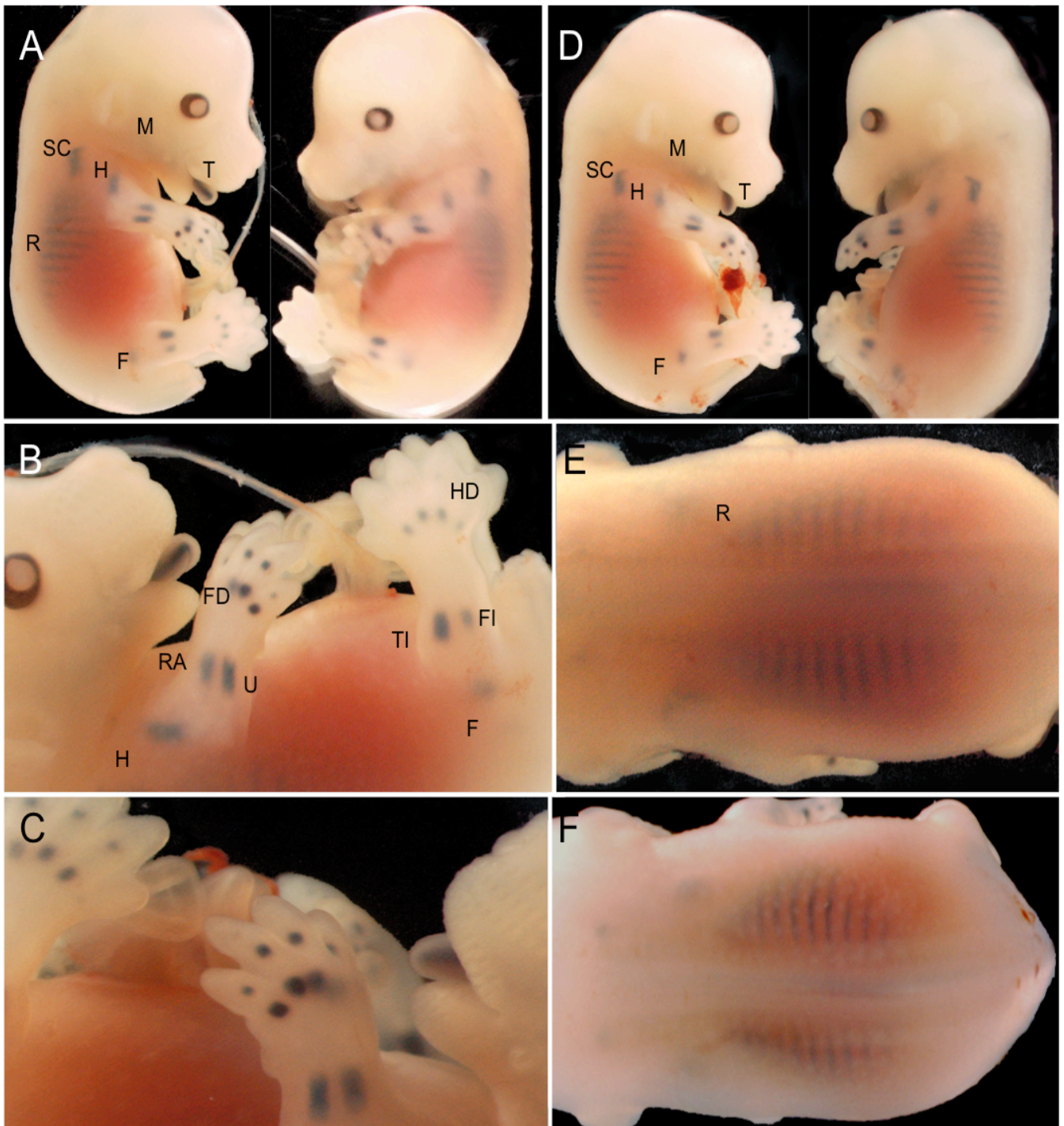


Figure 17: 2 representative E14.5 wholemount embryos expressing β -galactosidase (blue staining) driven by the β AGG20 promoter (-20,054bp - +375bp). A, B, and E are images from 1 transgenic embryo while C,D and F are images from the another. Both mice show identical and symmetrical staining. A,B,C and D are lateral images while E and F are dorsal images. F (femur), FD (forelimb digits), FI (fibula), H (humerus), HD (hindlimb digits), M (Meckel's cartilage), R (rib cartilage), RA (radius), SC (scapular cartilage). T (tongue), U (ulna).

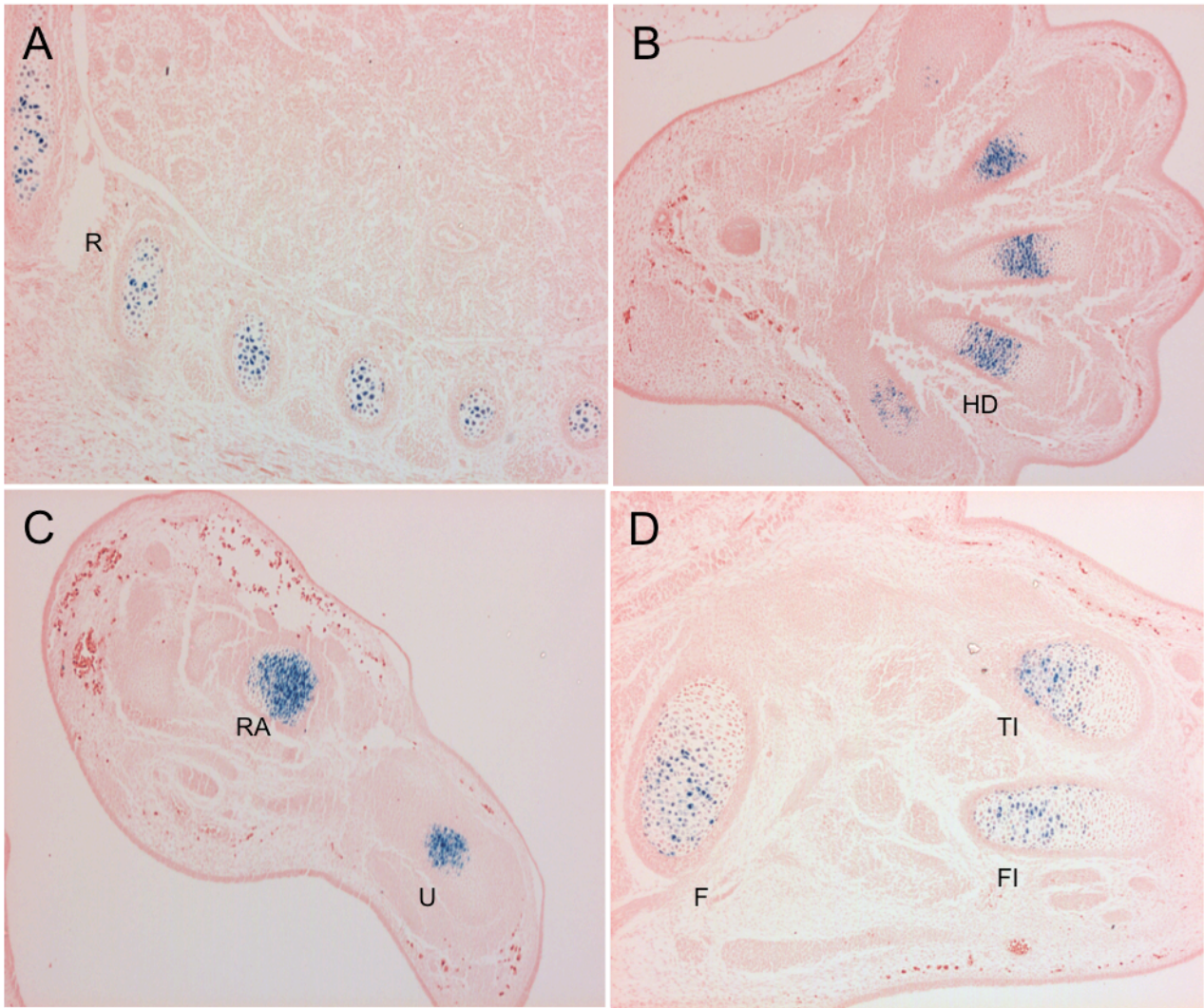


Figure 18: Histological sections of a blue stained transgenic embryo expressing β -galactosidase driven by the β AGG20 promoter (-20,054bp to +375bp). Blue staining localised in chondrocytes of the developing A) ribs (R). B) the digits of the hindlimb (HD). C) the ulna (U)/radius (RA) of the forelimb. D. the femur (F), tibia (TI) and fibula (FI). These sections are counterstained with eosin.

Table 7: Summary of transgenic and expressing mice driven by the β AGG20 promoter (-20,054bp to +375bp). ‘Positive embryos’ indicates the total number of embryos in which the transgene has become integrated throughout each session. ‘Expressing embryos’ indicates the total number of embryos in which the integrated reporter gene is being expressed. The ‘Percentage of embryos expressing the transgene’ was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β AGG20	7	3	42

Table 8: Average levels of expression for β AGG20 transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

Tissue	Intensity of staining	Tissue	Intensity of staining
Chondrocranium	-	Digits of the hindlimbs	++
Cartilage of the eye socket	-	Scapular cartilage	+++
Nasal cartilage	-	Cartilage of the humerus	+++
Meckels cartilage	+	Cartilage of the ulna	+++
Cartilage of the ribs (costal)	+++	Cartilage of the radius	+++
Cartilage of the femur	++	Intervertebral discs	-
Cartilage of the tibia	+++	Vertebral cartilage	-
Cartilage of the fibula	+++	Cartilage of the ear	-
Digits of the forelimbs	+++	Tongue	+++

β AGG20 was injected on another occasion (Table 9) and the embryos sacrificed and stained at E16.5 so as to verify whether the missing vertebral body expression gets switched on at a later stage in development. The pattern of β -galactosidase expression (blue staining) in the E16.5 β AGG20 transgenic embryo mimicked the pattern of expression in the E14.5 β AGG20

transgenic embryo in that all anatomical regions that were positive at E14.5 remained positive at E16.5 (Figure 19A+B+E+F) with the exception of the tongue. However, the intensity of the staining was much higher and was widened spatially to include more facial cartilages such as the nasal cartilage, the cartilage of the ear and the cartilage of the eye socket (Figure 19C) as well as the vertebral bodies and intervertebral disc (Figure 19D). It is important to note that at this stage of development, we would expect staining in the chondrocranium (Figure 13C). However, there was no staining detected in the chondrocranium. The intensity and localisation of X-gal staining is shown in Table 10.

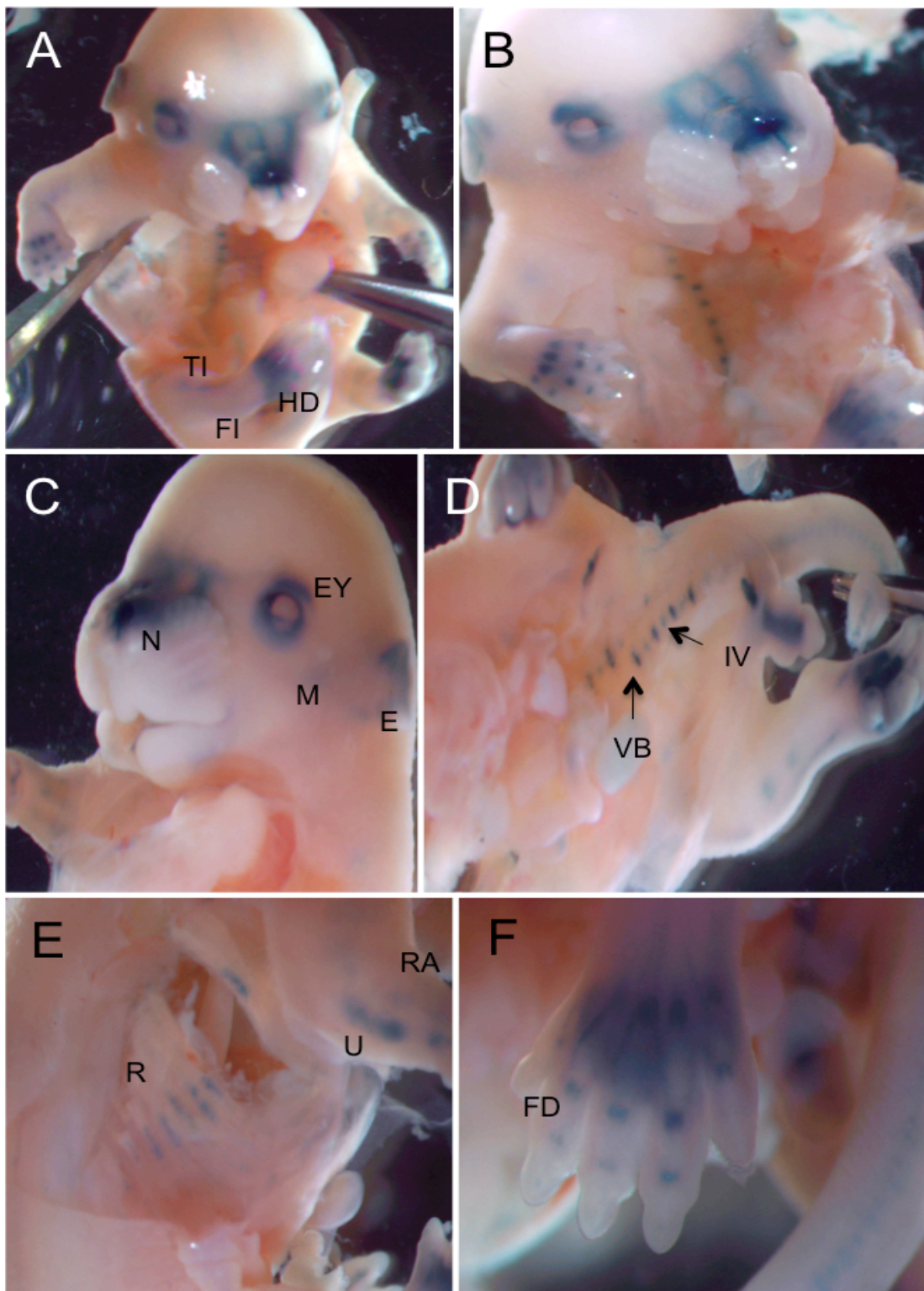


Figure 19: Representative E16.5 wholemount embryo expressing β -galactosidase (blue staining) driven by the β AGG20 promoter (-20,054bp to +375bp). A and B) ventral images of whole embryo C) craniofacial cartilages D) vertebral bodies and intervertebral discs E) ribs F) forelimb digits. E (ear cartilage), EY (cartilage of the eye socket), F (femur), FD (forelimb digits), FI (fibula), HD (hindlimb digits), IV (intervertebral discs), M (Meckel's cartilage), R (rib cartilage), RA (radius), U (ulna). VB (vertebral bodies).

Table 9: Summary of transgenic and expressing mice driven by the β AGG20 promoter (-20,054bp to +375bp). ‘Positive embryos’ indicates the total number of embryos in which the transgene has become integrated throughout each session. ‘Expressing embryos’ indicates the total number of embryos in which the integrated reporter gene is being expressed. The ‘Percentage of embryos expressing the transgene’ was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β AGG20	3	1	33

Table 10: Average levels of expression for β AGG20 transgenic mice at E16.5 after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, ++high level staining, ++++very high level staining.

Tissue	Intensity of staining	Tissue	Intensity of staining
Chondrocranium	-	Digits of the hindlimbs	++++
Cartilage of the eye socket	+++	Scapular cartilage	+++
Nasal cartilage	+++	Cartilage of the humerus	+++
Meckels cartilage	++	Cartilage of the ulna	+++
Cartilage of the ribs (costal)	+++	Cartilage of the radius	+++
Cartilage of the femur	+++	Intervertebral discs	+++
Cartilage of the tibia	+++	Vertebral cartilage	+++
Cartilage of the fibula	+++	Cartilage of the ear	+++
Digits of the forelimbs	++++	Tongue	-

Now that the presence of *cis*-regulatory element(s) within the ~20kb sequence had been established, the next step was to narrow this down in order to identify the minimal region required for the cartilage specific expression driven by this fragment. In parallel with this, I wanted to investigate the other potential regulatory regions identified in figure 14 (*cis* 6 + *cis* 7) since the *LacZ* expression driven by β AGG20 didn't exactly match endogenous aggrecan expression and showed lagging in the vertebral X-gal staining as well as missing staining in the chondrocranium.

1.2.5 Identifying the minimal region of regulatory sequence required for chondrocyte specific expression of aggrecan

1.2.5.1 *cis* 3 + 4

To try and identify the location of the putative regulatory element(s) within the ~20kb fragment in β AGG20 and decipher the function, if any, of *cis* 1-4, I made a series of constructs which contained different combinations of the '*cis*' elements that I had identified in the vista analysis (Figure 14). In order to identify which of these sequences was necessary, I used the program blast (Basic Local Alignment Search Tool), which is a widely available algorithm used to compare the nucleotides of two DNA sequences (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). The two sequences under comparison here were the ~20kb regulatory sequence found in β AGG20 along with the ~5kb rat non-coding sequence shown to act as a chondrocyte specific enhancer element when fused to the rat aggrecan 'basal promoter' (Doege et al., 2002).

The results showed an approximately 400bp highly homologous region (83% identity – 378/451 base pairs) located between -10.7kb to -10.4kb relative to the TSS of the human aggrecan gene (Figure 20). This sequence lies within *cis* 3. I removed ~7kb of DNA from the 5' end of β AGG20 via the restriction endonucleases *AhdI* and *SalI* leaving behind ~13kb of

human aggrecan non-coding DNA (corresponding to -13,030bp to +375bp relative to the TSS) upstream of the *LacZ* transgene. The resultant linear fragment of DNA was named β AGG13 and encompassed *cis* 3-5 without *cis* 1 and 2 (Methods section 5 construct 14).

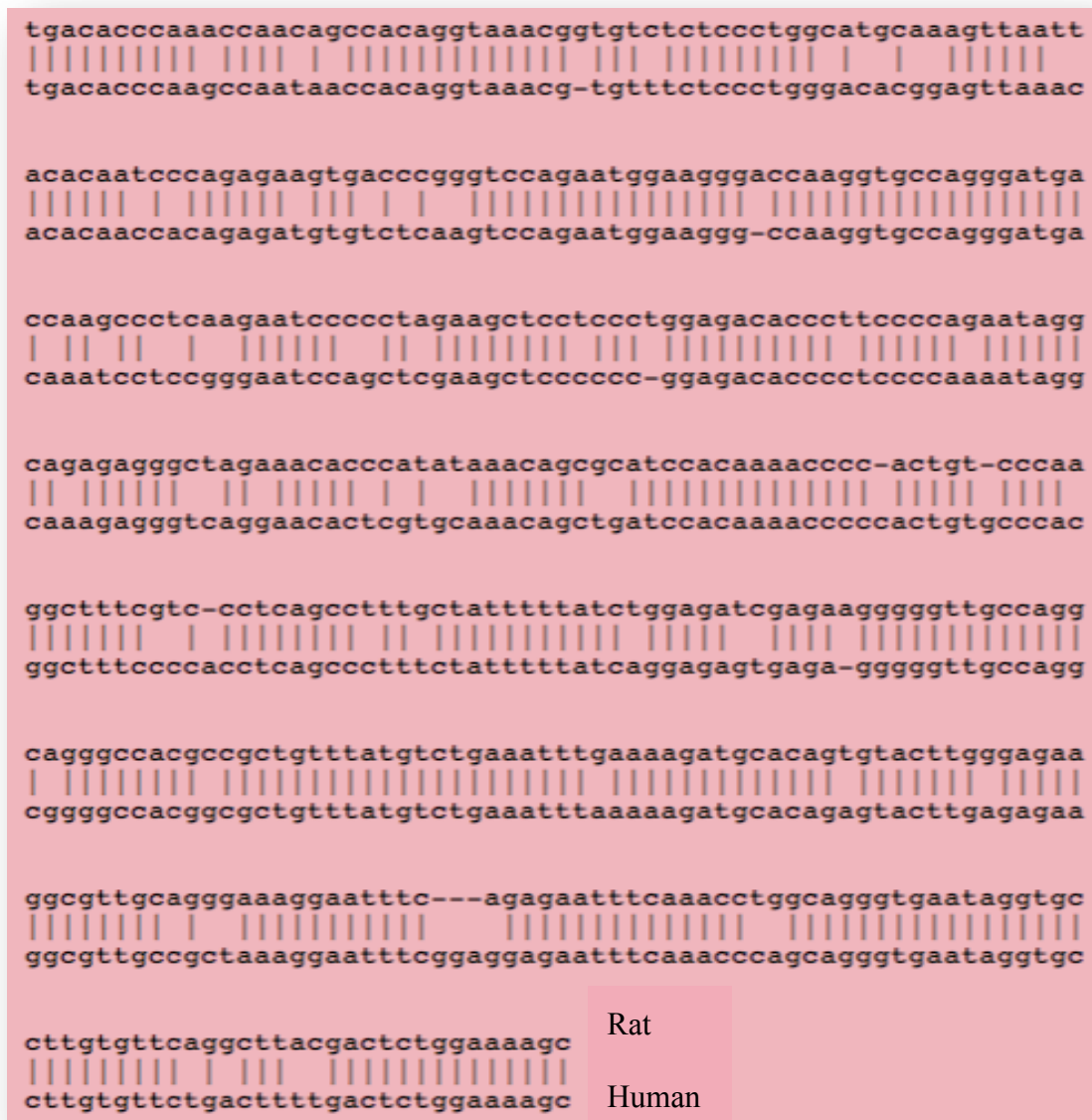


Figure 20: ~400bp sequence with 83% identity between Human DNA (-10.7kb to 10.4kb relative to the human aggrecan TSS) and Rat DNA (-9.2kb to -8.7kb relative to the rat aggrecan TSS). The sequence was shown to be the most highly conserved sequence identified by performing a blast Alignment between the ~20kb regulatory sequence found in β AGG20 (between -20,054bp and +375bp relative to the human aggrecan TSS) with the ~5kb rat non-coding sequence (between -12kb and -7kb relative to the rat aggrecan TSS) shown to act as a chondrocyte specific enhancer element when fused to the rat aggrecan ‘basal promoter’.

This shorter fragment was then injected into the pronuclei of fertilized murine eggs to generate β AGG13 transgenic embryos. The embryos were evaluated for *LacZ* expression by whole mount X-gal staining overnight as before. None of the positive embryos showed any kind of staining. This newly integrated DNA fragment therefore seemed to have lost vital sequences that enabled it to regulate the expression of the *LacZ* gene in cartilage. Table 11 shows the number of positive embryos containing the injected transgene and the number of transgenic embryos expressing the reporter gene.

Table 11: Summary of transgenic and expressing mice driven by the β AGG13 promoter (-13,030bp to +375bp). 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β AGG13	11	0	0

1.2.5.2 *cis* 1 + 2

Next, I decided to clone this seemingly important deleted ~7kb fragment (corresponding to -20,054bp to -13,030bp relative to the TSS) directly upstream of the human aggrecan basal promoter AB and into the β gal-basic vector. This would exclude *cis* 3 and 4 while including *cis* 1, 2 and 5 in the construct. As before, the cloning was carried out by recombineering (Figure 21). According to the recombineering method and to generate the retrieval plasmid, I cloned a 5' fragment of DNA corresponding to the 5' end of the deleted ~7kb fragment (same as the 5' arm used to clone the ~20kb fragment in β AGG20 and named EF) along with a 3' fragment corresponding to the 3' end of the deleted ~7kb fragment (named IJ) upstream of the minimal

promoter of the aggrecan gene (AB) and into the p β gal basic vector. This vector was named β galAB+IJ+EF (see Methods section 5 construct 7). This vector was then linearized by digesting between the 5' and 3' arms and co-electroporated with a **linear** target DNA sequence (the deleted ~7kb fragment) into recombination competent EL350 cells. The resultant vector was named β galAB + 7 (see Methods section 5 construct 15).

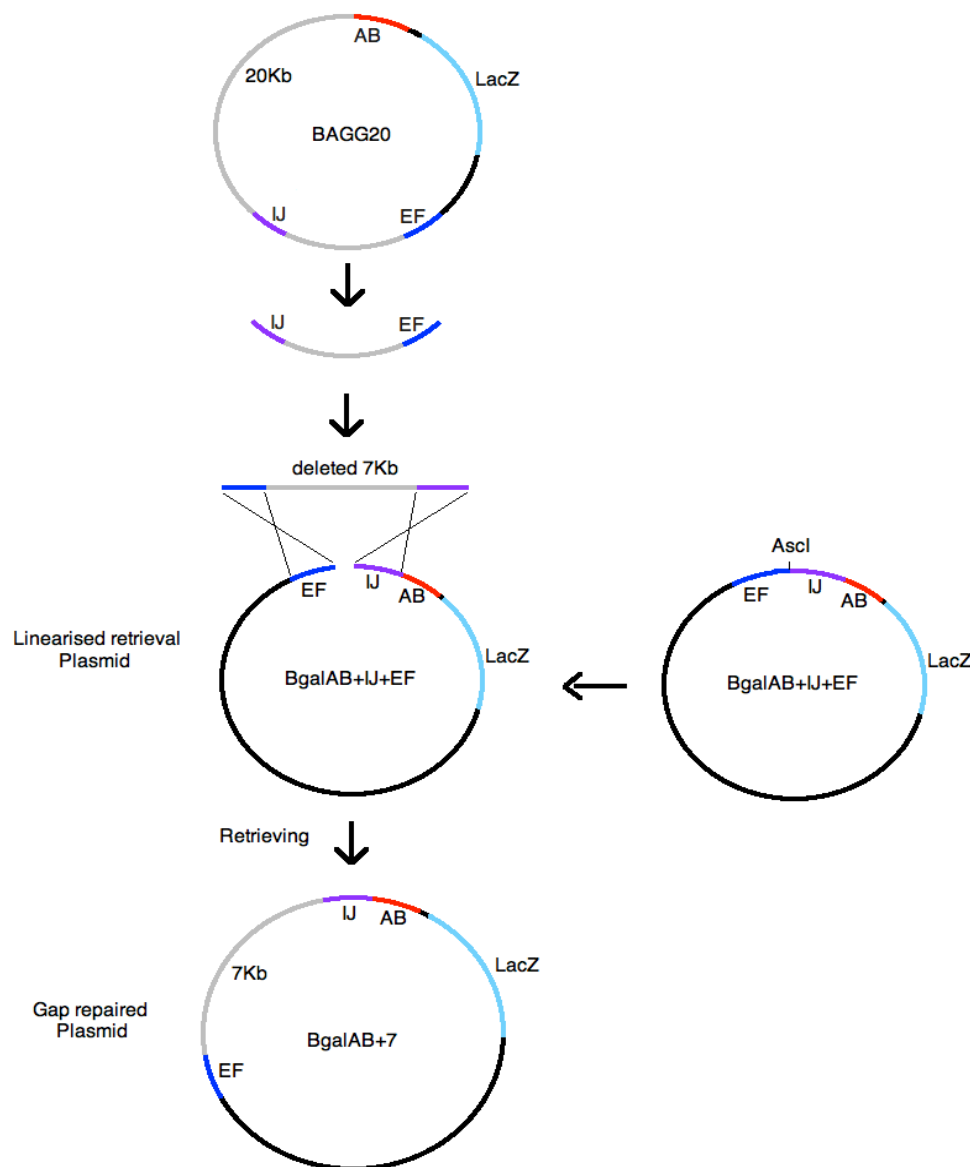


Figure 21: Diagram representing the recombineering strategy used to clone the deleted 7kb fragment into the p β gal-Basic vector upstream of the basal promoter and *LacZ* gene. The 5' arm (EF) and 3' arm (IJ) are coloured in dark blue and purple respectively while the 'basal promoter' (AB) is coloured in red. The *LacZ* gene is shown in light blue. Here, the targeting 7kb fragment is digested out of β AGG20 to be recombined into the p β gal-Basic vector upstream of the basal promoter of aggrecan.

The vector ' β galAB + 7' was then digested to release its backbone and subsequently injected into the pronuclei of fertilised murine eggs. Out of several injections and numerous transgenics, two E14.5 embryos showed similar and symmetrical cartilage specific expression in fewer developing cartilaginous regions compared to the E14.5 β AGG20 transgenic embryo (Table 12 + 13). Figure 22 Demonstrates one of these embryos. Staining was present in the cartilage of the long bones of the forelimbs - ulna and radius (Figure 22D), the nasal cartilage (Figure 22C) and minimal expression in the cartilage of the developing hindlimbs - femur and tibia (Figure 22B). Figure 23 shows histological sections of the hindlimb including the tibia and digits (Figure 23A), the elbow including the humerus, ulna and radius (Figure 23B) as well as the ribs (Figure 23C-E) of the representative β galAB + 7 embryo shown in Figure 22. In all histological sections, the staining is evident in both proliferating and hypertrophic chondrocytes but is lower, more random and scattered/mosaic compared to the staining present within the histological sections of the E14.5 β AGG20 embryo in Figure 18.

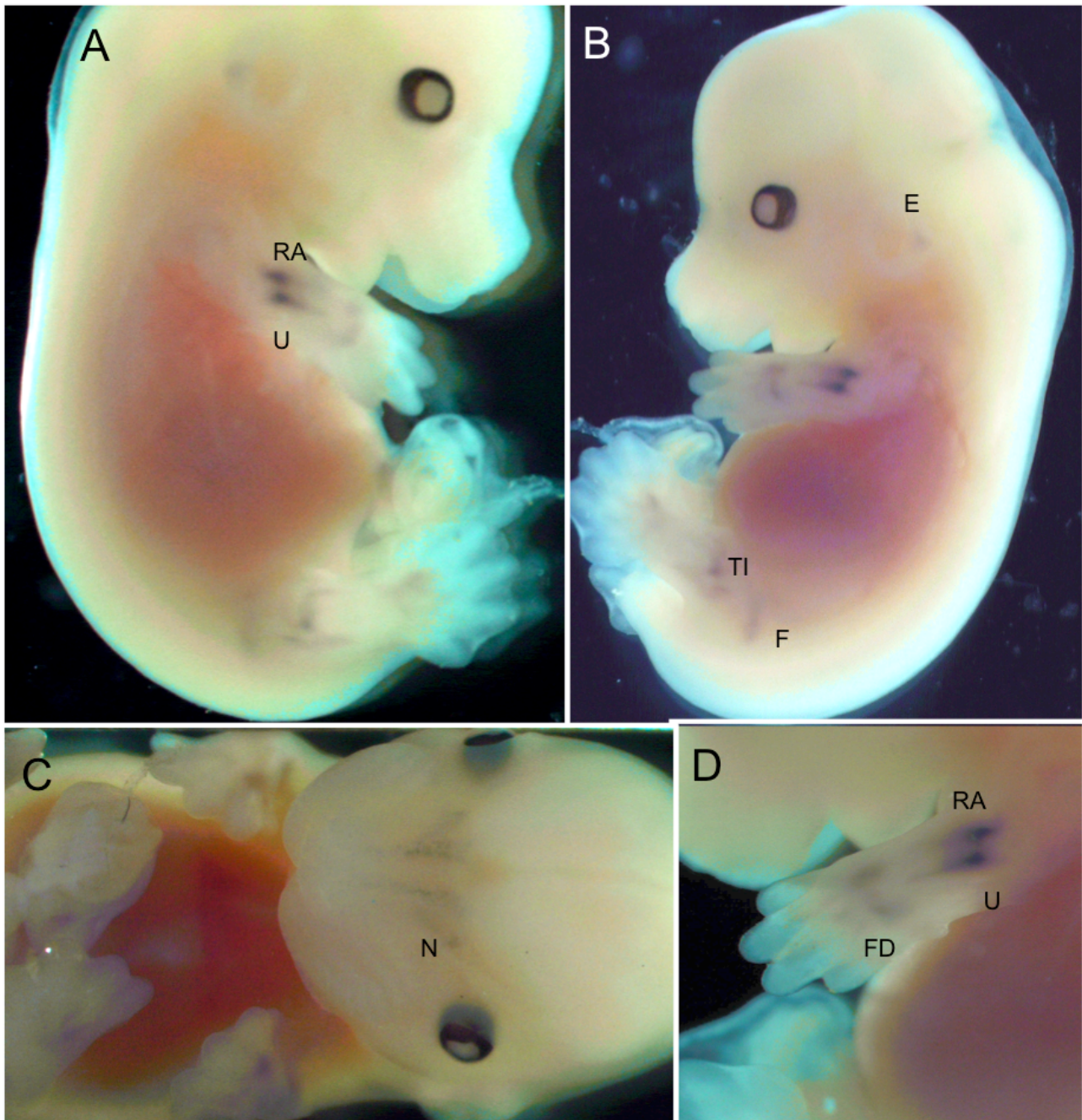


Figure 22: Representative E14.5 wholemount embryo expressing β -galactosidase (blue staining) driven by the β galAB+7 promoter (-20,054bp to -13,030bp fused to minimal promoter/-452bp to +375bp). A and B) embryo shows completely symmetrical staining. A,B and D are lateral images while C is a ventral image. E (ear cartilage), F (femur), FD (forelimb digits), N (nasal cartilage), RA (radius), TI (tibia), U (ulna).

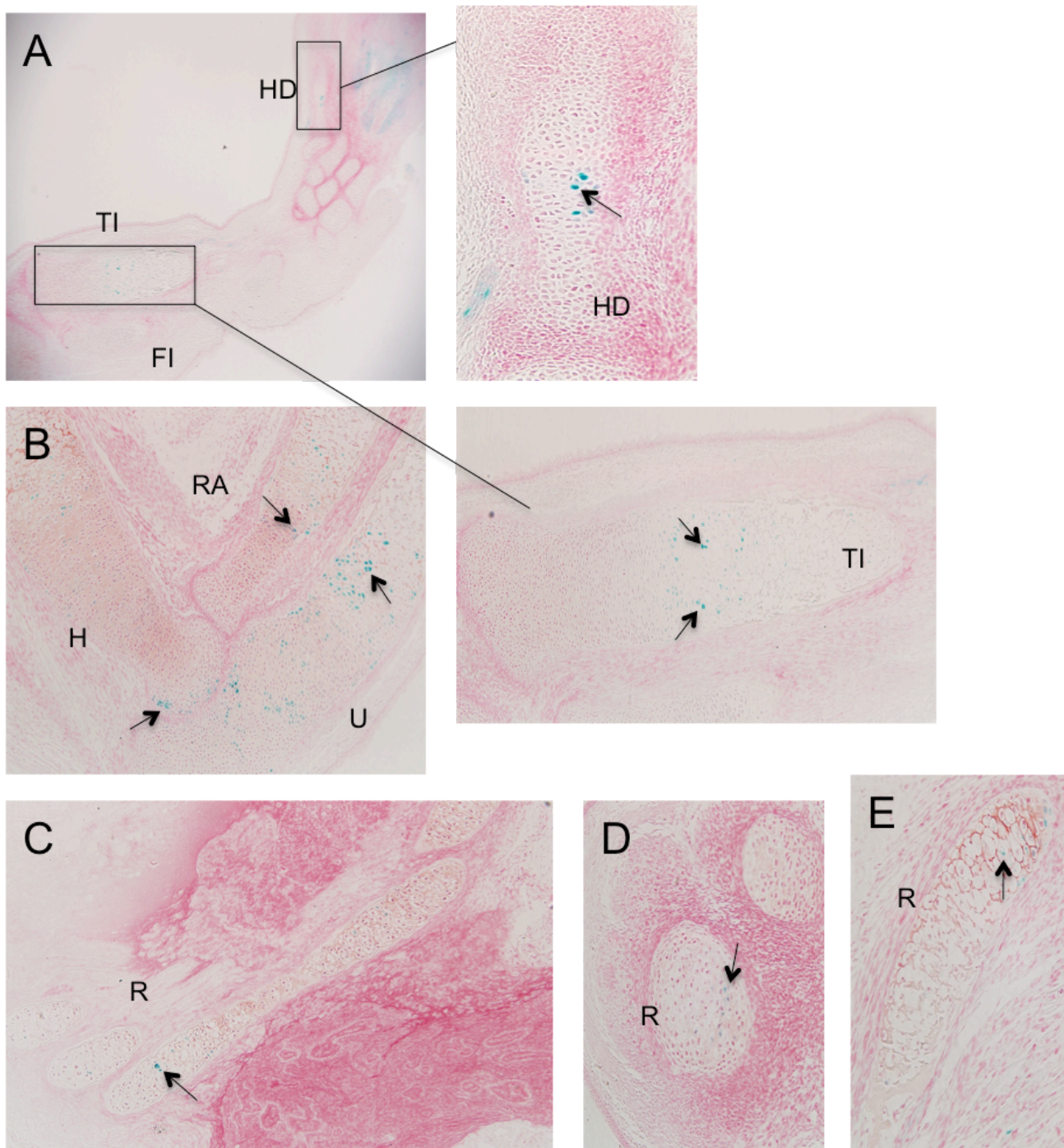


Figure 23: Histological sections of a representative $\beta\text{galAB}+7$ blue stained embryo. A) 10x magnification of the tibia (TI), fibula (FI) and digits of the hindlimb (HD) (left) with corresponding 20x magnification of boxed regions (right). Upper boxed region shows expression of the transgene in the hypertrophic chondrocytes of one of the digits while lower boxed region shows expression of the transgene in the proliferating and hypertrophic chondrocytes of the tibia. B) Humerus, ulna and radius all expressing β -galactosidase in a scattered fashion in the proliferating and hypertrophic chondrocytes. C-E) Ribs expressing β -galactosidase in a scattered fashion within hypertrophic chondrocytes. Arrows point to examples of blue stained cells. These sections were eosin counterstained.

Table 12: Summary of transgenic and expressing mice driven by the β galAB+7 promoter (-20,054bp to -13,030bp fused to minimal promoter/ -452bp to +375bp). 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β galAB+7	21	2	10

Table 13: Average levels of expression for β galAB+7 transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

Tissue	Intensity of staining	Tissue	Intensity of staining
Chondrocranium	-	Digits of the hindlimbs	-
Cartilage of the eye socket	-	Scapular cartilage	-
Nasal cartilage	++	Cartilage of the humerus	-
Meckels cartilage	-	Cartilage of the ulna	+++
Cartilage of the ribs (costal)	-	Cartilage of the radius	+++
Cartilage of the femur	+	Intervertebral discs	-
Cartilage of the tibia	+	Vertebral cartilage	-
Cartilage of the fibula	-	Cartilage of the ear	+
Digits of the forelimbs	++	Tongue	-

1.2.5.3 *cis* 1+2+4

In the next construct, I removed a fragment of DNA approximately 10kb in size within the approximately 20kb sequence in 'βAGG20' between ~ 14kb and 4kb upstream of the transcriptional start site by restriction enzymes *KpnI* and *BstBI* (Methods section 5 construct 13). This construct was named 'βAGG10' and excluded *cis* 3 while including *cis* 1,2,4 and 5. This yielded 18 transgenic embryos at E14.5 containing the transgene with no expression (Table 14). It seemed that the cartilage specific expression seen in the βAGG20 transgenics was lost.

Table 14: Summary of transgenic and expressing mice driven by the βAGG10 promoter (-20,054bp to -13,838bp fused to -4,182bp to +375bp). 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
βAGG10	18	0	0

1.2.5.4 Missing 92bp sequence

When βAGG20 was initially cloned, the sequencing of the minimal promoter AB within *cis* 5 on the 3' end of the ~20kb fragment showed a deletion of 92 base pairs between -68 to +23bp relative to the TSS. Upon further investigation, it was clear that the BAC RP11 343B18 used to create βAGG20 harboured this deletion. Nevertheless, βAGG20 transgenics displayed cartilaginous expression and the ~20kb sequence was manipulated to produce all subsequent clones in which some also activated *LacZ* expression in transgenic embryos. It was possible, however, that the 92 base pairs played a necessary role for the full endogenous expression of the gene and therefore had to be investigated. Validating a role for the missing 92 base pairs

would require cloning a new ~20kb fragment inclusive of this missing sequence within the vector backbone β gal basic in the same way. I acquired a new BAC (RP11-1123E8) and checked for the presence of the lacking sequence. Following this, a targeting vector containing the new 5' and 3' arms was cloned and called β galAB⁺⁹²+EF (see Methods section 5 construct 4). Ongoing work at the KIR will be done to perform the necessary recombineering experiments to produce a construct containing the ~20kb fragment in β AGG20 with the inclusion of the missing 92 base pairs (β AGG20⁺⁹²) and consequently generate β AGG20⁺⁹² transgenic embryos.

I also performed other experiments to investigate the role of the missing 92 base pairs. Around 2 years into my PhD, the paper by Yu Han and Véronique Lefebvre (2008) 'L-Sox5 and Sox6 Drive Expression of the Aggrecan Gene in Cartilage by Securing Binding of Sox9 to a Far-Upstream Enhancer' was published and it confirmed the presence of an approximately 400 base pair enhancer within *c/s* 3 that I had already identified (Figure 20). I designed and constructed two clones; one containing the minimal promoter AB fused to the highly homologous 400 base pair fragment within *c/s* 3 and the other containing the same sequences with the inclusion of the missing 92 base pairs (amplified from the new BAC RP11-1123E8). The constructs were named ' β galAB+*c/s* 3' (see Methods section 5 construct 5) and ' β galAB⁺⁹²+*c/s* 3' (see Methods section 5 construct 6) respectively.

Upon linearization of the construct ' β galAB+*c/s* 3', it was microinjected into the pronuclei of fertilized eggs and embryos were sacrificed at E14.5 days and stained with X-gal. Out of 8 transgenics, 1 embryo showed minimal cartilage specific expression in far fewer anatomical regions compared to the E14.5 β AGG20 embryos (Table 15 + 16). Figure 24 shows this embryo in which the β -galactosidase expression/blue staining is limited to the nasal cartilage (Figure 24B) and the cartilage of the distal digits of the forelimb (Figure 24A).

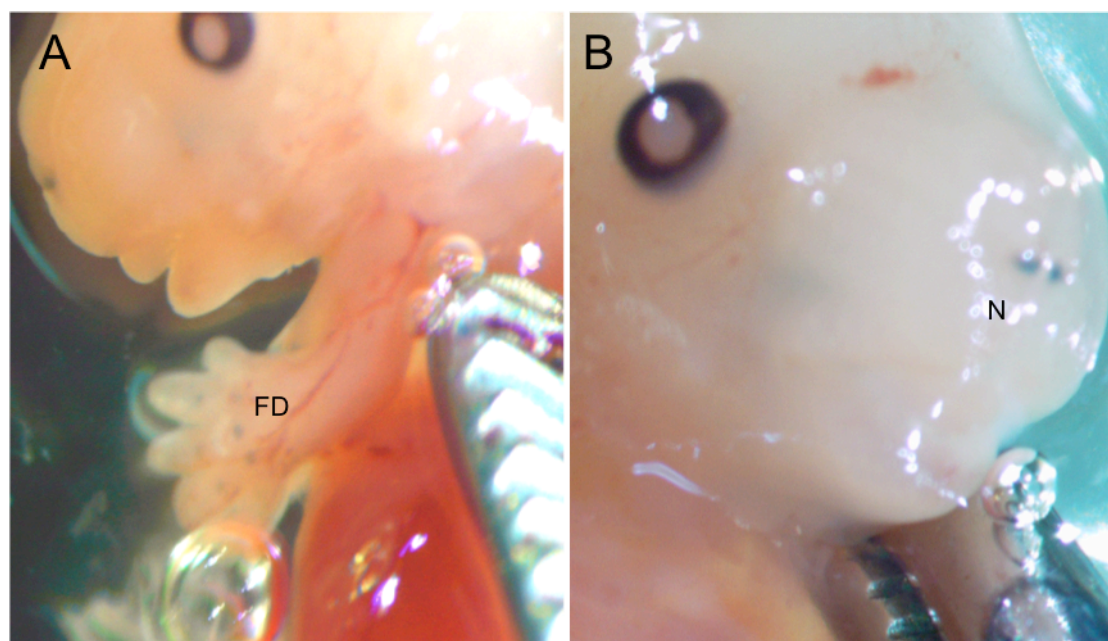


Figure 24: Representative E14.5 wholemount embryo expressing β -galactosidase (blue staining) driven by the β galAB + *cis* 3 promoter (-10,703bp - -10,343bp fused to minimal promoter/-452bp - +375bp). A) Lateral image of embryo B) Facial region. FD (forelimb digits), N (nasal cartilage).

Table 15: Summary of transgenic and expressing mice driven by the β galAB + *cis* 3 promoter (-10,703bp to -10,343bp fused to minimal promoter/-452bp to +375bp). 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

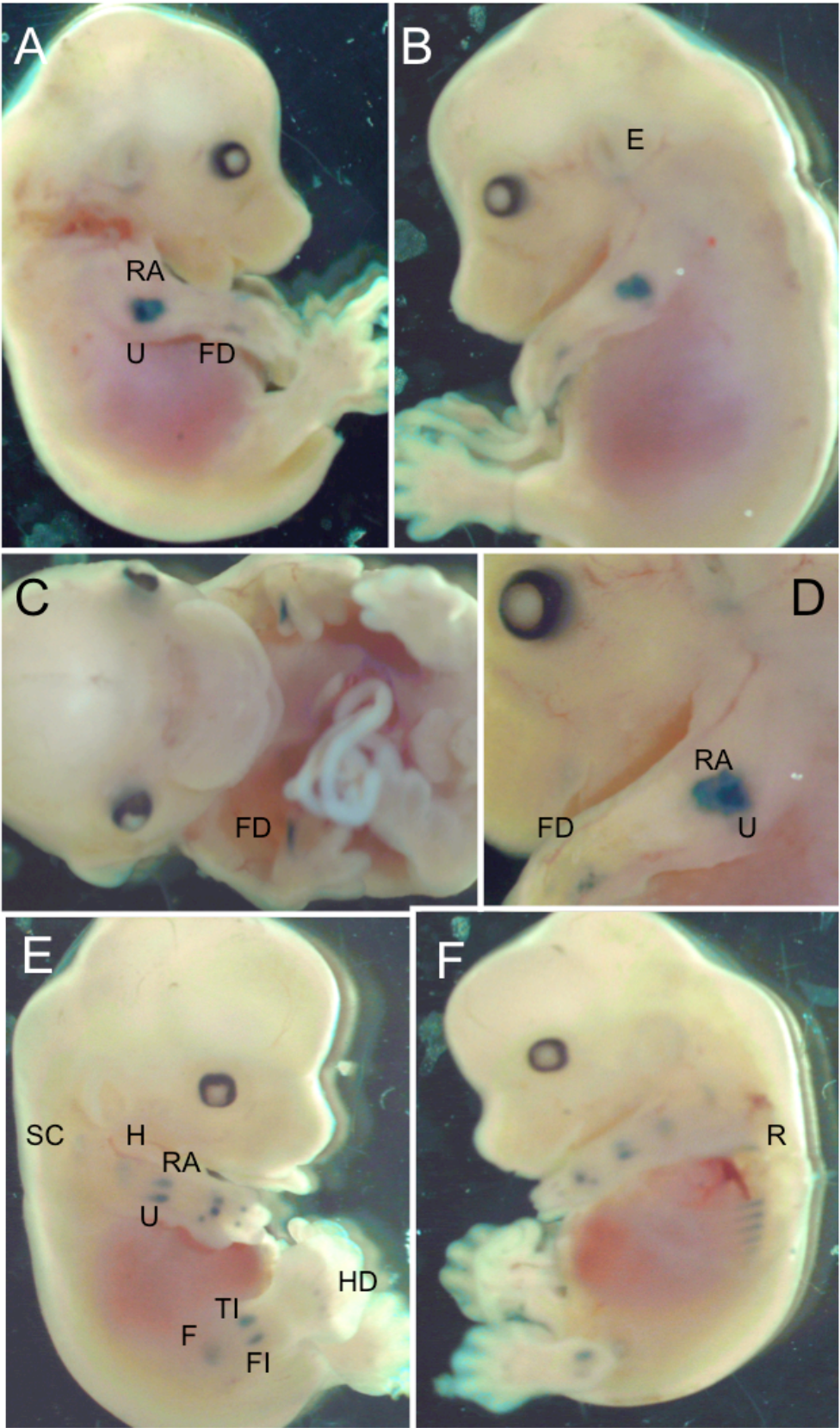
Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β galAB+ <i>cis</i> 3	8	1	13

Table 16: Average levels of expression for β galAB+cis 3 transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

Tissue	Intensity of staining	Tissue	Intensity of staining
Chondrocranium	-	Digits of the Hindlimbs	-
Cartilage of the eye socket	-	Scapular cartilage	-
Nasal cartilage	++	Cartilage of the humerus	-
Meckels cartilage	-	Cartilage of the ulna	-
Cartilage of the ribs (costal)	-	Cartilage of the radius	-
Cartilage of the femur	-	Intervertebral discs	-
Cartilage of the tibia	-	Vertebral cartilage	-
Cartilage of the fibula	-	Cartilage of the ear	-
Digits of the Forelimbs	++	Tongue	-

The construct ' β AGGAB⁺⁹²+ cis 3' was also linearized, microinjected into the pronuclei of fertilized eggs, and embryos sacrificed at E14.5 days and stained with X-gal. Four out of 11 positive transgenic embryos showed different levels/patterns of cartilaginous expression (Table 17 + 18). In 2 of the 4 expressing embryos, the staining was localized to the developing cartilage of the long bones of the fore and hind limbs, the ribs, the scapulae, and the digits (Figure 25E+F and Figure 25G+H). This somewhat resembled the staining pattern of the β AGG20 transgenic embryos but with a much lower intensity and with the exclusion of the staining found in the tongue and meckel's cartilage of the β AGG20 transgenic embryos. In the other 2 embryos, the staining pattern was also cartilage-specific but different. One embryo showed very strong X-gal staining in the ulna and radius with weaker staining in the first and fifth digit of the forelimbs (Figure 25A-D) and the other showed weak staining in the ribs, scapula, cartilage of the eye socket and the digits of the forelimb (Figure 25I+J). Figure 26

shows a representative histological sagittal section of the forelimb and torso of the positive expressing β AGGAB⁺⁹² *cis* 3 transgenic embryo in Figure 25A-D. It is clear from this section that the blue staining is located in the developing cartilage of the humerus and ulna and is very weak + scattered randomly across the proliferating chondrocytes and larger hypertrophic chondrocytes. However, in this embryo, although there is staining located within the chondrocytes of the developing long bones of the forelimb, it is also more strongly found in fibroblasts surrounding these developing bones. Aggrecan is not known to be expressed within fibroblasts at this stage of development.



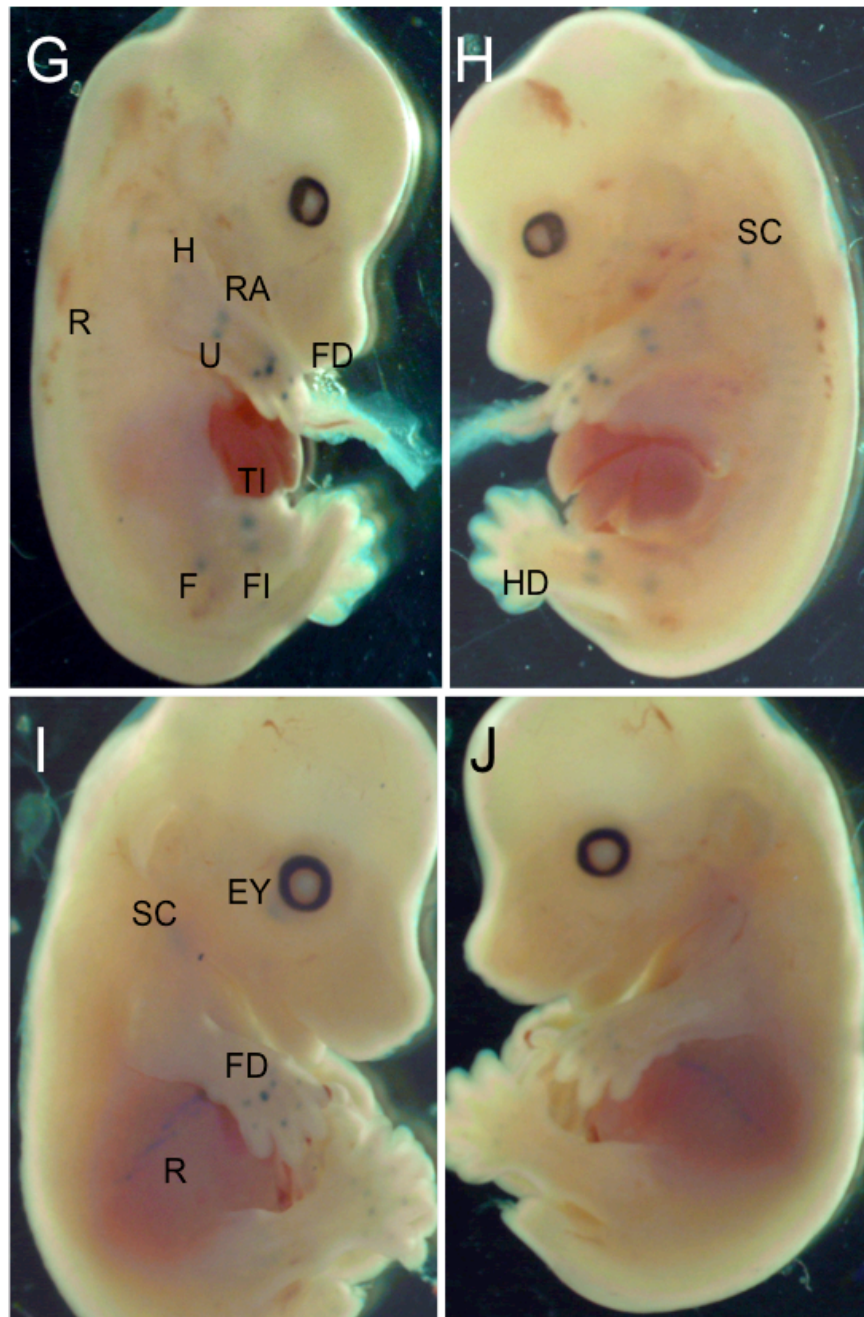


Figure 25: 4 representative E14.5 wholemount embryos expressing β -galactosidase (blue staining) driven by the β galAB⁺⁹²+cis 3 promoter (-10,703bp to -10,343bp fused to minimal promoter *including missing 92 base pairs* -452bp to +375bp). A, B, C and D are images from the 1st transgenic embryo. E + F are images from 2nd, G + H are images from the 3rd while I + J are images from the 4th transgenic embryo. All individual embryos show symmetrical staining. All images are lateral except for C, which is a ventral image. E (ear), EY (cartilage of the eye socket), F (femur), FD (forelimb digits), FI (fibula), H (humerus), HD (hindlimb digits), R (rib cartilage), RA (radius), SC (scapular cartilage), TI (tibia), U (ulna).

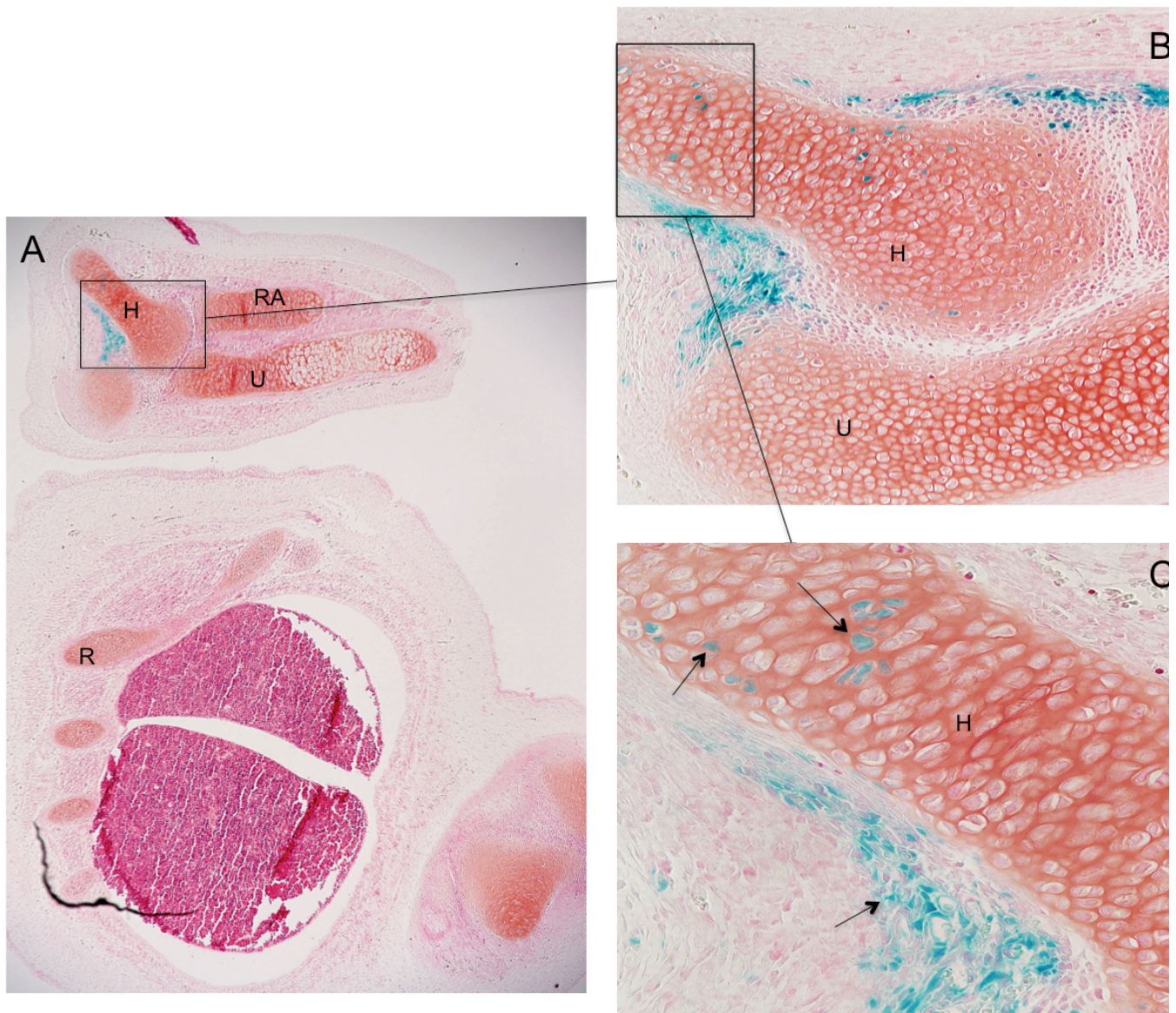


Figure 26: Histological section of a $\beta\text{galAB}^{+92}+\text{cis } 3$ blue stained embryos. A,B and C are different magnifications of the same section. A) 4x magnification of sagittal section of transgenic embryo showing the hindlimb and the upper torso. B) 10x magnification of humerus. C) 20x magnification of humerus showing scattered blue X-gal staining in small proliferating chondrocytes and large hypertrophic chondrocytes as well as non-chondrocytic fibroblastic cells. Section was eosin counterstained. H (humerus), R (rib cartilage), RA (radius), U (ulna).

Table 17: Summary of transgenic and expressing mice driven by the β AGGAB⁺⁹²+ *cis* 3 promoter (-10,703bp to -10,343bp fused to minimal promoter *including missing 92 base pairs*/-452bp to +375bp). . 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
' β AGGAB ⁺⁹² + <i>cis</i> 3'	11	4	36

Table 18: Average levels of expression for β AGGAB⁺⁹²+ *cis* 3 transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

a staining only found in 1 out of 4 expressing embryos

b staining of the ribs found in 3 out of 4 expressing embryos, one of which only shows staining in the distal part of the developing ribs

c staining found in 2 out of 4 expressing embryos

d staining found in all 4 expressing embryos, but localised to the first and fifth digit in one of the expressing embryos

e staining found in 3 out of 4 expressing embryos

f staining found in 3 out of 4 expressing embryos, one of which shows high level staining compared to low level staining in the other 2

Tissue	Intensity of staining	Tissue	Intensity of staining
Chondrocranium	-	Digits of the hindlimbs	- e
Cartilage of the eye socket	+ a	Scapular cartilage	+ e
Nasal cartilage	-	Cartilage of the humerus	+ c
Meckels cartilage	-	Cartilage of the ulna	++ f
Cartilage of the ribs (costal)	+ b	Cartilage of the radius	++ f
Cartilage of the femur	+ c	Intervertebral discs	-
Cartilage of the tibia	++ c	Vertebral cartilage	-
Cartilage of the fibula	++ c	Cartilage of the ear	+ a
Digits of the forelimbs	++ d	Tongue	-

1.2.6 Finding additional *cis*-acting element(s)

1.2.6.1 *cis* 6

To test the potential regulatory activity of *cis* 6, approximately 3kb of non-coding DNA within the first intron of the human aggrecan gene (+5 to +8kb) was cloned directly downstream of the 20kb fragment in β AGG20 and upstream of the *LacZ* reporter gene. According to the recombineering method, I cloned the 5' arm corresponding to the 3' end of AB from β AGG20 along with a 3' arm corresponding to the 5' end of the *LacZ* gene from β AGG20 on either end of *cis* 6 into the pBluescript KS+ vector (Figure 27.1 and 27.2). The resultant vector was called pBS/*cis* 6 recomb (Methods section 5 construct 9). The 5'arm/*cis* 6/3'arm fragment was then digested out of pBS/*cis* 6 recomb and used as a target linear sequence. β AGG20 was digested with *Xho*I and used as the retrieval targeting plasmid (Figure 27.3). The resultant gap-repaired vector was named ' β AGG20 + *cis* 6' (Methods section 5 construct 16). β AGG20 + *cis* 6 was then digested to release its backbone and subsequently injected into the pronuclei of fertilised murine eggs. The microinjection of β AGG20 + *cis* 6 yielded 3 transgenic embryos with identical and symmetrical patterns of reporter gene expression (Table 19 + 20). Figure 28 shows 3 of these embryos. The expression of the transgene was reduced in comparison to β AGG20 transgenic embryos and limited to the distal digits of the forelimbs as well as the long bones of the fore and hindlimbs where only minimal expression was noted. All other cartilage specific expression found in the β AGG20 embryos was lost.

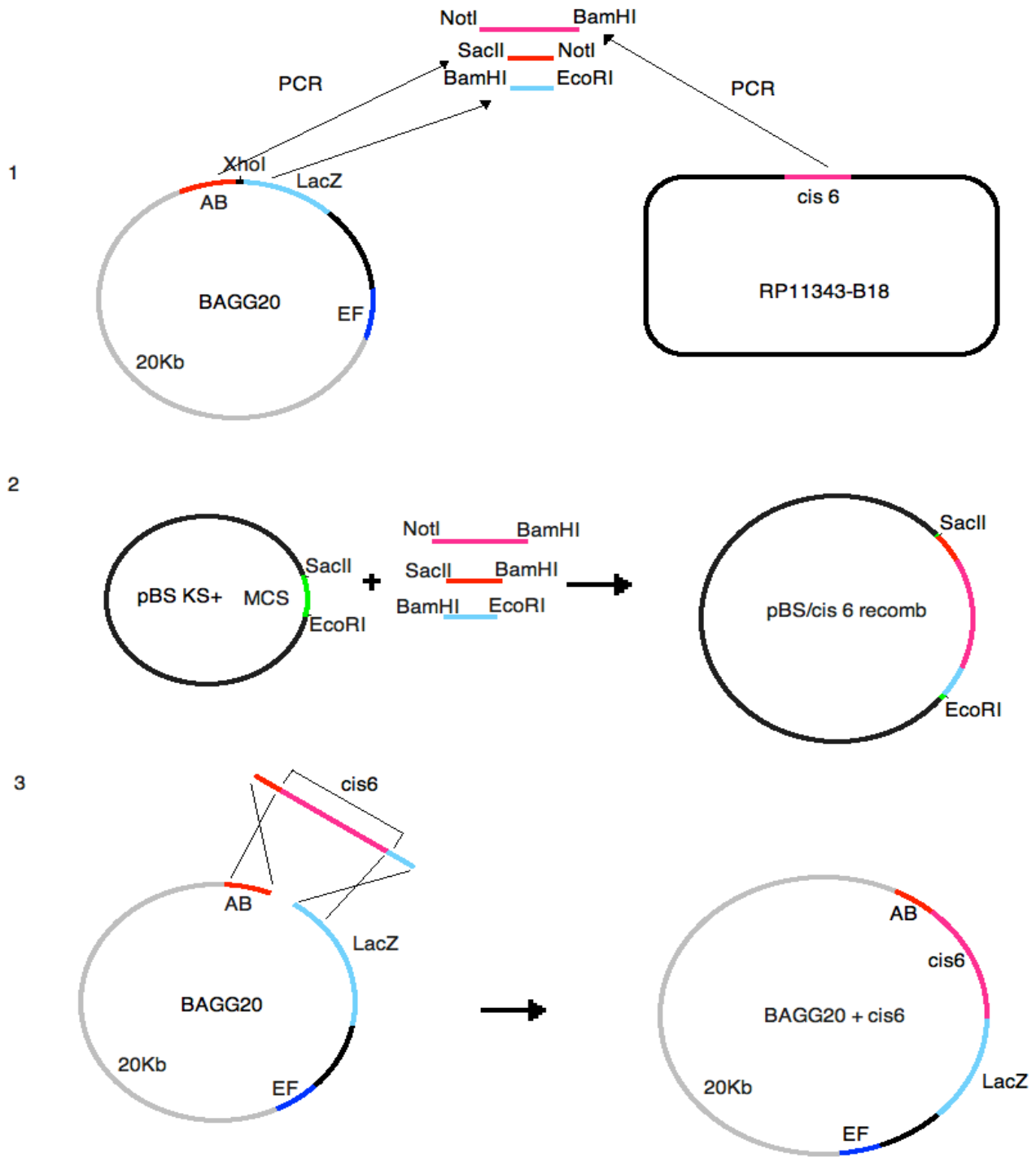


Figure 27: Diagram representing the recombineering strategy used to clone the 3kb intronic sequences (+5kb to +8kb) in an indirect way from the BAC RP11-343B18 into β AGG20 downstream of the ~20 kb insert and upstream of the *LacZ* gene. The indirect method uses pBS KS+ as an intermediate target vector. The 5' and 3' arms are coloured in red and blue respectively while the 3kb intronic sequence is coloured in pink. The pBS KS+ multiple cloning site is green. Note: *cis 6* was amplified by proof-reading pfx polymerase and used to generate the vector β AGG20 + *cis 6*.

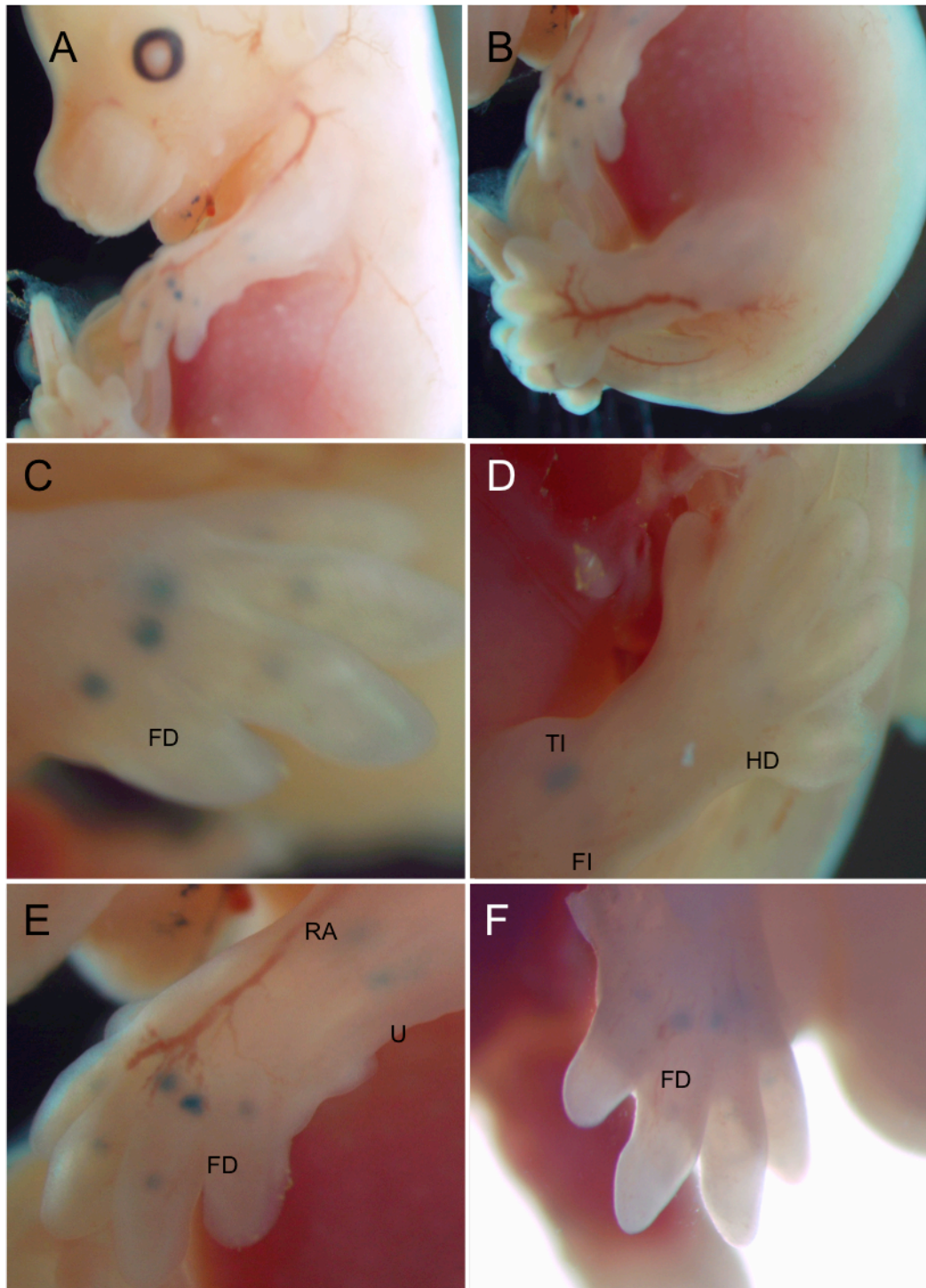


Figure 28: 3 representative E14.5 wholemount embryos expressing β -galactosidase (blue staining) driven by the β AGG20 + *cis* 6 promoter (-20,054bp to +375bp fused to +5,000bp to +7,813bp). A, B, C and D are images from 1 transgenic embryo while E and F are each from separate embryos. A) Saggital image of upper torso region including head. B) Saggital image of lower torso region including hindlimbs. C, E and F) forelimbs D) Hindlimb. FD (forelimb digits), FI (fibula), HD (hindlimb digits), RA (radius), TI (tibia), U (ulna).

Table 19: Summary of transgenic and expressing mice driven by the β AGG20 + *cis* 6 promoter (-20,054bp to +375bp fused to +5,000bp to +7,813bp). ‘Positive embryos’ indicates the total number of embryos in which the transgene has become integrated throughout each session. ‘Expressing embryos’ indicates the total number of embryos in which the integrated reporter gene is being expressed. The ‘Percentage of embryos expressing the transgene’ was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β AGG20 + <i>cis</i> 6	18	3	17

Table 20: Average levels of expression for β AGG20 + *cis* 6 transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

Tissue	Intensity of staining	Tissue	Intensity of staining
Chondrocranium	-	Digits of the hindlimbs	+
Cartilage of the eye socket	-	Scapular cartilage	-
Nasal cartilage	-	Cartilage of the humerus	-
Meckels cartilage	-	Cartilage of the ulna	++
Cartilage of the ribs (costal)	-	Cartilage of the radius	++
Cartilage of the femur	-	Intervertebral discs	-
Cartilage of the tibia	++	Vertebral cartilage	-
Cartilage of the fibula	-	Cartilage of the ear	-
Digits of the forelimbs	++	Tongue	-

1.2.6.2 *cis* 7

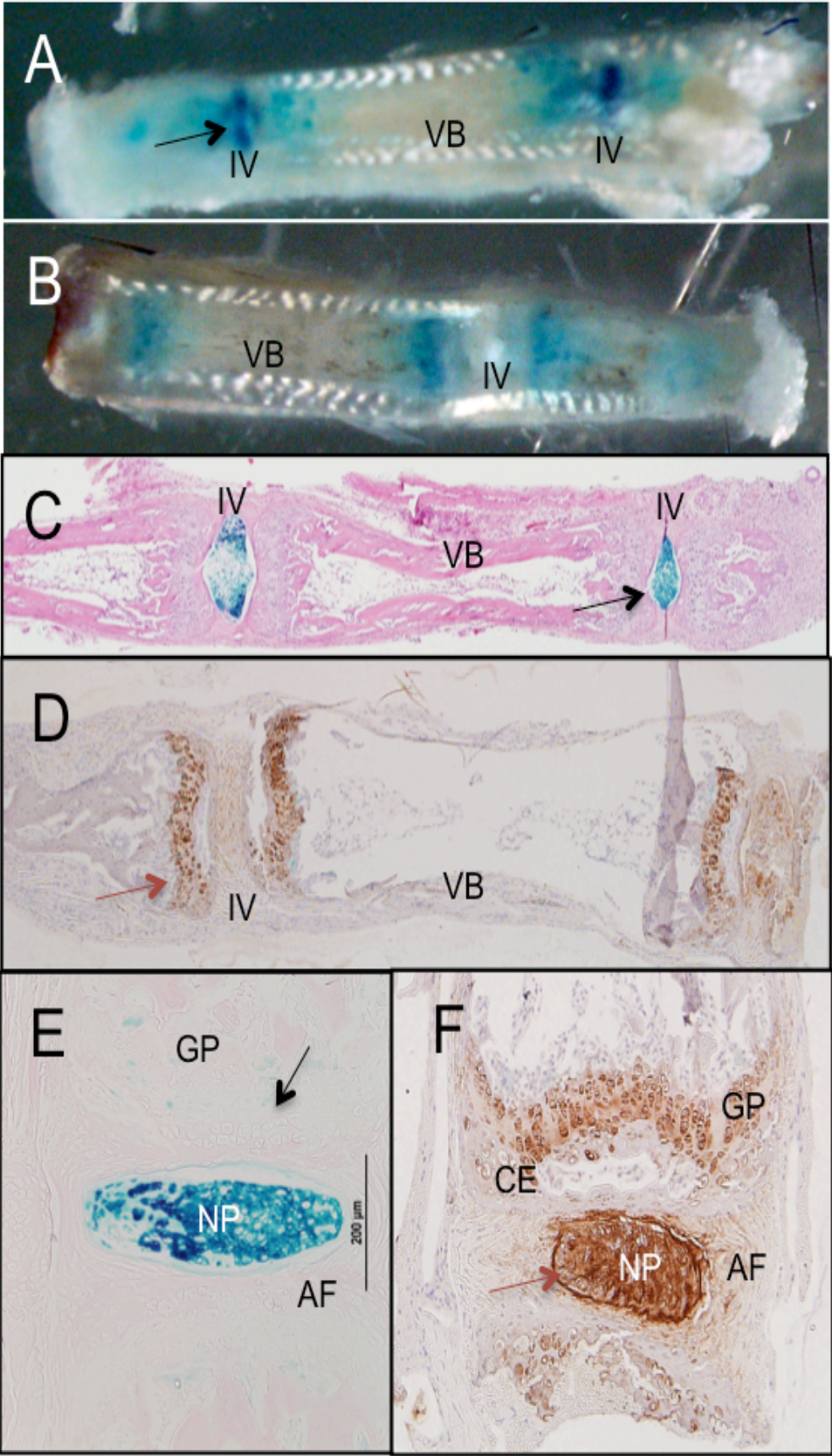
The highly conserved sequence *cis* 7 between humans and *fugu* within the 14th intron of the human aggrecan gene is around 150 base pairs (see Figure 15). I designed primers flanking a 400 base pair region, which encompassed the homologous sequence and cloned it

downstream of *cis* 3 and AB (including missing 92 base pairs). The resultant vector was called ' β AGGAB⁺⁹² + *cis* 3 + *cis* 7' (Methods section 5 construct 8). Unfortunately, although this construct was microinjected several times, no positive transgenic embryos have been yielded.

1.2.7 β AGG20 line

In order to monitor the expression driven by the ~20kb fragment in β AGG20 –with the missing 92 base pairs in the minimal/basal promoter- throughout adulthood, I created a line. To find transgenic expressors, I took tail tips of 5-week-old weaned pups and stained them with the substrate X-gal. I identified one transgenic pup in which the nucleus pulposus of the intervertebral discs, the superficial proliferating chondrocytes as well as a few deeper hypertrophic chondrocytes of the growth plate of the vertebral bodies stained blue. There was no blue staining in the annulus fibrosus or the cartilage end plate. Figure 29 shows images (Figure 29A) and sections (Figure 29C + E) from this 5-week-old transgenic mouse's tail tip. A wildtype tail was also stained with X-gal to identify background endogenous β -galactosidase expression in bone (Figure 29B). The positive expressor that was identified was bred to produce a homozygous line named line 12 (after the founder). Tail tips of 5-week-old wildtypes were also stained with the BR1 antibody to compare endogenous aggrecan expression with that of the reporter gene in line 12. Endogenous aggrecan expression was similarly found to be present within the nucleus pulposes of the intervertebral discs as well as the proliferating and hypertrophic chondrocytes of the surrounding growth plate cartilage (Figure 29D + F). However, unlike the pattern of expression seen in the X-gal stained transgenic tail tips of line 12, there was some antibody staining present in the annulus fibrosus and the cartilage end plate of the wildtype tail tips and the staining present in the hypertrophic chondrocytes was more intense.

Figure 29 (next page): Tail tips of 5-week-old weaned pups stained with the substrate X-gal. A, C, D, E and F are tails tips from line 12 β AGG20 transgenics while B, G, H and I are tail tips from wild type mice. A + B) whole mount X-gal stained tail tips of a line 12 β AGG20 transgenic (A) and a wildtype littermate (B). C) 4x magnification of eosin counterstained section of line 12 β AGG20 transgenic tail tip. D) 4x magnification of a BR1 antibody stained wildtype tail tip. E) 10x magnification of eosin counterstained sections of line 12 β AGG20 transgenic tail tips. F) 10x magnification of a BR1 antibody stained wildtype tail tip. AF (annulus fibrosus), CE (cartilaginous end-plate), GP (growth plate), IV (intervertebral disc) NP (nucleus pulposus), VB (vertebral body). Black arrows are pointed to blue stained β -galactosidase positive cells/regions. Red arrows are pointed to brown BR1 antibody stained cells/regions.



Next, I planned an experiment in which I could monitor the activity of the regulatory ~20kb sequence in line 12 in a 4-month-old adult mouse while verifying its activity at the embryonic stage (already tested). A 4 month-old homozygous male was crossed with 2 F1 females. The females had plugs on two consecutive days and were sacrificed approximately 2 weeks later alongside the homozygous male. Along with the E13.5 and E14.5 embryos retrieved from the pregnant mothers, the heart, lungs, kidneys, brain, liver, spleen, spine, foot pad, thorax, hips, knees and tail of the homozygous male were also dissected and stained overnight with X-gal.

As expected, the embryos at various stages of development (Figure 30.1 A-D) were all positive for cartilage-specific expression and with the exception of 1 embryo, were comparable to the 1st 'βAGG20' transgenic embryos acquired through microinjections. This particular embryo (Figure 30.1A, embryo on the left +Figure 30.1C) had blue staining in the nasal cartilage which was probably due to it being slightly more developed than the E14.5 βAGG20 transgenic embryos generated by microinjections (Figure 17). Regarding the adult male, the heart, lungs, kidneys, brain, liver, spleen, spine and thorax were all negative (data not shown). Figure 30.2 A-D shows wholemount images of the knee (Figure 30.2A), footpad (Figure 30.2B), tail (Figure 30.2C) and hip (Figure 30.2D) from this adult male. There was no blue staining in the footpad. However, the hips, knee and tail all showed some positive staining. Histological sections of these tissues were prepared in order to verify whether the macroscopic blue staining seen in the different tissues was due to *LacZ* transgene expression or simply due to endogenous β-galactosidase expression that occurs in bone marrow tissue.

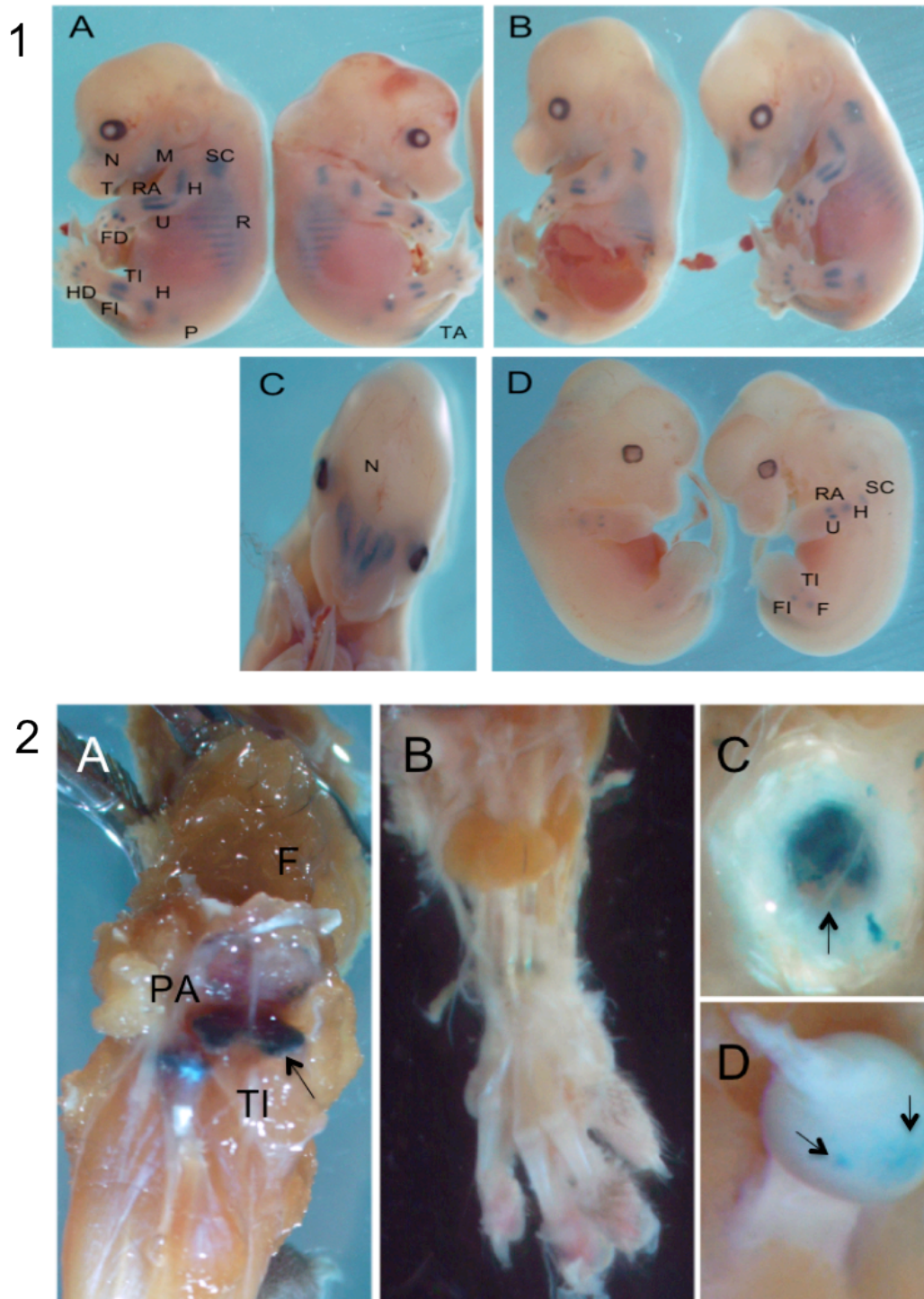


Figure 30.1 + 30.2 (next page): 30.1) Wholemount X-gal blue staining of heterozygous E12.5-E14.5 embryos from line 12 (with β AGG20 transgene). A and B) lateral images of E14.5 line 12 embryos C) Ventral image of E14.5 line 12 embryo D) lateral images of E12.5 (left) and E13.5 (right) line 12 embryos. **30.2) Wholemount X-gal blue staining of homozygous 4 month old adult mouse from line 12 (with β AGG20 transgene).** A) Knee of 4 month old adult from line 12 B) Footpad of line 12 4 month old mouse C) Cross section of tail from line 12 4 month old mouse D) Hip of 4 month old adult from line 12. N (nasal cartilage), F (femur), FD (forelimb digits), FI (fibula), HD (hindlimb digits), M (Meckel's cartilage), PA (Patella), P (Pelvis), R (rib cartilage), RA (radius), SC (scapular cartilage), T (tongue), TA (tail), TI (tibia), U (ulna).

Figure 31 shows sagittal histological sections of the tail and hip shown in Figure 30.2C + 30.2D respectively. These sections were also counterstained with the BR1 antibody in order to distinguish between the endogenous expression of the gene as well as the expression driven by the 20kb transgenic sequence in line 12. The tail section in Figure 31A showed similar blue staining to what was previously seen in Figure 29. As was the case before, the blue X-gal staining and the brown BR1 antibody staining colocalised in the nucleus pulposus and the proliferating + hypertrophic cells of the vertebral growth plate. However, there was only BR1 antibody staining in the annulus fibrosus and the cartilage end plate. There was a big difference between the BR1 antibody staining and the blue staining in the hip sections. In Figure 31B, only 3 of the deeper chondrocytes showed blue staining whereas all the chondrocytes were positive for BR1 antibody staining.

The coronal sections of the mouse knee in Figure 30.2A are shown in Figure 32. There was no blue staining evident in the lateral or medial articular cartilage and there was very minimal blue staining in the deep hypertrophic chondrocytes of the medial and tibial growth plates (Figure 32D+E).

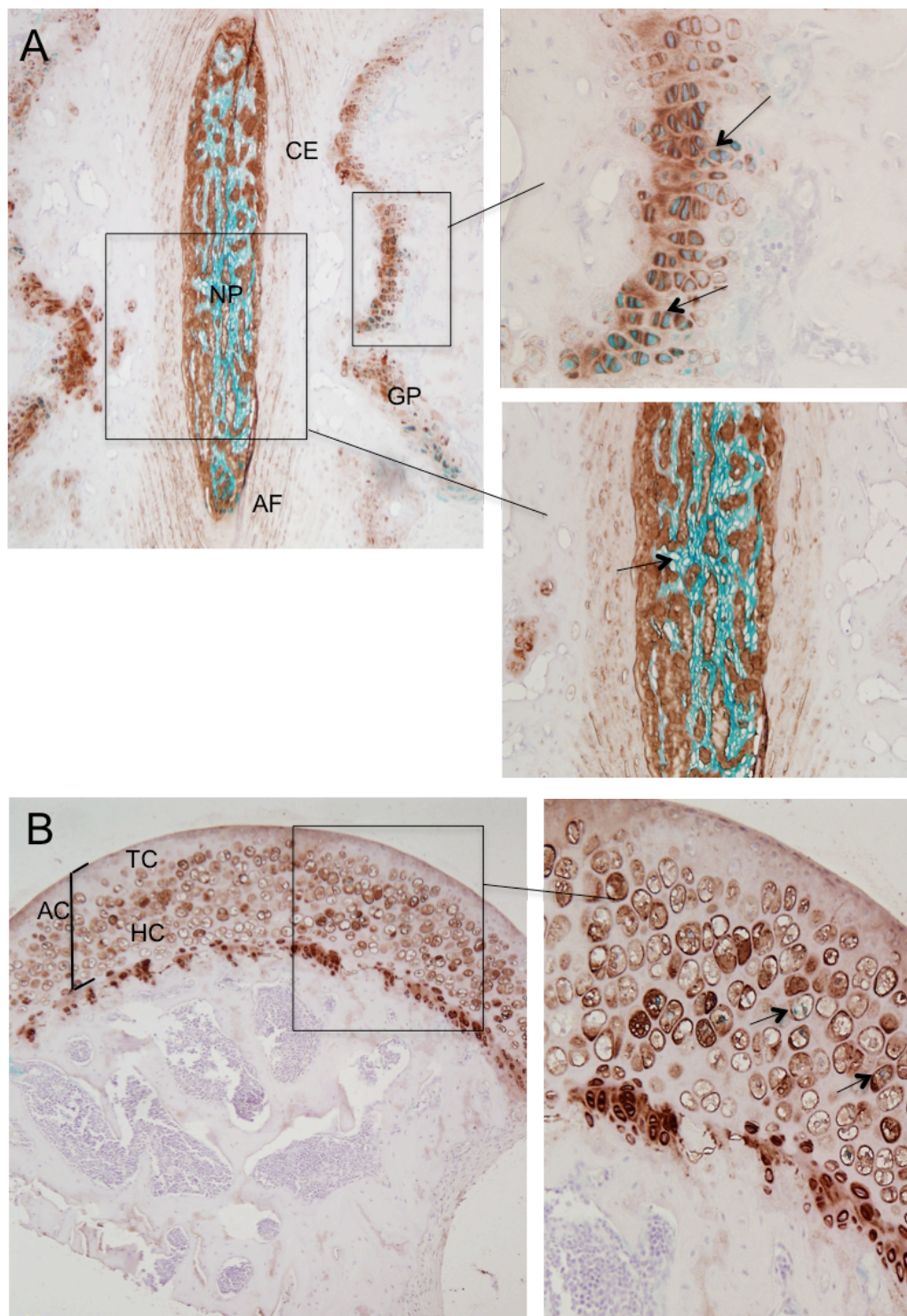


Figure 31: Paraffin sections of the line 12 (β AGG20 transgene) mouse tails (A) and hip (B) stained with the X-gal substrate and counterstained with the BR1 antibody. A) 10x magnification of a transgenic tail tip section (left) with corresponding 20x magnification of boxed regions (right). B) 10x magnification of a transgenic hip section (left) with corresponding 20x magnification of boxed region (right). AC (articular cartilage), AF (annulus fibrosus), CE (cartilaginous end-plate), GP (growth plate), HC (hypertrophic chondrocytes) NP (nucleus pulposus), TC (terminal chondrocytes). Black arrows are pointed to blue stained β -galactosidase positive cells/regions.

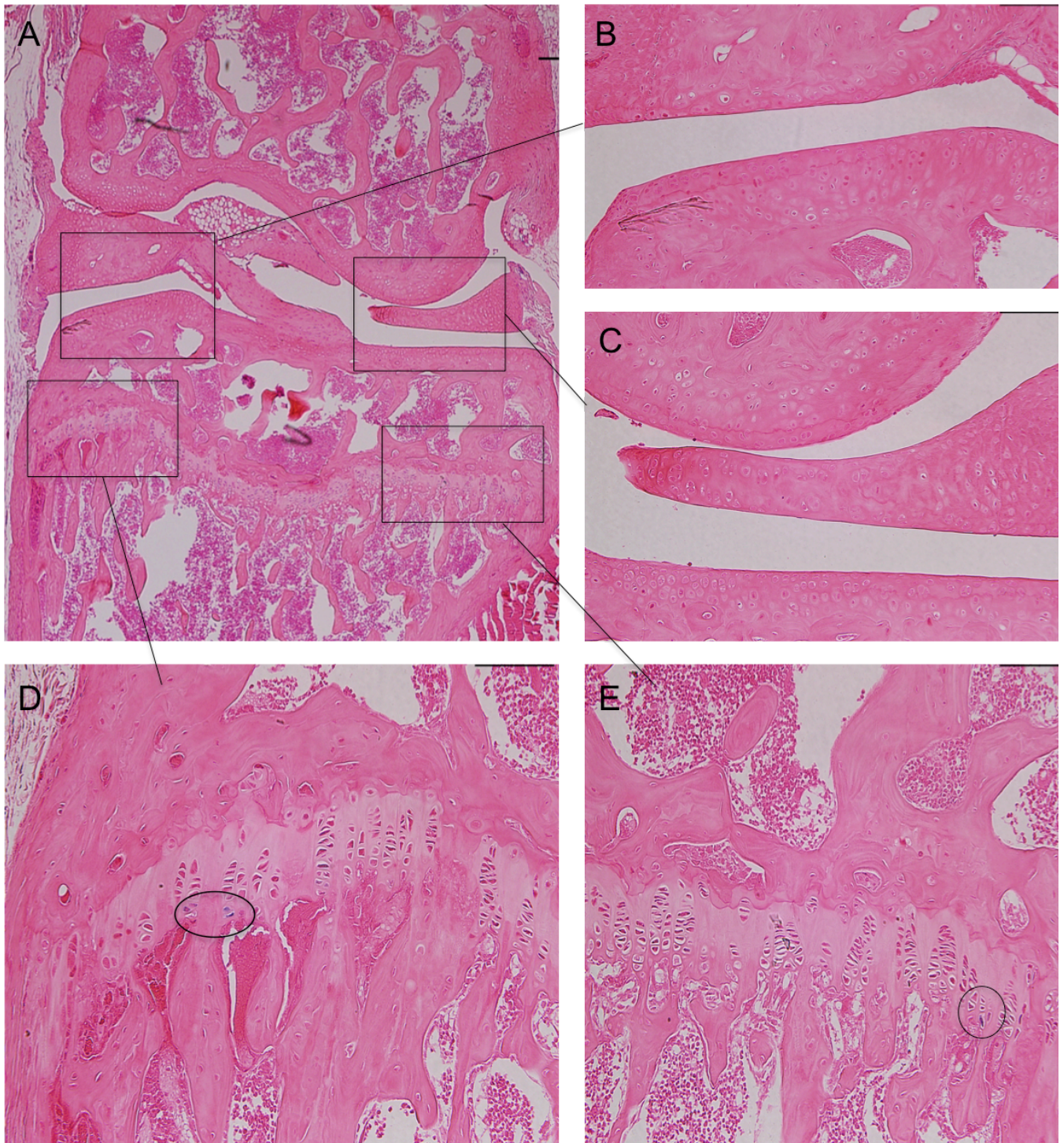


Figure 32: Coronal sections of a representative 4 month old line 12 (β AGG20) mouse knee in which the hypertrophic blue-stained chondrocytes have been circled. A) 4x magnification of whole knee B) 20x magnification of lateral articular cartilage C) 20x magnification of medial articular cartilage D) 20x magnification of lateral tibial growth plate E) 20x magnification of medial tibial growth plate. Section was counterstained with eosin.

1.2.8 Summary of Results

Table 21: Summary of mice transgenic and expressing all generated constructs for microinjection. 'Positive embryos' indicates the total number of embryos in which the transgene has become integrated throughout each session. 'Expressing embryos' indicates the total number of embryos in which the integrated reporter gene is being expressed. The 'Percentage of embryos expressing the transgene' was derived by dividing the number of expressing embryos by the number of positive embryos and multiplying this figure by 100.

Construct	Number of embryos positive by PCR	Number of embryos expressing the transgene	Percentage of embryos expressing the transgene (%)
β galAB	10	0	0
β AGG20 (E14.5)	7	3	42
β AGG20 (E16.5)	3	1	33
β AGG13	11	0	0
β galAB+7	21	2	10
β AGG10	18	0	0
β galAB+ <i>cis</i> 3	8	1	13
β galAB ⁺⁹² + <i>cis</i> 3	11	4	36
β AGG20 + <i>cis</i> 6	18	3	17

Table 22 (next page): Average levels of expression for all transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

a staining only found in 1 out of 4 expressing embryos

b staining of the ribs found in 3 out of 4 expressing embryos, one of which only shows staining in the distal part of the developing ribs

c staining found in 2 out of 4 expressing embryos

d staining found in all 4 expressing embryos, but localised to the first and fifth digit in one of the expressing embryos

e staining found in 3 out of 4 expressing embryos

f staining found in 3 out of 4 expressing embryos, one of which shows high level staining compared to low level staining in the other 2

Tissue	β galAB	β AGG20 (E14.5)	β AGG20 (E16.5)	β AGG 13	β galAB +7	β AGG10	β galAB+cis 3	β galAB ⁺⁹² +cis 3	β AGG20+cis 6
Chondro- cranium	-	-	-	-	-	-	-	-	-
Cartilage of the eye socket	-	-	+++	-	-	-	-	+ a	-
Nasal cartilage	-	-	+++	-	++	-	++	-	-
Meckels cartilage	-	+	++	-	-	-	-	-	-
Cartilage of the ribs (costal)	-	+++	+++	-	-	-	-	+ b	-
Cartilage of the femur	-	++	+++	-	+	-	-	+ c	-
Cartilage of the tibia	-	+++	+++	-	+	-	-	++ c	++
Cartilage of the fibula	-	+++	+++	-	-	-	-	++ c	-
Digits of the forelimbs	-	+++	++++	-	++	-	++	++ d	++
Digits of the hindlimbs	-	++	++++	-	-	-	-	- e	+
Scapular cartilage	-	+++	+++	-	-	-	-	+ e	-
Cartilage of the humerus	-	+++	+++	-	-	-	-	+ c	-
Cartilage of the ulna	-	+++	+++	-	+++	-	-	++ f	++
Cartilage of the radius	-	+++	+++	-	+++	-	-	++ f	++
Intervert- ebral discs	-	-	+++	-	-	-	-	-	-
Vertebral cartilage	-	-	+++	-	-	-	-	-	-
Cartilage of the ear	-	-	+++	-	+	-	-	+ a	-
Tongue	-	+++	-	-	-	-	-	-	-

1.3 Discussion

1.3.1 Using transgenic mouse embryos as a model to study the regulation of the aggrecan gene

There are several reasons for my choice to use transgenic mouse embryos rather than transient-transfection assays to study the regulation of aggrecan gene expression. Firstly, there are a lot of published data demonstrating the accurate correlation between the endogenous expression of genes and their expression in transgenic mouse embryos in which transgenes have been introduced alongside a reporter gene such as *LacZ*. An example of this comes from a study carried out by Whiting J et al. using a Hox-2.6-lacZ reporter gene in transgenic mice; they showed that the overall expression pattern of the endogenous Hox-2.6 gene could be reconstructed. The transgene was expressed with the correct spatial distribution and temporal pattern. Furthermore, direct comparison by *in situ* hybridization revealed that the levels of transgene expression were also similar to those of the endogenous gene (Whiting et al., 1991).

Secondly, the use of transient-transfection assays is very limited due to the restrictive nature of primary chondrocytic cell lines. Immediately after isolation from their natural protective environment, chondrocytes show hardly any proliferative activity. If they do however start to proliferate, this usually indicates the loss of a differentiated phenotype. This involves a change in cell shape as well as significant differences in gene expression pattern. Even the immortalised chondrocyte cell lines C-28/I2, T/C-28a2 and T/C-28a4, which have become a common tool in cartilage research, are still quite problematic since they show less expression of genes of matrix synthesis and turnover (Finger et al., 2003). Analysing the expression of aggrecan *in vitro* will therefore be quite challenging with culture conditions differing greatly from the *in vivo* conditions that make transgenic mouse experiments so useful.

Thirdly, transgenic mouse experiments shouldn't take longer than transient- transfection assays. It takes 14.5 days to grow transgenic embryos to the level at which they can be analysed for aggrecan gene expression. This is roughly the same amount of time it takes to attain a transiently transfected cell line. And finally, with transgenic experiments, you can visualise the expression of a gene throughout the organism as a whole in a temporal and spatial manner. For example, it is possible to determine the different cartilaginous tissues in which aggrecan is expressed and the time-points at which they begin to show this expression.

1.3.2 *In silico* analysis of the aggrecan gene

The web-based program Vista was used to look at identities between different non-coding sequences of the aggrecan gene across two different species, rats and humans. This seems to be a good way to identify evolutionarily conserved and therefore important regions within the genome. In support of this, in a study carried out by Pennacchio et al, extreme evolutionary sequence conservation was used as a filter to identify putative gene regulatory elements. This led to the characterisation of the *in vivo* enhancer activity of a large group of non-coding elements in the human genome that are conserved in human vs pufferfish fugu (*Takifugu rubripes*), or ultraconserved in human vs mouse vs rat. Of the 167 extremely conserved sequences tested in transgenic mouse enhancer assays, 45% functioned reproducibly as tissue-specific enhancers of gene expression at embryonic day 11.5 (Pennacchio et al., 2006). This serves as an argument for the investigation of human DNA regulatory sequences in a mouse system.

When comparing human and rat aggrecan non-coding sequences, there are several relatively highly conserved sequences to be found upstream of and around the transcriptional start site. For my purposes, these regions were named *cis* 1-5. *cis* 1-4 lie upstream of the start site and are in the approximate positions '-20- -19.7kb', '-14.5 - -14.4kb', '-10.9 - -10.2kb', and '-3.8 - -

3.7kb' respectively. *cis* 5 (-390 - +375bp) lies around/includes the 1st exon/5'UTR/previously determined basal promoter of the human aggrecan gene. Unlike *cis* 1-5, which are relatively highly conserved as indicated by the coloured peaks in Figure 14, *cis* 6 has a lower level of conservation between humans and rats and was selected due to the presence of 3 putative *sox9* binding sites within it. It is located at the 5' end of the large 1st intron of the gene (between +5 and +8kb relative to the TSS). The *sox9* putative binding sites were identified using the program genomatrix (<http://www.genomatix.de/>) and are located at +5.3kb, +6.8kb and +7.4kb relative to the TSS. Finally, *cis* 7 is positioned within the 14th intron of the human aggrecan gene between +64.6kb and +65.0kb relative to the TSS. My interest in *cis* 7 was due to the fact that it was the only sequence within the investigated region in which the conservation down to fugu remained relatively high (Figure 15).

1.3.3 The human aggrecan minimal promoter within *cis* 5 (-452bp - +375bp)

As discussed, the human aggrecan minimal promoter used was based on the data published by Valhmu et al along with the information from my *In silico* analysis in which *cis* 5 includes this minimal promoter. Although the published data points to the 5'UTR (0bp - +375bp relative to the TSS) of the gene as the primary regulatory sequence able to act as the minimal promoter required to drive basal expression of the aggrecan gene, the vista analysis shown in Figure 14 showed high sequence identity between humans and rats from -390 - +375bp. Thus, as the human aggrecan minimal promoter, I used the sequence corresponding to -452 - +375bp, which was named AB. This ensured the inclusion of the 5'UTR as well as the upstream conserved sequence. Although the 5'UTR was sufficient to drive reporter gene expression by about 8 fold in chondrocytes and 40 fold in NIH 3T3 cells (*in vitro*) (Valhmu et al., 1998a), this was not the case for the transgenic embryos harbouring this sequence (*in vivo*). There was no blue staining in any of the positive transgenics (Table 6). This result corresponded somewhat

with the data from Han and Lefebvre (2008) where the 'mouse aggrecan minimal promoter' corresponding to the sequence -596bp - +90bp relative to the mouse TSS was completely inactive when tested alone in RCS cells (Han and Lefebvre, 2008). This sequence was most probably selected as a 'basal promoter' due to it containing two glucocorticoid receptor-binding sequences (TGTTCT/C) and one GGGCGG sequence (SP-1 site) as well as several homologous direct repeat sequences. In addition to this, a region that shows identity to a sequence of the promoter of the rat type-II collagen gene, which itself is highly conserved in both the rat and mouse type-II collagen genes and is known to be important for type-II collagen gene promoter activity (Watanabe et al., 1995). It would be interesting to test whether the activity of this promoter is switched on *in vitro* with the inclusion of the mouse 1st exon/5'UTR which is 509bp in size (+1 - +509bp relative to the mouse TSS). Nevertheless, the combination of this data supports the conclusion that the selected human aggrecan minimal promoter AB is unable to drive reporter gene expression *in vivo*. At this point, it was therefore assumed that any reporter gene activity was due to sequences that lie outside this minimal promoter interacting with it to drive transcription of the gene.

1.3.4 Investigating the potential regulatory activities of *cis* 1-4

1.3.4.1 cis 1-4 (located at -20 to -19.7kb, -14.5 to -14.4kb, -10.9 to -10.2kb, and -3.8 to -3.7kb respectively).

At the time I started this work, the only *in vivo* experiment that had been performed in order to locate *cis*-regulatory 'enhancer' sequences was by Doege and colleagues whose work focused on the rat aggrecan gene (Doege et al., 2002). They showed that an *Agc1* 4.7kb (P3) upstream region was able to activate an *Agc1* 3.4kb (P0) promoter in transgenic embryos, but weakly and only in hypertrophic chondrocytes in a subset of cartilage elements. Taking this into account as well as the *in silico* homology search in Figure 14, I cloned a DNA fragment

around 20 kilobases in size (ranging from -20,054bp to +375bp relative to the TSS) into the β gal-basic vector upstream of the *LacZ* reporter gene. This included *cis* 1-4 as well as the minimal promoter AB within *cis* 5. By cloning and analyzing such a large fragment, I also took into account any shifting of *cis*-regulatory elements that may have occurred during evolution. It is important to note that any statements made below assume that the *cis*-regulatory sequences that drive the expression of the human aggrecan gene endogenously fall only within the highly homologous sequences illustrated by the vista analysis, *cis* 1-5 (Figure 14).

The vector containing *cis* 1-5 was named β AGG20 and microinjected into the pronuclei of fertilised murine oocytes. The results of the transgenic mouse experiments revealed the existence of *cis*-acting elements other than the basal promoter upstream of the TSS of the aggrecan gene that are responsible for the chondrocyte specific expression of the gene. Approximately 40% (3 embryos) of the positive transgenic mice expressed the transgene identically, symmetrically and specifically in the developing cartilage/differentiated chondrocytes of the long bones of the hind and fore limbs, the ribs, the scapulae, some facial cartilaginous structures (meckel's cartilage) and the digits (Figures 17). The histological sections in Figure 18 showed that the β -galactosidase expression was present in both the proliferating and hypertrophic chondrocytes of the developing cartilaginous structures. The presence of the β -galactosidase protein in the hypertrophic chondrocytes is most probably due to the stability of the protein and the progression of the cells. Henry et al (2009) also came across this when they compared aggrecan mRNA expression (detected by RNA *in situ* hybridization) that was present in the proliferating chondrocytes of the ulna growth plate but was not observed in the hypertrophic chondrocytes with the expression of aggrecan in a reporter mouse in which aggrecan expression coincides with *LacZ* expression. In this latter mouse, positive X-gal staining was observed in both the proliferating and hypertrophic chondrocytes of the ulna (Henry et al., 2009). The observation of the same expression pattern

in three of the positive β AGG20 transgenic embryos makes the possibility of insertional position artefacts extremely unlikely.

It is obvious that the pattern of transgene expression is different when comparing the transgenic mice expressing β AGG20 with those expressing the P3/P0 transgene (Doege et al., 2002). The β AGG20 transgenics displayed a more intense and spacially increased pattern of transgene expression when compared with the P3/P0 transgenics. Unlike the P3/P0 transgenics, they expressed *LacZ* in the developing cartilage of the digits and in addition to hypertrophic chondrocyte expression, β -galactosidase expression was present in proliferating chondrocytes. From this, I propose the following of which at least one statement is possible:

- 1- This ~20kb fragment must have additional necessary *cis*-regulatory elements that the rat -12kb - -7kb/basal promoter fused DNA does not.
- 2- Enhancer element(s) found within P3 might require a minimal distance from the P0 element in order for it/them to function normally. Since the ~20kb fragment in β AGG20 was not modified from what is present within the human genome, the location and relative distances between the 'regulatory sequence(s)' in the construct will therefore reflect what is true endogenously.
- 3- There may be other elements such as silencers within P3 or P0, which weaken the regulatory activity of the entire sequence. These silencers could be neutralised by the presence of other sequences within the ~20kb fragment.

1.3.4.2 cis 3+4 (-10.9 to -10.2kb and -3.8 to -3.7kb respectively).

In order to minimise the size of the DNA fragment that can drive the expression of aggrecan specifically in cartilage. β AGG20 was shortened by ~7kb from its 5'end (generating β AGG13) and used to produce transgenic mice. In doing this, the DNA sequence containing *cis* 1 + 2

was removed leaving behind *cis* 3 + 4 along with the minimal promoter AB, within *cis* 5. This cut off point was primarily selected in order to include the ~400 bp sequence with high identity across humans and rats that is present within *cis* 3 and the fragment P3 (Doege et al., 2002) located between -10.7kb and -10.4kb relative to the human aggrecan TSS.

Despite the generation of 11 positive β AGG13 transgenic embryos, none showed any form of X-gal staining (Table 11). On its own, the ~13kb stretch of DNA did not seem to drive transcription of the reporter gene. Therefore, with regards to *cis* 3+4, it was possible that one or both of these regions were either non-essential (are not real *cis*-regulatory sequences), necessary regulatory sequences but insufficient (require other fragment(s) to interact with) or negative regulatory sequences that downregulate the activity of the reporter gene.

1.3.4.3 cis 1+2 (-20 to -19.7kb and -14.5 to -14.4kb respectively).

To test whether the regions corresponding to *cis* 3 and/or *cis* 4 were either insufficient or unnecessary, the deleted 7kb fragment containing *cis* 1 + 2 was fused immediately upstream of the minimal promoter AB (generating the construct β galAB+7) and used to produce transgenic mice. A similar embryonic expression pattern to that generated by the β AGG20 transgene would indicate that *cis* 3 and/or *cis* 4 are unnecessary while a different expression pattern would indicate their insufficiency.

The results of generating E14.5 β galAB+7 transgenic embryos revealed the existence of *cis*-acting elements other than the basal promoter within the fused 7kb stretch of DNA containing *cis* 1 and 2. Approximately 10% (2 embryos) of the positive transgenic mice showed similar and symmetrical cartilage specific expression in the cartilage of the long bones of the forelimbs (ulna and radius), the nasal cartilage and minimal expression in the cartilage of the developing hindlimbs (femur and tibia) (see Figure 22 and Table 12+13). Although the expressing

β galAB+7 transgenic embryos showed cartilage-specific expression of the transgene, the *LacZ* expression in the β AGG20 transgenics was more intense with a wider range of positive developing cartilaginous structures (digits, ribs, scapulae, etc). It was also clear when comparing the histological sections of the β galAB+7 transgenic embryos (Figure 23) with the β AGG20 transgenic embryos (Figure 18) that although there was also blue *LacZ* staining in the proliferating and hypertrophic chondrocytes of the β galAB+7 embryos, this staining was more scattered/mosaic across the cells with far fewer cells stained blue. It was possible, therefore, that although the 7kb stretch of DNA containing *cis* 1 and 2 had a regulatory sequence(s) that could drive cartilage specific expression of the transgene, other regulatory sequences within β AGG20 which could correspond to *cis* 3 and/or 4 were also necessary but probably insufficient. This is further supported by the fact that only 10% of the positive β galAB+7 transgenic embryos expressed the transgene compared with 40% for the β AGG20 transgenic embryos. From this, I propose the following of which at least one statement is possible:

- 1- This ~7kb fragment must have necessary *cis*-regulatory element(s) to drive endogenous aggrecan expression.
- 2- Like the *cis*-regulatory elements in β AGG13 (*cis* 3+4), these elements are necessary but insufficient to drive endogenous cartilage specific aggrecan expression.

1.3.4.4 cis 1+2 and 4 (-20 to -19.7kb, -14.5 to -14.4kb and -3.8 to -3.7kb respectively).

The fact that the blast alignment of the P3 and P0 rat fragments (Doege et al., 2002) with the β AGG20 sequence revealed a highly conserved approximately 400 bp sequence within *cis* 3 led me to believe that it was *cis* 3 that was the relevant and necessary regulatory sequence within β AGG13 that potentially interacted with the regulatory sequence(s) in β galAB+7. To test this, I generated a construct (called β AGG10) lacking the *cis* 3 sequence but containing the *cis*

1,2 and sequences along with the minimal promoter AB by removing a fragment of DNA approximately 10kb in size within the 20kb sequence in β AGG20 between $\sim$ -14kb and -4kb relative to the TSS. Out of 18 positive E14.5 transgenic embryos, none showed any reporter gene expression as indicated by the lack of blue staining (Table 14). It is important to note that unlike β galAB+7, which was able to drive minimal cartilage specific expression in 10% of the positive transgenic embryos, β AGG10 (containing all the fragments present in β galAB+7 in addition to *cis* 4) showed expression in 0% of the positive transgenic embryos. From this I concluded that:

- 1- The deleted 10kb sequence contains a potential *cis*-regulatory element (*cis* 3) that is necessary but insufficient to drive endogenous expression of aggrecan.
- 2- It is possible that *cis* 4 has a negative regulatory effect on the expression of the reporter gene.

1.3.4.5 *cis* 3 (-10.9 to -10.2kb)

Given the negative results from the β AGG10 transgenics and the fact that the blast alignment between the P3/P0 rat fragments (Doege et al., 2002) and the β AGG20 sequence revealed a highly conserved $\sim$ 400bp sequence within *cis* 3 (between -10.7kb and -10.4kb), it was highly likely that *cis* 3 could be indispensable for the endogenous expression of aggrecan. At this point in my PhD, Veronique Lefebvre's group published data on the mouse aggrecan promoter that showed that a short regulatory sequence, which they named A1 and that corresponded to the $\sim$ 400bp sequence that I identified within *cis* 3, could function as an enhancer that can drive cartilage specific expression of a reporter gene (Han and Lefebvre, 2008). However, although they showed that when multimerised (x4), this sequence was able to direct expression of the *Col2a1* minimal promoter (-309 bp - +308 bp relative to the TSS) in both embryonic and adult cartilage in transgenic mice in a manner that matches aggrecan expression, they showed no

evidence that this was true for the *Agc1* minimal promoter fused to 1 copy of the A1 sequence. The only data shown on the gene regulation driven by the A1 sequence fused to the *Agc1* minimal promoter was in RCS cells in which they showed that the *Agc1* promoter was weakly but significantly activated by 1 copy of the A1 sequence and more powerfully activated by 4 copies of A1.

In order to test the regulatory role of the highly conserved ~400bp sequence within *cis* 3, I constructed a clone containing the minimal promoter AB fused to the highly homologous 400 base pair fragment within *cis* 3. The construct was named ' β galAB+*cis* 3'. Out of 8 transgenics, 1 embryo (12.5%) showed minimal cartilage specific expression limited to the nasal cartilage and the cartilage of the distal digits of the forelimb (Table 15 + 16 and Figure 24). From this, I concluded that *cis* 3 definitely is a *cis*-regulatory sequence that is necessary but insufficient as only 12.5% of the positive ' β galAB+*cis* 3' transgenic embryos expressed the transgene compared with 40% for the β AGG20 transgenic embryos. This expression was also a lot weaker and limited to 2 developing cartilagenous elements.

1.3.4.6 Hypothesised roles of *cis* 1-5

From the above data, I hypothesized the following:

- 1- *cis* 5, in which the minimal promoter of the human aggrecan gene lies, is unable to drive reporter gene expression *in vivo* and requires other regulatory sequences to interact with in order to drive the endogenous expression pattern of the aggrecan gene.
- 2- *cis* 4 is a negative regulatory sequence (silencer) as is shown by the fact that unlike β galAB+7 (containing *cis* 1,2 and 5), which was able to drive minimal cartilage specific expression, β AGG10 (containing all the fragments present in β galAB+7 in addition to *cis* 4) showed no expression in any of the positive transgenic embryos. Also, β AGG13,

which contained *cis* 3,4 and 5 showed no expression of the transgene unlike β galAB+*cis* 3 (*containing cis 3 and 5 and excluding cis 4*), which showed minimal cartilage-specific expression in 12.5% of the positive transgenic embryos. This is further supported by the paper by Pirok et al in which the deletion of two copies of the sequence CACCTCC (CIIS2) approximately 870 and 720 base pairs upstream of the chicken aggrecan start site reduced the cell type specificity of reporter gene expression but increased the overall promoter activity (Pirok et al., 1997).

- 3- *cis* 3 is an enhancer sequence that is necessary but insufficient to drive endogenous aggrecan expression as is shown by Han and Lefebvre's (2008) work, the ability of the β galAB+*cis* 3 transgenic embryo to drive cartilage specific expression and the fact that none of the β AGG10 transgenics (which excluded *cis* 3) drove the expression of the reporter gene.
- 4- *cis* 1 and/or *cis* 2 are also necessary but insufficient to drive the reporter gene expression seen in the β AGG20 transgenic embryos as is shown by their ability to activate the minimal promoter in the β galAB+7 transgenics and the inability of β AGG13 (which excludes *cis* 1 and 2) to drive reporter gene expression. They could possibly interact with *cis* 3 to drive full endogenous aggrecan expression and counteract the negative regulatory effect that *cis* 4 has on expression of the reporter gene.

To test this further, it would be useful to create a construct composing of *cis* 1,2,3 and 5 and excluding *cis* 4.

1.3.5 Finding additional *cis*-regulatory elements in the aggrecan gene

Although these results clearly confirmed a major role for the ~20kb fragment containing *cis* 1-5 in the tissue-specific regulation of aggrecan, there were some indications that it may not be sufficient for all aspects of aggrecan regulation. It was quite surprising that the reporter gene

expression in the vertebral bodies of the β AGG20 transgenic embryos was switched on as late as E16.5. Also, even at E16.5, there was no X-gal staining in the chondrocranium. The expectations of the temporal and spatial pattern of aggrecan gene expression were initially based on a comparison with the expression pattern driven by the type II collagen p3000i3020 promoter/enhancer in E14.5 transgenic mouse embryos (Zhou et al., 1995). However, the generation of the *Agc1*^{tm(IRES-CreERT2/+)} transgenics in De Crombrughe's laboratory (Henry et al., 2009), in which the *Cre recombinase* gene was knocked into the 3' untranslated region of the endogenous mouse aggrecan gene as well as the BR1 antibody stained mouse sections in Figure 13 enabled the use of other more comparable ~ results. This was necessary since there is a big difference in the intensity and a smaller difference in the spatial orientation of *LacZ* gene expression between the tamoxifen induced *Agc1*^{tm(IRES-CreERT2/+)}; *Rosa26* Cre reporter double transgenics and the 'p3000i3020' transgenic embryos at E14.5. There are different possible explanations for this. Firstly, it has been suggested that there would be differences in developmental expression between type II collagen and aggrecan (Sandell, 1994), (Han and Lefebvre, 2008). Also, even where these genes are coexpressed, there are differences of at least 5-fold in the steady-state levels of their mRNAs (Oxford et al., 1994). Therefore, we only used the *Agc1*^{tm(IRES-CreERT2/+)}; *Rosa26* Cre reporter double transgenic embryos + the BR1 antibody stained mouse sections as reliable comparative tools. The tamoxifen induced *Agc1*^{tm(IRES-CreERT2/+)}; *Rosa26* Cre reporter double transgenic embryos, in which the vertebral body expression of the *Cre recombinase* gene started as early as E13.5, showed more intense X-gal staining that was localized to a wider range of cartilaginous regions (such as the chondrocranium) in E15.5 embryos when compared to the E14.5 and E16.5 β AGG20 transgenic embryos. The BR1 antibody stained mouse sections also showed aggrecan staining in the vertebral bodies as early as E13.5 as well as staining in the chondrocranium at E15.5. It can therefore be concluded that the activity of β AGG20 may just represent a specialised subprogram for aggrecan's expression so there was a need to identify

additional areas of the gene, which can supplement its function.

1.3.5.1 Missing 92 base pair sequence

When β AGG20 was initially cloned, the sequencing of the minimal promoter AB within *cis* 5 on the 3' end of the ~20kb fragment showed a deletion of 92 base pairs (-68 to + 23bp) which was also found in the BAC RP11 343B18 used to retrieve human aggrecan promoter DNA. Fresh stocks of the BAC were ordered but unfortunately all contained this deletion. In the meantime, transgenic embryos were generated using the β AGG20 construct and they displayed cartilaginous expression that was similar but not identical to endogenous aggrecan expression. β AGG20 was further manipulated to produce other constructs some of which also drove cartilage specific expression of *LacZ*. On the one hand, the missing 92 base pairs could be inconsequential. This is somewhat supported by the work done by Valhmu et al to elucidate the role of the minimal aggrecan promoter by using luciferase reporter constructs driven by the human aggrecan promoter (taken as -701bp - +52bp relative to the human aggrecan TSS, 87,147,670 on chromosome 15, Assembly: March. 2006 NCBI36/hg18). The sequential deletions of a large portion of this so-called human aggrecan proximal promoter from -701bp - -52bp had very little effect on promoter activity when the 5'UTR (+1 - +375bp) was present. However, when the 5'UTR was removed, leaving a 5' flanking sequence of -701bp - +25bp or -52bp - +25bp, luciferase activity was reduced 85-100%. It is possible that the whole sequence from -701bp up to +25bp, which includes the missing 92 bp, is unnecessary since, on its own, it was not able to drive substantial luciferase activity. On the other hand, the sequence between -52bp and +25bp could be necessary but insufficient. This is more likely as the 5'UTR, which can act as the minimal promoter required to drive basal expression of the aggrecan gene, does include the sequence between +1 and +25bp. Therefore, within the

missing 92bp (-68 to +23bp), 23bp are part of the 5'UTR and have been excluded from β AGG20 and all subsequent clones generated.

Within the human aggrecan promoter, there are several putative SP-1/AP-2 binding sites (CCCGCC), including a cluster of SP-1/AP-2 binding sites around the TSS (Figure 33). SP-1 plays a key role in regulating transcription initiation of TATA-less and CAAT-less promoters by possibly recruiting the TFIID protein to the transcription initiation complex. Three overlapping SP-1/AP-2 binding sites span the TSS of the human aggrecan promoter. One of these overlapping sites is present within the missing 92 base pairs, whereas the second and third are located at positions -142bp and -246bp. In the rat gene, there is only one copy of the SP-1/AP-2 site in a position equivalent to that of the overlapping SP-1/AP-2 sites in the human promoter (Doege et al., 1994).

It is possible that the minimal promoter in all the constructs that I have produced was able to compensate for the lack of one of the overlapping SP-1/AP-2 sites due to the presence of two others. It is clear from the rat promoter that the presence of one overlapping site is sufficient for gene expression. Having said that, in both the rat and human genes, transcription is initiated 23-24bp downstream of a conserved TATG direct repeat (located at -31bp in the human gene), which could be important for the correct transcriptional initiation of the gene as this sequence may be the equivalent of a TATA box in TATA-less promoters such as the aggrecan promoter (Figure 33). However, the bovine aggrecan gene can undergo transcriptional initiation at four different sites upstream of this TATG repeat suggesting that the promoter can find alternative routes to initiate transcription (Watanabe et al., 1995).

```

-701          ACTGTAGAGACACTGGAGAATTCCTGCTTCATATGAGATGTCCTTGGTCTCTCCCGGGGAATTTGAAGACC
                                     GR/PR          SSRE
-630 CAGAGACTCGCAGCTCCCGTGAGCCTCCCGCTGACGCCTTACCTCCCGTCTCCACCATCCCCTGCCATCCCGGTGTACGCCACCCCC
          AP-4          AP-1/CREB          SIF
-540 AAAAAAATTTTACGCCAGCTCCTCGGTGAGGGTCGTTCAAATCGTGGCCGACAGTAGCGGGCTGCACCCTCCTTAGATCGCGTCGTGGC
          NF-kB          Gfi-1
-450 CAGCCTCAGCTGCAGAACCCCGCCGGGCGCCCGGAGCCCTGTCCCGCCCGCGCGGGCTCAGAGTCCGGGCACTTGGGGATTCTCTGAAG
          AP-4          AP-2          AP-2          SP-1          NF-kB          STAT
-360 GTGCAGCCCTCCTGGGCCCCCTCCGACTGCGGAGCTGCCGACCGCAATGCAGATGCGGGCCTTCCAGCCCTTCTCGCGCCGCTTCTCTCCC
          E47          E2F          AP-2
-270 GCCAGCCCCGGAGCTGCCGGGTCCGCGCCGCCCGGGGAGCGCTCCTCTCCCGCCCTGAGCGTAGGCCGGCTTCCCCATCGGCGCGCCGG
          AP-2/SP-1          SP-1
-180 TCCGAGCCAGGGTCCAGCGCGCTCCAGACGCTCTGCCTTCCCCCTCCCCCTGTCTGGCCGCCCTCGGTCCCCTGGGGGTGGGGTTTCCCT
          SP-1          AP-2          SP-1          NF-kB
-90  TTGCGCTCGCCCCCTCCCGCCGCCACCCCTCACGGGCCCTCCCTTCCCGGCCGTCCTATGTATGTGTACAGCGCGCCATGCCCGCC
          AP-2          AP-2          SIF          AP-1          AP-2/SP-1
+1   CGCCCGCCACCTACCTCCCGCCGCTCCAGAGGGGGCTCGCAGAGCTGAGGACGCGCGCAGCGCTGCTCAAGGTCTCTCTCTCAGCA
          AP-2/SP-1          δEF1          SP-1          SSRE
+91  CCCTCGCCGGCCGGCGTCTGACGCGGGTGCCAGGGTCTCCGGGCACCTTTTCAAGTGTCCATTCCCTCAGCCAGCCAGGACTCCGCAACCCA
          AP-1/CREB          SSRE          δEF1
+181 GCAGTTGCCGCTGCGGCCACAGCCCGAGGGGACCTGCGGACAGGACCGCGCAGGAGGAGGGGTGCGCAGCGCCCGCGCAGAGCGTCTCC
          AP-4/c-Myb
+271 CTCGTACGCAGCGAGACCCGGGCTCCCGGCCAGGAGCCCCAGCTGCCTCGCCAGGTGTGTGGGACTGAAGTTCTTGGAGAAGGGA
          SSRE          AP-4          USF/δEF1
+361 GTCCAACCTCTTCAAG

```

Figure 33: Sequence of the human aggrecan minimal promoter. The missing 92bp sequence in the RP11-343B18 BAC is shown in the purple box. Several putative binding sites for ubiquitous transcription factors such as SP-1, AP-2, and AP-4 are located in both the promoter and exon 1. Recognition sequences for growth factor- and cytokine-inducible factors are also present. The promoter lacks a TATA box, but a TATGTATG sequence that is present at -31bp (highlighted in black) might function as a TATA element. Adapted from (Valhmu et al., 1998a).

Having made all these points, it was still possible that the lack of this sequence due to the missing 92bp/SP-1/AP-2 sites in the constructs that I have generated may compromise the full expression of the gene in all cartilaginous elements of developing embryos. This notion is further supported by the work carried out by my supervisor George Bou-Gharios and his group on the Col1a2 promoter. They set out to examine the transcriptional contribution of SP-1 sites in the proximal promoter within the context of the interaction with the far upstream enhancer by mutating the three SP-1 binding sites of the cytokine responsive element within the proximal promoter. Transgenic embryos harbouring these mutations were found to express the *LacZ* reporter gene significantly less than the embryos harbouring the wild-type version of the proximal promoter/enhancer. Histological examination of the transgenic embryos harbouring

these mutations revealed that this was due to a mosaic pattern in the expression of the *LacZ* reporter gene. Also, some areas that were supposed to express the transgene showed no reporter gene expression. These results suggested that the SP-1 mutations impair transgene expression in addition to reducing the overall strength of the enhancer/promoter interaction for the *Col1a2* gene (Tanaka et al., 2004). The lack of one of the SP-1/AP-2 overlapping sites in the transgenics may therefore compromise the interaction between the aggrecan minimal promoter and the *cis* 3 enhancer or a possible regulatory sequence present within *cis* 1 or *cis* 2.

Validating a role for the missing 92 base pairs required the cloning of a new ~20kb fragment inclusive of this missing sequence within the vector backbone β gal basic in the same way. Although a new BAC containing the missing 92bp was finally acquired, I was not able to complete the recombineering experiments that are necessary to construct the clone due to a lack of time. I did, however, manage to construct a clone containing the minimal promoter AB including the missing 92bp sequence fused to the highly homologous 400 base pair fragment within *cis* 3 (called β galAB⁺⁹²+*cis* 3). Compared to the results from the β galAB+*cis* 3 transgenic, in which 1 embryo out of 8 positive transgenic (12.5%) showed minimal cartilage specific expression limited to the nasal cartilage and the cartilage of the distal digits of the forelimb (Table 16 and Figure 24), four of the 11 positive β galAB⁺⁹²+*cis* 3 transgenics (36%) showed a more intense cartilage-specific pattern of *LacZ* expression (Figure 25) that in two of the embryos, mimicked the expression pattern of the β AGG20 transgenics (Figure 17). However, although it mimicked this expression pattern, it was still less intense and not present within the tongue and the facial meckel's cartilage. Also, the regulatory activity driven by the β galAB⁺⁹²+*cis* 3 promoter was inconsistent and slightly different in all expressing transgenic embryos unlike the expressing β AGG20 transgenics which all had the same expression profiles. Furthermore, the histological sections of a β AGG20 transgenic embryo (Figure 18), in

which most of the proliferating and hypertrophic chondrocytes were stained blue are very different to a histological section of a $\beta\text{galAB}^{+92} + c/s\ 3$ transgenic embryo (Figure 26), in which the blue staining was much weaker and very scattered/mosaic in the proliferating and hypertrophic chondrocytes. There was also staining in non-chondrocytic fibroblastic cells, which is not what we would expect at this stage of development.

It is possible that along with the 7kb (-20kb to -13kb relative to the TSS and containing *c/s* 1 and 2) sequence's hypothesized role in interacting with *c/s* 3 to neutralize the negative regulatory effects that *c/s* 4 potentially has on aggrecan gene expression, it may likely contain binding sites that can partially compensate for the missing 92 base pairs. I hypothesized this because transgenic embryos that harboured the 7kb sequence along with the 92 base pair deletion still showed cartilage-specific expression of the *LacZ* reporter gene (βAGG20 transgenics, $\beta\text{galAB}+7$ transgenics but not βAGG10 transgenics possibly due to the absence of the *c/s* 3 enhancer combined with the presence of *c/s* 4). However, when both the 7kb and the 92base pairs were missing, the transgenic mice either did not have any expression (βAGG13 transgenics) or the expression was extremely low but still present in the forelimb digits and nasal area possibly due to the absence of the potential negative regulator *c/s* 4 ($\beta\text{galAB} + c/s\ 3$ transgenic). In the $\beta\text{galAB}^{+92} + c/s\ 3$ transgenics, there was also no '7kb sequence' but in comparison to the $\beta\text{galAB} + c/s\ 3$ transgenic, the expression was much higher most probably due to the presence of the 92 base pairs. However, It was still less than the βAGG20 transgenics but this was potentially due to the fact that there was no insulation of the short $\beta\text{galAB}^{+92} + c/s\ 3$ regulatory sequence from the genome as is indicated by the lack of chondrocytic specificity in one of the embryos. Unlike the results of Han and Lefebvre (2008), the human *c/s* 3 regulatory enhancer element is necessary but still insufficient in an endogenous setting in which it has not been multimerised and lies upstream of potential negative regulatory sequences. It would be extremely interesting to see what the inclusion of

cis 1 and 2 to the β galAB⁺⁹²+*cis* 3 construct (excluding *cis* 4) would do for the expression pattern in transgenic embryos. Since the ~20kb regulatory sequence in β AGG20 has the regulatory potential to drive reporter gene expression in the vertebral bodies and intervertebral discs (as is seen in the E16.5 transgenic embryos), it is possible that the inclusion of the 92 base pairs would be all that is required to boost the activity of the sequence further and enable earlier expression of the reporter gene in these elements.

1.3.5.2 Regulatory role of *cis* 6 (+5,000bp - +7,813bp)

As described in the Introduction section, transcription factors bind to *cis*-acting sequences to control gene expression. Those factors, such as sox9, which bind to consensus module sequences can bind to any promoter that contains these sequences. The binding of multiple factors may be essential for transcription initiation. Enhancers have sites for multiple *trans*-acting factors to bind. Thus genes with enhancers may require several complexes to be constructed for gene expression to be initiated. In order to find potential additional *cis*-regulatory elements that interact with the regulatory elements in the ~20kb fragment in β AGG20, I used the transcription factor sox9 for guidance since there was a lot of evidence for its involvement in the regulation of chondrocyte specific genes including collagen type II (Lefebvre et al., 1997).

In order to find additional *cis*-regulatory element(s) that can supplement the activity of the β AGG20 fragment and intensify/widen the areas of reporter gene expression, I decided to examine DNA fragments in the first intron of the gene. Although a sox9 interaction site that is essential for enhancer activity has already been identified within the mouse A1 sequence (corresponding to *cis* 3 in humans), it is possible that another site is also necessary for the full endogenous expression of the gene. This sox9 interaction site was identified first by a ChIP

assay in which direct binding of the sox trio (sox9, L-sox6 and L-sox6) to this A1 sequence was recognized. Secondly, an EMSA approach was used to delineate the binding sites of the proteins in A1 (Han and Lefebvre, 2008). Whereas L-sox5/sox6 bind 3 elements in the A1 sequence, sox9 binds a single element that is essential for enhancer activity. This element is a pair of sox-like sites separated by 4 nucleotides and oriented back to back. This configuration was proved to be critical. It seems that the consensus sox9 site in chondrocytes is a pair of sox-like sites positioned tail-to-tail and separated by 3 to 5 nucleotides. Interestingly, sox9 sites vary greatly amongst enhancers but not across species for any gene. These differences in sox9 sites amongst different enhancers may affect the efficiency of sox9 binding to DNA or its interaction with other proteins which is possibly what contributes to differences in expression levels and patterns between cartilage-specific genes. Another identified sox9 binding site in an additional enhancer sequence could therefore have a very different consensus binding site. The presence of an additional sox9 binding site within the aggrecan gene is supported by the fact that a) intron 1 of the mouse aggrecan gene has been shown to be responsive to sox9 (Sekiya et al., 2000). This group suggested that the possible interplay between sox9 and intron 1 could be what gives rise to chondrocyte specific regulation of mouse aggrecan b) By hybridising ChIP DNA to a microarray covering 80 genes, **2** hybridization peaks were detected for aggrecan (Oh et al.) and c) sox9 regulates the expression of collagen type II via a consensus site residing in the first intron of the gene (Lefebvre et al., 1997). This could be quite significant since it is possible that the regulatory mechanism by which matrix-specific genes are regulated has been conserved through evolution.

Using the internet-based program genomatrix, a tool for investigating possible transcription binding sites within a desired sequence, I found 3 potential putative sox 9 binding sites within the first intron of the gene located at +5.3kb, +6.8kb and +7.4kb relative to the TSS. An approximately 3kb sequence (+5 - +8kb) called *cis* 6 and containing all these potential binding

sites was cloned downstream of the ~20kb regulatory sequence in β AGG20 in order to test its potential regulatory activity. Three out of 17 positive transgenic embryos showed identical and symmetrical patterns of reporter gene expression (Figure 28). The expression of the transgene, however, was reduced and limited to the distal digits of the forelimbs as well as the long bones of the fore and hindlimbs where only minimal expression was noted (Table 20). It seems that *cis* 6 behaves as a silencer as opposed to an enhancer that can repress the expression of the transgene. Whether this is truly a silencer element that downregulates the expression of the aggrecan gene endogenously under certain conditions or whether the downregulation seen is just an artefact of fusing *cis* 6 directly downstream of the ~20kb regulatory sequence in β AGG20 is unknown.

1.3.5.3 *cis* 7 (+64,630bp to +65,005bp).

Cis 7, a 150bp highly conserved sequence located in intron 14 of the human aggrecan gene in which the homology down to fugu is relatively high, was cloned downstream of *cis* 3 and AB (including missing 92 base pairs). The resultant vector was called β AGGAB⁺⁹² + *cis* 3 + *cis* 7 but unfortunately, no positive transgenic embryos were yielded. This sequence can possibly serve as a *cis*-regulatory sequence or even a microRNA target site. Although microRNAs usually target the 3'UTRs of target mRNA sequences, they can have complementary target sites within the coding sequence. microRNA 140 (miR-140), the only microRNA that has been shown to have a cartilage-specific expression pattern, is expressed in a differentiation-dependent pattern in mesenchymal stem cells and articular chondrocytes. Also, its levels change in OA cartilage and in response to IL-1 β stimulation. The increased expression of miR-140 leads to the increased expression *SOX9* and *AGGRECAN* (Miyaki et al., 2009). Although, studies on microRNAs have shown that they regulate gene expression by repressing translation or directing sequence-specific degradation of complementary mRNA, new evidence

has shown that some miRNA may also function to directly induce gene expression (Place et al., 2008). If not miR-140, *cis* 7 could be a target site for another microRNA. Another possibility is the sequence within *cis* 7 that it could be involved in RNA splicing.

1.3.6 Adult β AGG20 transgenic line 12

In order to monitor the expression driven by the ~20kb fragment in β AGG20 –with the missing 92 base pairs in the minimal/basal promoter- throughout adulthood, I created a line. The tail tips of 5-week-old positive transgenic pups were used to identify positive expressors by staining them with X-gal. The blue staining pattern of the tail tips was then compared to BR1 antibody stained histological sections of 5-week-old wild type tail tips. The blue staining was mostly present in the nucleus pulposus as well as the resting and proliferating chondrocytes of the growth plate of the vertebral bodies. There was some weak staining in the deeper hypertrophic cells as well. The brown staining in the BR1 antibody stained wild type tail tip sections was also located in the nucleus pulposus as well as the resting and proliferating chondrocytes but was more predominant in the deeper hypertrophic cells. Another discrepancy in the staining patterns was due to the additional weak BR1 antibody staining in the annulus fibrosus and the cartilage end plate of the wild-type tail tips which was not found in the X-gal stained β AGG20 tail tips. These differences could have been due to a couple of things: Firstly, the X-gal staining method is a far less sensitive method of protein detection compared to antibody staining as it requires a certain threshold of protein to be present in order for it to be enzymatically active and catalyse the substrate. Secondly, there is a difference in turnover of aggrecan compared to β -galactosidase protein. The half-life of aggrecan is much longer than the half-life of the β -galactosidase protein (thought to be approximately 24-48 hours – see Results Chapter 3, Discussion section 3.3.1). This may be why the BR1 antibody staining in the hypertrophic chondrocytes of the wildtype tail tip vertebral growth plate is much stronger

than the X-gal staining within the hypertrophic chondrocytes of the β AGG20 tail tip vertebral growth plate.

Nevertheless, it is possible that this line does not express the reporter gene in the same way to what was previously seen in β AGG20 transgenic embryos due to the inevitable variation in the location of the inserted transgene within the chromatin. It was important to verify that in this particular line, the transgene wasn't inserted into a silenced region of the chromatin. To check this, a 4 month-old homozygous male was crossed with 2 F1 females. The females had plugs on two consecutive days and were sacrificed approximately 2 weeks later alongside the homozygous male. Along with the E13.5 and E14.5 embryos retrieved from the pregnant mothers, the heart, lungs, kidneys, brain, liver, spleen, spine, foot pad, thorax, hips, knees and tail of the homozygous male were also dissected and stained overnight with X-gal.

The embryos at various stages of development were all positive for cartilage-specific expression and were comparable to the 1st ' β AGG20' transgenic embryos acquired through microinjections (Figure 30 A-D). Regarding the adult male, the heart, lungs, kidneys, brain, liver, spleen, spine and thorax were all negative (data not shown). Although the results from the lungs, kidneys, and spleen were expected, the lack of X-gal staining in the heart, liver, brain, spine and thorax posed some questions (see Introduction section 3.3.2 for expression pattern of aggrecan). Aggrecan is known to be expressed in the brain postnatally as well as the hyaline cartilage of the vertebral end and growth plates, the fibrocartilage of the intervertebral discs and the hyaline costal cartilage in the thorax. Also, it was strange that there was no X-gal staining in the cartilaginous segments of the footpad and minimal staining in the hip and the knee. The few cells that were stained blue in the hip were deeper cells and the only cells stained blue in the knee were the hypertrophic cells of the tibial growth plate. None of the cells of the articular cartilage were positive. The inconsistency between the X-gal

staining and the BR1 antibody staining in the hip was too big to rule it down to a difference in turnover of the proteins. As before, the tail was the only examined anatomical structure to show a high level of blue staining.

There could be a number of reasons why the X-gal staining pattern in the different tissues and organs of the adult β AGG20 transgenics did not match the same staining profile seen embryonically. Firstly, although we verified that the transgene had not landed in a region of the chromatin that was silenced embryonically, it was not possible to know whether this region of chromatin was still active in adults due to the epigenetic modifications that can occur throughout different stages of development. It is possible that due to a difference in methylation pattern of histones that can occur post-development, the region of chromatin in which the β AGG20 transgene landed was silenced (Pedram et al., 2006), (Williams et al., 2008). The only way to verify whether this is the case or not is to produce multiple other β AGG20 transgenic lines and compare the adult expression of the reporter gene in different tissues across the different lines.

Another possibility is the involvement of the missing 92 bp in adult-specific expression of aggrecan. It is possible that the promoter is able to compensate for the lack of 92 bp sequence embryonically but not in adults due to a difference in transcriptional mechanisms that operate in embryonic vs adult cartilage. Han and Lefebvre (2008) stated that the fact that the A1 enhancer sequence drives gene expression with a pattern that is different from that of endogenous Col2a1 and Col2a1 reporters in that it is not expressed in skeletogenic mesenchymal cells, the perichondrium, and presumptive joint cells and is thus strictly specific for differentiated chondrocytes, the growth plate and articular cartilage through adulthood, can be useful to elucidate the specific transcriptional mechanisms operating in adult articular cartilage which remain unknown. They suggested that A1 constitutes a new tool to investigate

mechanisms that control gene expression differentially according to cartilage type, age, and physiological or pathological conditions (Han and Lefebvre, 2008). If the future generated β AGG20⁺⁹² transgenic lines are able to drive expression of the reporter gene with a similar pattern to endogenous aggrecan gene expression in adult cartilage, the ~20kb $\pm$ the missing 92 bp would serve as a new tool to elucidate the different transcriptional mechanisms that govern the expression of matrix genes such as aggrecan in differentiated embryonic chondrocytes vs differentiated adult chondrocytes. It is also possible that it is yet another *cis*-regulatory sequence not found within the ~20kb including the missing 92 bp that is important for the adult-specific regulation of aggrecan.

The fact that the *LacZ* gene expression in the cartilaginous elements of the tail remains relatively high in the β AGG20 transgenic adults from line 12 could possibly mean that the transcriptional mechanisms that govern the expression of aggrecan in the tails of mice could be different to those that govern the expression of the gene in other tissues.

Chapter 2: Conditional/inducible cartilage-specific model

2.1 Background

As mentioned earlier, aggrecan is a chondrocyte-specific protein and its expression is detected in growth plate and articular chondrocytes of long bones and other cartilage tissues in the body (Murphy et al., 1999). Unlike collagen type II, aggrecan is expressed continuously in cartilage (Chambers et al., 2002). To generate a chondrocyte-specific transgenic mouse model, the enzyme Cre recombinase can be fused to a mutated ligand-binding domain of the human estrogen receptor (ER) and driven by the aggrecan promoter. This fusion protein will be expressed in chondrocytes and translocated into the nuclei in response to the estrogen antagonist tamoxifen or 4-hydroxy tamoxifen (Feil et al., 1996).

Characterizing the spatial and temporal pattern of Cre expression in a transgenic mouse line is generally achieved using a Cre reporter transgenic mouse that contains a marker gene (e.g. *LacZ*, *GFP*, *CFP* or *YFP*) within its genome that is only expressed following Cre-mediated recombination. For these purposes, we used the global double-fluorescent Cre reporter mouse line. Within this Cre reporter line, a marker is expressed not only in cells that have undergone Cre-mediated recombination but also in those that have not. The markers are membrane-targeted tdTomato ('mT') and membrane-targeted EGFP ('mG'). Prior to Cre excision, 'mT' is synthesised in cells while following excision, there is a switch to the production of 'mG' thereby allowing live visualisation and differentiation between recombined and non-recombined cells. The 'mT/mG' construct combines the use of a strong ubiquitous promoter pCA (Niwa et al., 1991), with the *ROSA26* targeting locus (Soriano, 1999). Figure 34 illustrates the mechanism by which cells switch from expressing mT before Cre-mediated excision to mG after Cre-mediated excision.

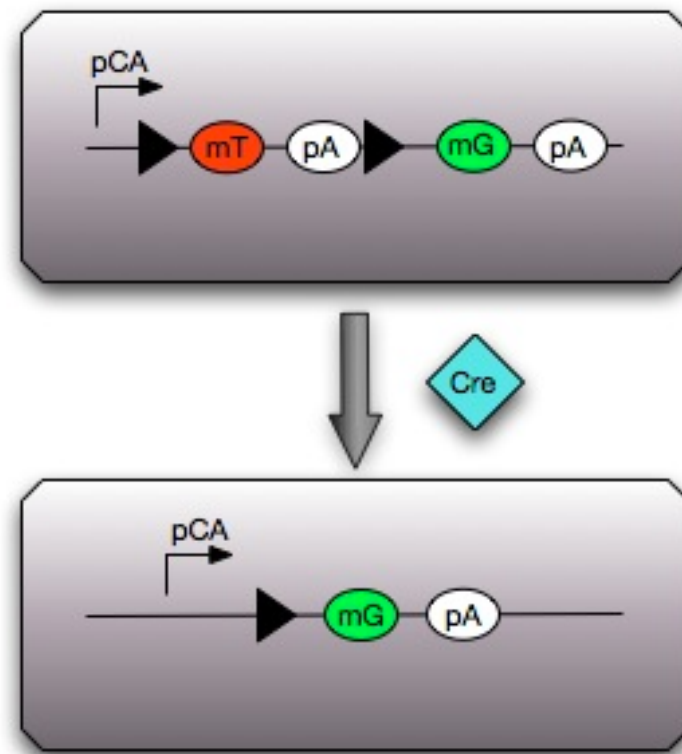


Figure 34: Schematic diagram of the mT/mG construct before and after Cre-mediated recombination. mT/mG consists of a chicken β -actin core promoter with a CMV enhancer (pCA) that drives a loxP-flanked coding sequence of membrane-targeted tandem dimer Tomato (mT), which results in tdTomato expression with membrane localization. After Cre-mediated intrachromosomal recombination, the mT sequence is excised allowing the pCA promoter to drive the expression of membrane-targeted enhanced green fluorescent protein (mG). Arrows denote the direction of transcription. Triangles represent loxP target sites for Cre-mediated recombination. PA denotes polyadenylation sequences. Adapted and modified from (Muzumdar et al., 2007).

2.2 Results

In order to generate a conditional/inducible cartilage-specific transgenic mouse in which particular genes can be knocked out/over-expressed preceeding/following injury (that can lead to osteoarthritis), it was important to use a promoter that would drive cartilage-specific expression throughout adulthood, as osteoarthritis is a late-onset disease. For this reason, I opted against the use of the ~20kb fragment in the β AGG20 transgenic line 12 which proved insufficient in driving adequate reporter gene expression in adult knee articular cartilage.

A more appropriate promoter would be the regulatory sequence created by Han and Lefebvre (Han and Lefebvre, 2008). The construct contained 4 copies of the mouse 'aggrecan enhancer' which they named A1 (corresponding to *cis* 3 in my work) fused downstream of the *Col2a1* minimal promoter (-309 - +308 base pairs relative to the mouse *Col2a1* transcriptional start site), i.e, the proximal promoter, first exon (with the translation initiation site mutated into CTG), as well as 70p of the first intron (not including enhancer sequences). They also included a splice donor and splice acceptor site immediately upstream and downstream of the 4xA1 sequence respectively. For these purposes, and with the help of Hiro Nakamura at the KIR, this *ColII*4xA1 sequence was cloned upstream of the CreER^{T2} fusion gene, IRES sequence, the firefly luciferase reporter gene and polyA DNA sequences (Figure 35A). The firefly luciferase gene serves as a reporter gene to locate the expression of CreER^{T2} by *in vivo* bioluminescence imaging while the IRES (internal ribosomal entry site) sequence ensures the transcription of bi-cistronic mRNA in which two proteins are translated from one mRNA sequence. This construct was named 'ColII4xA1Cre' (Methods section 5 construct 17).

To generate a transgenic mouse line, this sequence was microinjected into fertilised oocytes as described earlier and positives identified via real-time PCR. A positive expressor was

selected by detecting transgene expression through whole animal imaging. The Kodak imaging system was used to detect any bioluminescence (Methods section 7.2.2.2). Upon injection of transgenic mice synthesising luciferase with the substrate luciferin, a reaction occurs in which light that is emitted from luciferase-producing live cells is captured as light photons and photographed by this machine. This mouse was then bred into a homozygous line named line 39 after the founder. The light emission from this transgenic line was detected as a black pseudo-colour and confined to cartilaginous regions as expected. Figure 35B displays an example of a 'ColII4xA1Cre' mouse in which light is detected in the footpads, paws, knees as well as the segmented tail. Light emission is not detected in all other known cartilaginous regions due to the effect that mouse fur has on blocking this. Black fur blocks weak signal from internal organs or unexposed cartilages and bones

Although the spatial and temporal characterisation of the pattern of CreER^{T2} expression was achievable via the use of the Kodak imaging system, examining functionality of the Cre recombinase required a different system. To do this, we used the global double-fluorescent Cre reporter mouse (Muzumdar et al., 2007).

Upon crossing a homozygous female 'mT/mG' global double-fluorescent reporter mouse with a heterozygous male mouse from line 39 (ColII4xA1Cre), the intraperitoneal injection of the pregnant mothers or F1 progeny with tamoxifen would theoretically lead to the Cre-producing cells (chondrocytes) in half the F1 progeny to be labelled with bright green membrane-bound fluorescence. All other cells should be labelled with bright red membrane-bound fluorescence. The localization of fluorescent proteins to the cell membrane enables cell morphology to be outlined with fine resolution.

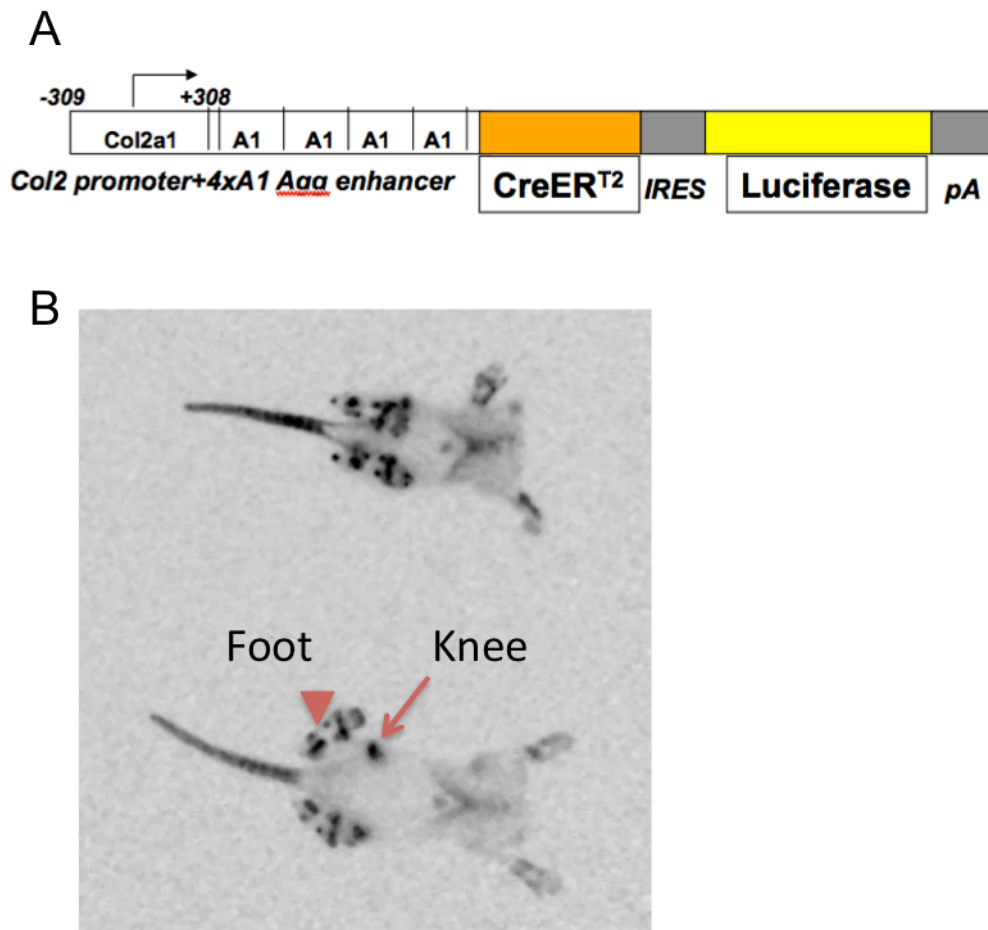


Figure 35: Bioluminescence imaging of luciferase activity in an adult transgenic mouse from the Col14xA1Cre transgenic line. A) Diagram of the Col14xA1Cre transgene. B) Dorsal view of an image of two representative Col14xA1Cre adult transgenic mouse showing luminescence/luciferase activity (black regions) in cartilaginous areas such as the knees, the tail, the foot pads and the paws.

To assess whether what is expected is in fact true, the pregnant mother in which half the F1 progeny were double transgenic 'mT/mG/Col14xA1Cre' while the other half were 'mT/mG' single transgenics was injected with 4 doses of 5mg tamoxifen (20mg in total) on the final 4 days of gestation (E15.5-E19.5). The hips and knees of the F1 progeny were then snap frozen 1 or 2 weeks after the final tamoxifen injections (at postnatal day 7 for one litter and at postnatal day 14 for another litter). Figure 36 demonstrates the hips of sacrificed pups at postnatal day 7 (P7 - 1 week after the final dose of tamoxifen was injected) while Figure 37 demonstrates the hip and knees of sacrificed pups at postnatal day 14 (P14 - 2 weeks after the final dose of tamoxifen was injected). In both cases, there has been a conversion from red-membrane bound fluorescence to green that is specific to the cartilage of the

'mT/mG'/Col14xA1Cre' double transgenics as demonstrated by the negative controls in which none of the cells underwent conversion (Figure 36 D1-3 and Figure 37 E1-3 and F1-3). Figure 36 D1-3 and Figure 37 E1-3 show hip cartilage 1 week after the final tamoxifen injection and knee cartilage 2 weeks after the final tamoxifen injection respectively of mTmG single transgenics. There was no conversion in any of the chondrocytes. Figure 37 F1-3 shows muscle from one of the P14 mT/mG'/Col14xA1Cre' double transgenics in which none of the cells have undergone conversion. The labelling in the hips from the 'mT/mG'/Col14xA1Cre' double transgenics at the two timepoints was mosaic but also mutually exclusive (Figure 36 A1-3, B1-3 and C1-3 for P7 double transgenics and Figure 37 A1-3 and B1-3 for P14 double transgenics). The mG labelling was present throughout the different zones of the hip articular cartilage ranging from the terminal cells on the surface to the deeper hypertrophic cells. The deeper hypertrophic chondrocytes were identified by their size and location adjacent to the autofluorescing bone marrow. At a higher magnification, it was possible to see that the fluorescence was targeted to the membranes of the cells. The labelling in the knees from the P14 'mT/mG'/Col14xA1Cre' double transgenics was also mosaic and mutually exclusive. Like the hips, mG labelling was present in the hyaline cartilage of the articular cartilage throughout the terminal chondrocytes of the articular cartilage as well as the fibrocartilage of the meniscus. However the labelling did not seem to extend below the tidemark to the hypertrophic cells of the articular cartilage. In both the knees and the hips, the bone marrow showed autofluorescence as the fluorescence overlapped from both the green and red channels and was not targeted to the membranes of any cells. There are no knee images for the 'mT/mG'/Col14xA1Cre' double transgenics at the '1 week' timepoint due to the unfortunate fragmentation of the sections.

Figure 36 (next page): Ubiquitous mT and mG labelling of fixed hip sections in ‘mT/mG’/Col14xAlCre’ double transgenic mice (A1-3, B1-3 and C1-3) and an ‘mT/mG’ single transgenic (D1-3) post Cre-mediated recombination (1 week after final injection of tamoxifen). A1-D1 are bright field confocal images, A2-D2 are mT labelled confocal images while A3-D3 are mG labelled confocal images. The 3 images on each row are the same image captured using different filters. All images are at 10x magnification except for C1-3, which are at 20x magnification. AC (articular cartilage), BM (bone marrow), HC (hypertrophic chondrocytes), TC (terminal chondrocytes).

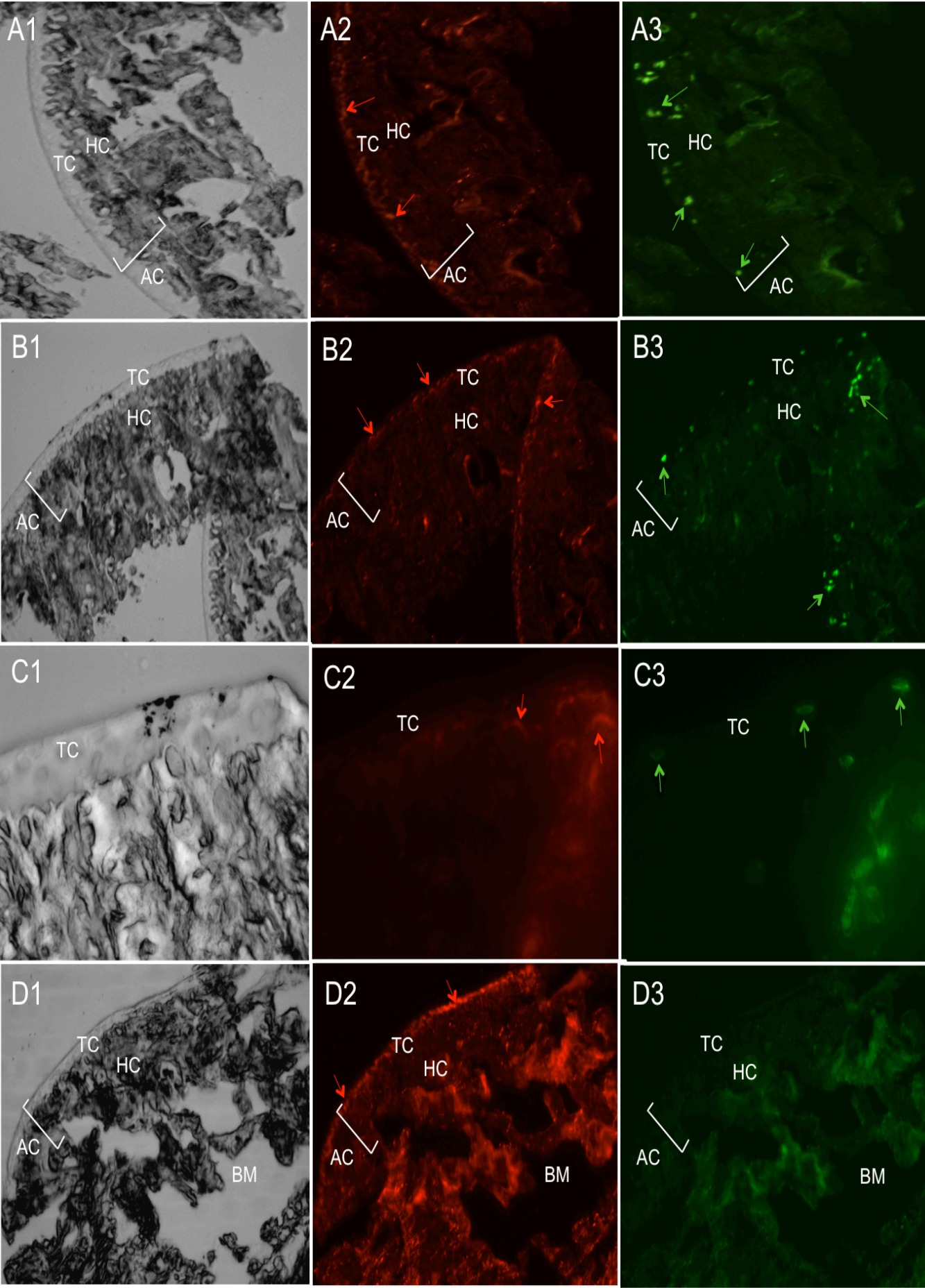
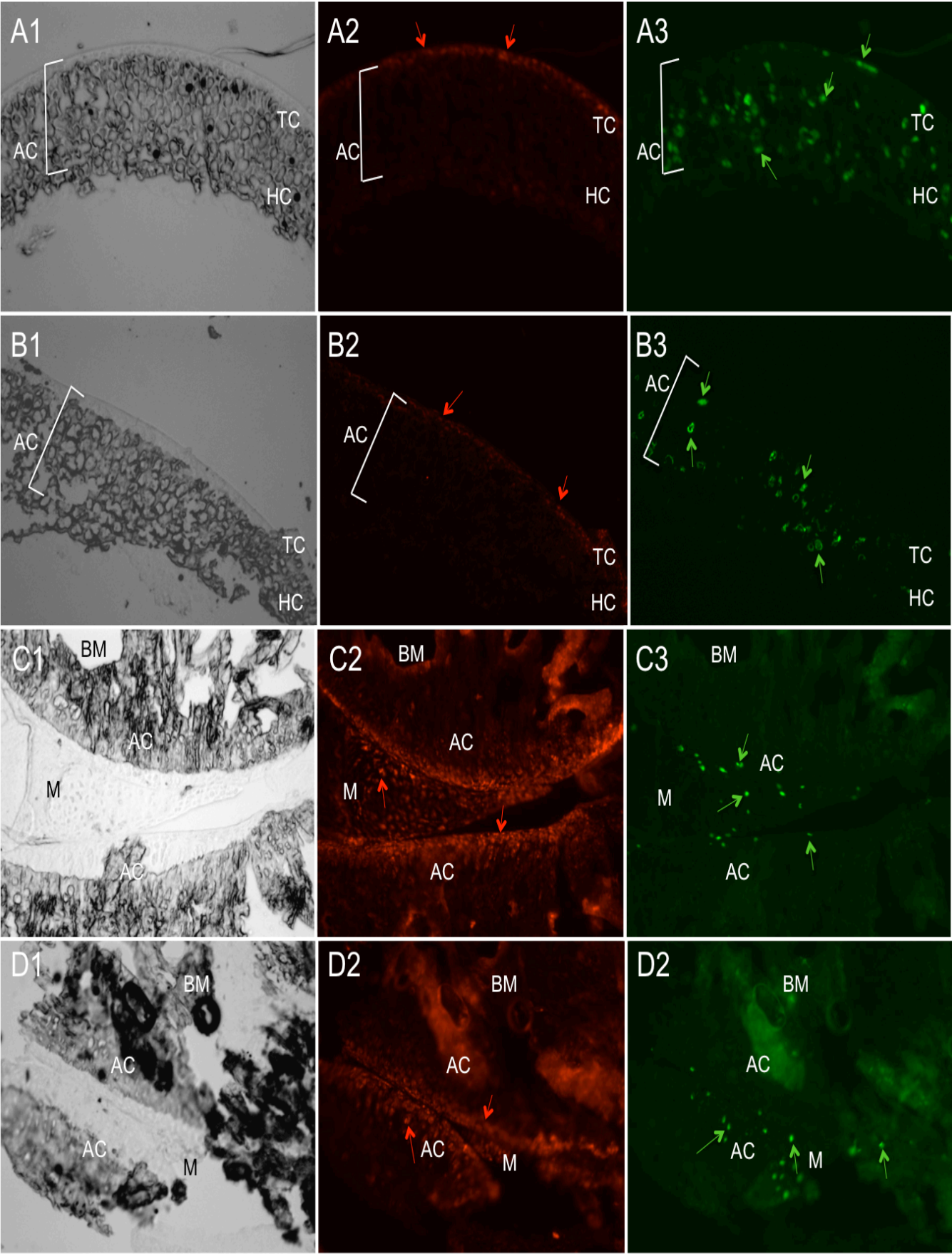
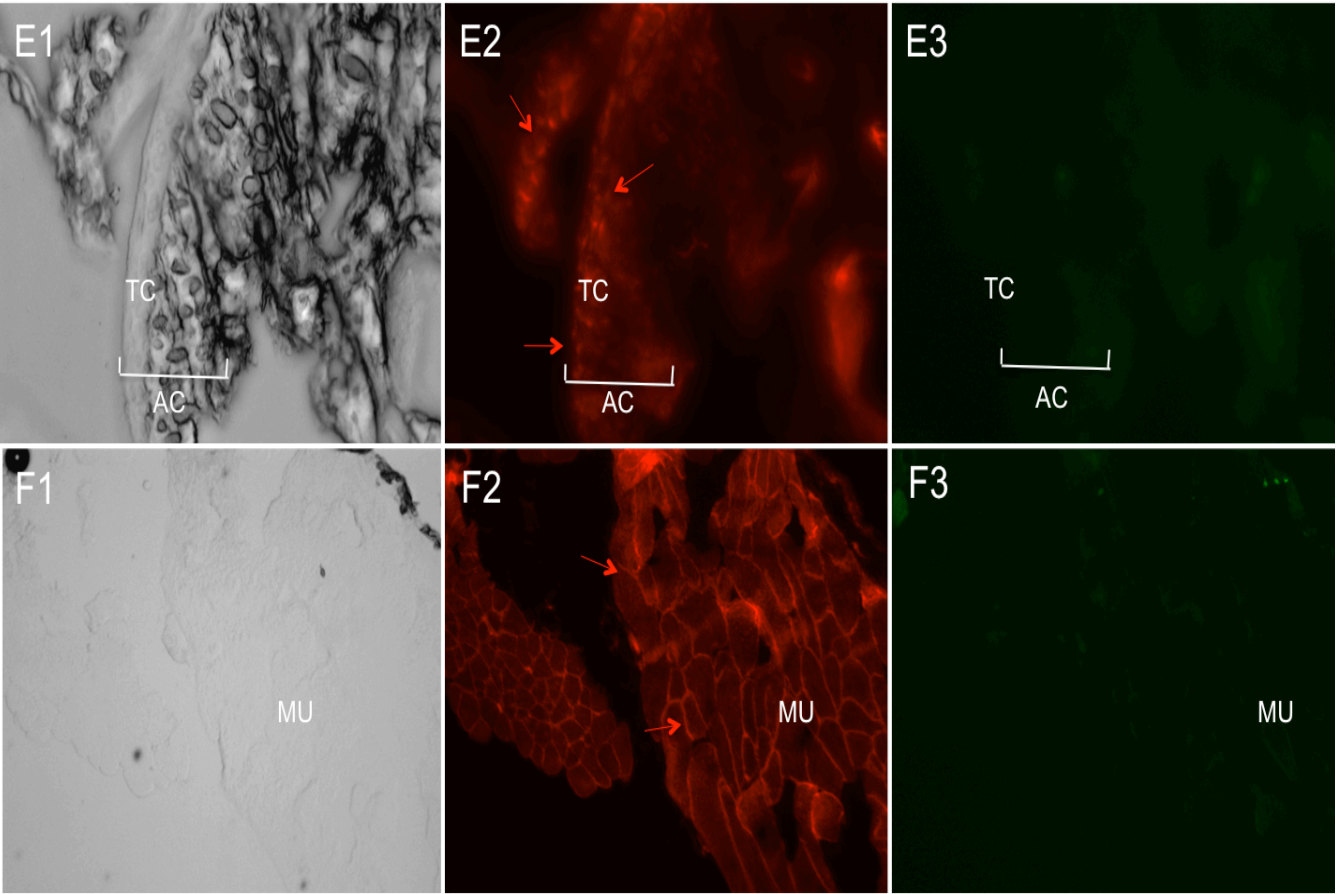


Figure 37 (next 2 pages): Ubiquitous mT and mG labelling of fixed hip (A1-3, B1-3), knee (C1-3, D1-3) and muscle (F1-3) sections in 'mT/mG'/'Col114xA1Cre' double transgenic mice and hip sections of an 'mT/mG' single transgenic (E1-3) post Cre-mediated recombination (2 week after final injection of tamoxifen). A1-E1 are bright field confocal images, A2-E2 are mT labelled confocal images while A3-E3 are mG labelled confocal images. The 3 images on each row are the same image captured using different filters. All images are at 10x magnification except for E1-3, which are at 20x magnification. AC (articular cartilage), HC (hypertrophic chondrocytes), M (meniscus), MU (muscle), TC (terminal chondrocytes).





2.3 Discussion

Aggrecan is a major ECM protein of both growth plate and articular cartilage in long bones and other cartilage tissues in the body. In articular cartilage, and particularly in adults, aggrecan is expressed more robustly than collagen type II. This was shown through a previous study of mouse knee joints in which endogenous Col2a1 mRNA was detected in the growth plate by *in situ* hybridization but not in articular cartilage or in the meniscus of adult mice ranging from 7 to 10 months of age. In addition to this, aggrecan mRNA was detected in the articular, growth plate and meniscus cartilage of these adult mice, indicating that the aggrecan gene maintains its expression in articular cartilage of adult mice at a time when Col2a1 RNA is no longer evident (Chambers et al., 2002). So although collagen type II is expressed specifically in cartilage, it is switched off in adult mice.

For this reason, the inducible cartilage-specific Cre recombinase transgenic mice lines that have already been generated using the Col2a1 promoter have certain limitations when analyzing OA models. Although inducible Cre recombinase activity was detected in the cartilage of postnatal animals in these lines, they each had their own limitations. In one of these, in which expression of Cre recombinase is under control of both the type II collagen promoter (3kb promoter and 3kb first intron of Col2a1 (Zhou et al., 1995)) to allow cartilage-specific expression and a doxycycline response element to permit temporal control of expression, inducible Cre recombinase activity was observed in articular and growth plate cartilage of 1-month-old mice but no longer at later timepoints (Grover and Roughley, 2006). In a second transgenic mouse, in which CreER^{T1} expression was driven by a minimal promoter and an enhancer element of Col2a1 (Zhou et al., 1995), tamoxifen-induced Cre recombinase activity was detected in the articular cartilage of mice 3 weeks post-natally, but this activity gradually decreased with age and was no longer detected in 3-month-old animals (Nakamura

et al., 2006). In a third transgenic line (Chen et al., 2007; Zhu et al., 2008a), in which promoter and enhancer elements of Col2a1 were directing expression of CreER^{T2}, efficient tamoxifen-induced Cre recombination activity was shown in articular chondrocytes of adult 6-month-old animals, but the Cre recombinase activity was much less efficient in growth plate cartilage (Zhu et al., 2008b). It has been suggested that the discrepancy between the expression of Col2a1 mRNA in adult mice and that of the Col2a1-CreER^{T2} transgene in the third transgenic mice line might be a consequence of the transgene integration site. It would therefore be challenging to use these models for exploring osteoarthritis pathogenesis since it is a disease only affecting adult/aging mice.

The promoter of the aggrecan gene was selected here to drive cartilage specific expression of the CreER^{T2} fusion protein due to aggrecan being the most abundant proteoglycan in the ECM of cartilage as well as its continuous expression throughout adulthood. Another group generated an aggrecan-CreER^{T2} mice, in which CreER^{T2} expression very closely paralleled expression of the endogenous aggrecan gene. In these mice, a Cre-ER^{T2} cassette was introduced by homologous recombination into the 3'UTR of the aggrecan gene. The group used a Cre reporter line to localize the expression of the CreER^{T2} transgene and validate the activity of the recombinase. They found the expression of the CreER^{T2} transgene to be highly specific and efficient in postnatal mice at 6 and 12 months of age in articular chondrocytes, in growth plate chondrocytes, and in fibrocartilage cells of the meniscus and of the nucleus pulposus and annulus fibrosis of intervertebral disks. The excision was also very efficient at different times during embryonic development and during the earlier postnatal growth period (Henry et al., 2009).

For our purposes, I used an aggrecan promoter approach as opposed to the knockin mechanism used by Henry et al. Although the work I carried out to identify the regulatory

region(s) responsible for endogenous aggrecan regulation in chondrocytes led to the discovery of a reproducible cartilage-specific ~20kb sequence that was active in developing embryos, this sequence was not as robust in adults. As discussed in the previous results chapter, β -galactosidase expression driven by the ~20kb regulatory sequence in line 12 was only very weakly detected in the articular cartilage of the hips and the growth plate cartilage of the tibia as well as the fibrocartilage of the intervertebral discs of the tail. For the former 2 tissues, the level of expression was random, inconsistent, extremely low and only present within a few chondrocytes. There seemed to be no expression in the articular cartilage of the knees either. For this reason, we opted to use the established A1 mouse sequence (Han and Lefebvre, 2008), which was equivalent to the *α1* 3 human sequence in my work. Since a single copy of the A1 sequence only weakly activated the *Agc1* -596 - +90bp minimal promoter (pAgc1) and *Col2a1* -86 - +6bp minimal promoter (pCol2), we obtained the multimerised sequence in which 4xA1 copies were cloned downstream of a lengthened *Col2a1* minimal promoter (-309 - +308bp). This sequence was shown to drive gene expression in transgenic mice with the characteristic spatial and temporal pattern of aggrecan. In particular, reporter gene expression was revealed to be tissue-specific with expression found in all craniofacial, axial and appendicular cartilage elements but lacking in noncartilaginous tissues, prechondrocytes, the perichondrium and presumptive joints. Embryonic sections showed expression in differentiated articular and growth plate chondrocytes as well as in notochord-derived nucleus pulposus cells. Importantly, when tested postnatally, the sequence was able to drive high reporter gene expression in all growth plate and articular cartilage structures of actively growing 2-week old pups. This remained true for the same structures in 6-month-old adult mice.

This *coll4xA1* sequence was cloned upstream of the CreER^{T2} fusion protein, an IRES sequence, the firefly luciferase reporter gene as well as a mouse polyA signal. The IRES sequence was used to allow for the production of two proteins, CreER^{T2} and firefly luciferase,

out of one mRNA sequence. Firefly luciferase catalyzes the bioluminescent oxidation of the substrate luciferin in the presence of ATP, magnesium and oxygen. The firefly luciferase used was that of the North American firefly *Photinus pyralis*. The advantages of using the firefly luciferase as a reporter gene are the high sensitivity in detection of expression of the gene, the absence of luciferase activity inside most cell types, the wide dynamic range, rapidity and low costs. Also, the protein does not require any posttranslational modification to be enzymatically active and is not toxic in high concentrations *in vivo* (de Wet et al., 1987). For these purposes, the use of this reporter gene allowed me to carry out genotyping and pinpoint the location of *CreER^{T2}* expression at different ages and under different conditions e.g injury in the Col14xA1Cre mice (Chapter 3 'Results').

The bioluminescent images of the transgenic mice from the Col14xA1Cre line (line 39) showed enzyme activity within some cartilaginous elements such as the cartilage of the tail, foot pads, paws and knees (Figure 35B). The remaining cartilaginous elements of the axial and appendicular skeleton did not reveal enzyme activity due to the presence of fur around these regions. As described in the methods section, the fur was only shaved in the knee/leg area as it was an important region of interest due to the future applications of this conditional/inducible model for investigating the roles of genes within wildtype and OA-induced knees (Chapter 3 'Results').

To test for the functionality, specificity and potential leakiness of the *CreER^{T2}* fusion protein and establish a protocol for activation of cartilage-specific Cre in the Col14xA1Cre mouse line, the global double fluorescent Cre reporter mouse was used (Muzumdar et al., 2007). The advantages of using this mouse model are that it allows for direct visualization of both recombined and nonrecombined cells at a single-cell resolution, offering an internal control for phenotypic analysis of Cre-induced mosaic mutants. All this without the need for assaying for

enzymatic activity or antibody staining. However, one of the potential drawbacks of the double-fluorescent system is the possible bleedthrough between the red and green fluorescent channels. A cell that is labeled using the red channel filter may also be labeled using the green channel filter. Also, some tissues autofluoresce and it is important to take this into consideration. Nevertheless, if the mT and mG labeled cells do not overlap, it is safe to say that there is minimal bleedthrough between the red and green fluorescent channels. Also, because the fluorescent proteins are membrane-targeted, it is possible to rule out autofluorescence that is not due to *mT/mG* expression by visualizing the cells at higher magnifications.

In theory, Cre-mediated excision should occur within 24 hours after administration of Tamoxifen to the mouse. However, it is important to allow for a certain amount of time between Cre induction and analysis. This is to allow for the turnover of mT in cells that have undergone Cre-mediated excision. Examining tissues in a relatively short time interval after Cre induction would reveal double labeled cells (these cells appear yellow) due to the rapid onset of mG labeling and the perdurance of mT mRNA and protein. (Muzumdar et al., 2007) generated mosaic mice with the *actin-CreER* line (Muzumdar et al., 2007). The injection of 2 mg of TM at P7 followed by dissection at P21 revealed a mosaic and mutually exclusive pattern of mT and mG labeling in multiple tissues. They concluded that dissection 2 weeks after recombination induction with *actin-CreER* was sufficient to observe minimal numbers of double-labeled cells in most tissues examined. Because the rate of mT loss varies from tissue to tissue based on cell type-specific mRNA and protein turnover properties, I chose to analyse hips and knees 1 and 2 weeks after the final administration of Tamoxifen (20mg in total).

Figure 36 & Figure 37 show ubiquitous mT and mG labelling of fixed hip and knee sections in 'mT/mG/Col14x1Cre' double transgenic mice and 'mT/mG' single transgenic (control) post

cre-mediated recombination 1 week (Figure 36) and 2 weeks (Figure 37) after final injection of tamoxifen. For both the hip and the knee articular cartilage at both timepoints, there has been a conversion from red-membrane bound fluorescence to green that is specific to the chondrocytes of the 'mT/mG'/Col114xA1Cre' double transgenics as demonstrated by the negative controls in which none of the cells underwent conversion. The hip sections of the 1 and 2 week timepoints revealed conversion in the terminal and hypertrophic chondrocytes whereas the knee section of the 2 week timepoint (no one week timepoint due to deterioration of sections) revealed conversion in the superficial and deeper layers of the menisci as well as the superficial chondrocytes of the uncalcified articular cartilage above the tidemark. This was to be expected as the same was noted in the $Agc1^{tm(IRES-creERT2)/+}$ mice produced by Henry et al. By *in situ* hybridization, they found that the mRNA expression pattern of aggrecan was limited to the uncalcified articular cartilage above the tidemark (Henry et al., 2009). The two negative controls used were a) the cartilaginous tissues of the mTmG single transgenics whereby the lack of conversion from mT to mG fluorescence indicated that this conversion occurs indeed due to the activity of the Cre recombinase in the cells as opposed to leakiness in the system and b) non cartilaginous tissue (muscle) of the 'mT/mG'/Col114xA1Cre' double transgenics whereby the lack of conversion from mT to mG fluorescence indicated that the Cre recombinase was produced specifically in chondrocytes. The labelling in the hips and the knees from the 'mT/mG'/Col114xA1Cre' double transgenics at the two timepoints was membrane-targeted (indicating real fluorescence as opposed to autofluorescence), mosaic (individual chondrocytes showed either green or red fluorescence) and mutually exclusive (chondrocytes were not doubly labeled which would have appeared yellow). This means that the chondrocytes examined required as little as 1 week to switch on mG expression and labeling and turnover the mT mRNA and protein. The mosaic pattern in both the knees and hips at both timepoints indicates that the Col114xA1 promoter is active at different times across individual chondrocytes, which possibly reflects the way the endogenous aggrecan promoter

regulates aggrecan gene expression. Aggrecan may not be expressed continually in every cell but rather, as and when it is needed to maintain the ECM.

To take this model further, it is important to perform these experiments on mice at 1, 3 and 6 months of age. Although the promoter was shown to be active at 6 months of age (Han and Lefebvre, 2008), it is necessary to verify this in line 39. Also, in addition to articular cartilage, it would be useful to test for conversion within the hyaline cartilage of the growth plate. For these purposes, *Rosa26* β -galactosidase reporter mice will be used in order to ensure high quality knee sections. The visualisation of the membrane targetted mT and mG polypeptides in the cartilage of knees requires them to be frozen and eventually cut without decalcification. This can lead to the fragmentation of sections. However, the visualisation of β -galactosidase in knees does not require them to be frozen and they can be decalcified and paraffin embedded. Assuming the expression and activity of CreER^{T2} in line 39 remains the same at older ages and in other cartilaginous tissues (growth plate and intervertebral discs), this model should provide a very valuable tool to analyze the role of specific proteins in the maintenance of articular cartilage, growth plate and intervertebral disks at different timepoints postnatally in intact animals. These proteins include regulatory molecules such as key transcription factors, signaling molecules, and ECM proteins. The characterization of mice harboring the Col14xA1Cre transgene in combination with conditional alleles for these proposed proteins will give rise to new rational models for human degenerative cartilage diseases and will enable the dissection of the *in vivo* mechanisms that govern the renewal, stability, and integrity of the complex, multicomponent ECM of different cartilages.

Chapter 3: Induction of osteoarthritis in the knees of line 39 ('Col14xA1Cre') and line 12 (' β AGG20')

3.1 Background

In a bid to define the biological mechanisms through which mechanical stress leads to the development of OA, experimental mouse models have been developed whereby knee joint instability can be induced through different methods. One of these methods is the destabilization of the medial meniscus (DMM) through surgery (Glasson et al., 2007) while another is the axial loading of the mouse tibias (De Souza et al., 2005).

The DMM model in mice was developed due to the great ease and reproducibility with which it could be performed in guinea pigs. Another advantage of this model is the relatively slow speed with which OA develops in operated mice knees (closer resemblance to the slow development of OA in the older human population). It has been reported that, following surgery, the mouse knees develop mild to moderate OA with lesions primarily located on the anterior-central weight bearing area of the medial femoral condyle as well as the medial tibial plateau that increase in severity over time. Normally, there is no posterior erosion of the tibial plateaus and chondrogenesis is negligible (Glasson et al., 2007). Figure 38 illustrates a diagram of the right knee of a mouse. The medial meniscotibial ligament (MMTL), which anchors the medial meniscus (MM) to the tibial plateau, is transected in order to generate the destabilisation of the medial meniscus (DMM).

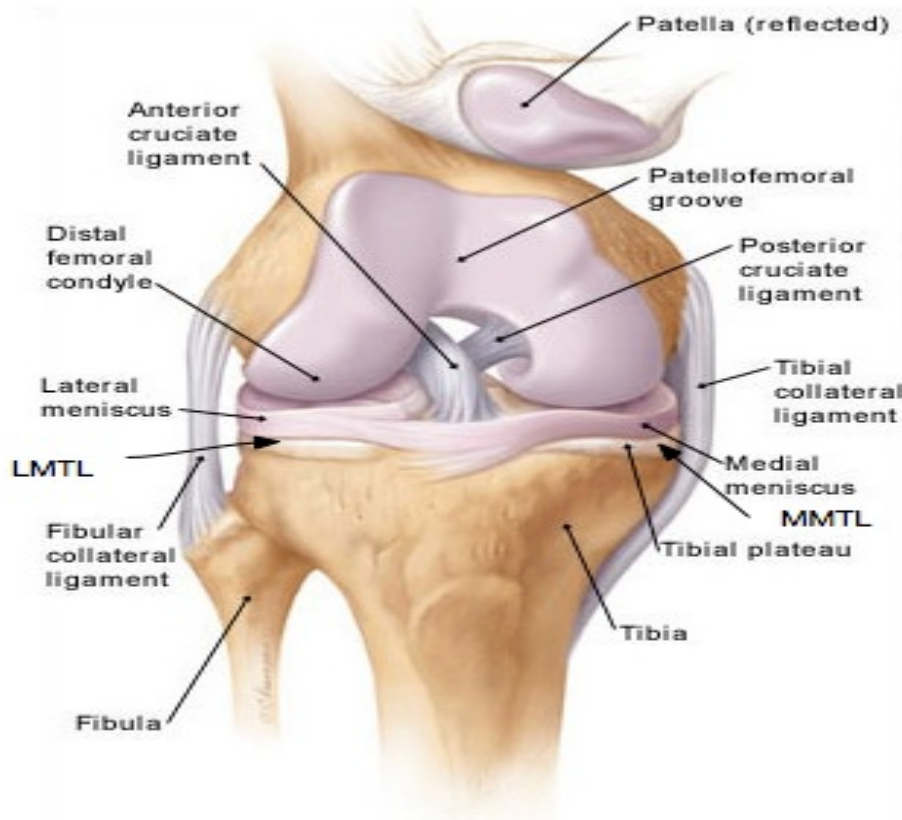


Figure 38: Basic anatomy of the knee. The main ligaments of the knee are: The ACL (Anterior Cruciate Ligament). It runs from the anterior tibia to the posterior femur and it keeps your knee from anteriorly translating: The PCL (Posterior Cruciate Ligament) runs posterior tibia to anterior femur and it controls the posterior translation; The LCL (Lateral Collateral Ligament); the MCL (Medial Collateral Ligament) run on the lateral and medial side respectively of the tibia and femur; the MMTL (medial meniscotibial ligament) which anchors the medial meniscus (MM) to the tibial plateau and the LMTL (lateral meniscotibial ligament). (<http://www.ncssm.edu/library/dirt/Valerie%20Todd/Basic%20Anatomy.html>).

Another model, the non-invasive axial loading of mouse tibias, is one which enables the study of how loading modulates (re)modelling of the knee joint. The mechanisms that initiate and coordinate these events can also be measured. The model is achieved by applying axial compressive loads to the mouse tibia through natural sites of articulation. The modifications that occur in cancellous and cortical bone can then be quantified, in response to controlled exogenous loading under conditions of continual use or disuse.

It has been shown that short periods of externally applied dynamic loading of the mouse tibia for 2 weeks (on alternate days) is an effective means of promoting adaptive modification in the cortical bone architecture and variation in trabecular structure in the proximal tibia. It induces periosteal formation at all diaphyseal sites within the central third of the tibia. In contrast to this periosteal response to loading, the load-induced endosteal formation is less widespread, spreading only distal to the mid-shaft. Consequently, this model of loading promotes radial expansion along a segment of the tibial mid-shaft that is coupled to marked increases in cortical thickness mostly at, and distal to the mid-shaft. The changes in endosteal, periosteal and trabecular bone are consistent with a regionalised osteogenic response (De Souza et al., 2005).

As opposed to characterising the biological components that can lead to the development of OA, the application of the above-described models to the transgenic lines 'βAGG20' and 'Col14xA1Cre' would enable us to further understand/define the response that either promoter sequence possibly has to injury.

3.2 Results

3.2.1 DMM of line 12 (β AGG20 transgenics)

DMM was performed on transgenic mice from line 12 based on the described method in Results sections 8.1 & 8.1.1. Knees were then analysed based on the described method in Results sections 8.1.1. At 4 weeks post-operatively, the whole mount X-gal staining of the right operated knees in all 6 of the sacrificed animals showed localised blue staining on the medial side of the knee. However, the left contralateral control knees of these mice showed no staining (Figure 39A&B). At 8 weeks post-operatively, the whole mount X-gal staining of both right operated knees and control left unoperated knees showed no staining (Figure 39C&D).

For histological analysis of reporter gene expression, all knees (operated and unoperated) from both the 4 and 8-week groups were analysed microscopically and operated knees with the highest score (indicating a more biologically aggressive disease) based on The OARSI Osteoarthritis Cartilage Histopathology scoring system were photographed alongside the contralateral unoperated knees and are shown below in Figure 40 A-E. Figure 40 A-E shows representative images corresponding to the sections with the maximum score (highest grade/stage) for the operated right knees of the 4 and 8-week groups as well as their contralateral control left knees. The histological analysis of the 4-week post-operative group revealed low levels of OA in the left unoperated control knees (grade 1, stage 1) (see Figure 40 A-C). At 8 weeks, the scores were similar for the knees that had not undergone surgery (also grade 1, stage 1) (see Figure 40 A-C). Erosion and free cells were not observed in these control knees.

The right knees, which had undergone the DMM surgery, had significantly more OA than the left control knee and increased scores were observed. The histological analysis of the 4-week

post-operative group revealed mild-moderate levels of OA (grade 3.5, stage 3) (see Figure 40 A-C). At 8 weeks, there was a progression in the disease from mild-moderate to moderate (grade 4.5, stage 3) (see Figure 40 A-C). Lesions were most apparent on the medial side of the joint and mainly on the tibia plateau. Safranin O staining showed severe loss of proteoglycans, fibrillation and erosion of the tibial plateau cartilage. Chondrogenesis, patellar dislocation, free cells in the synovial cavity and ectopic bone formation are not usually observed in this model.

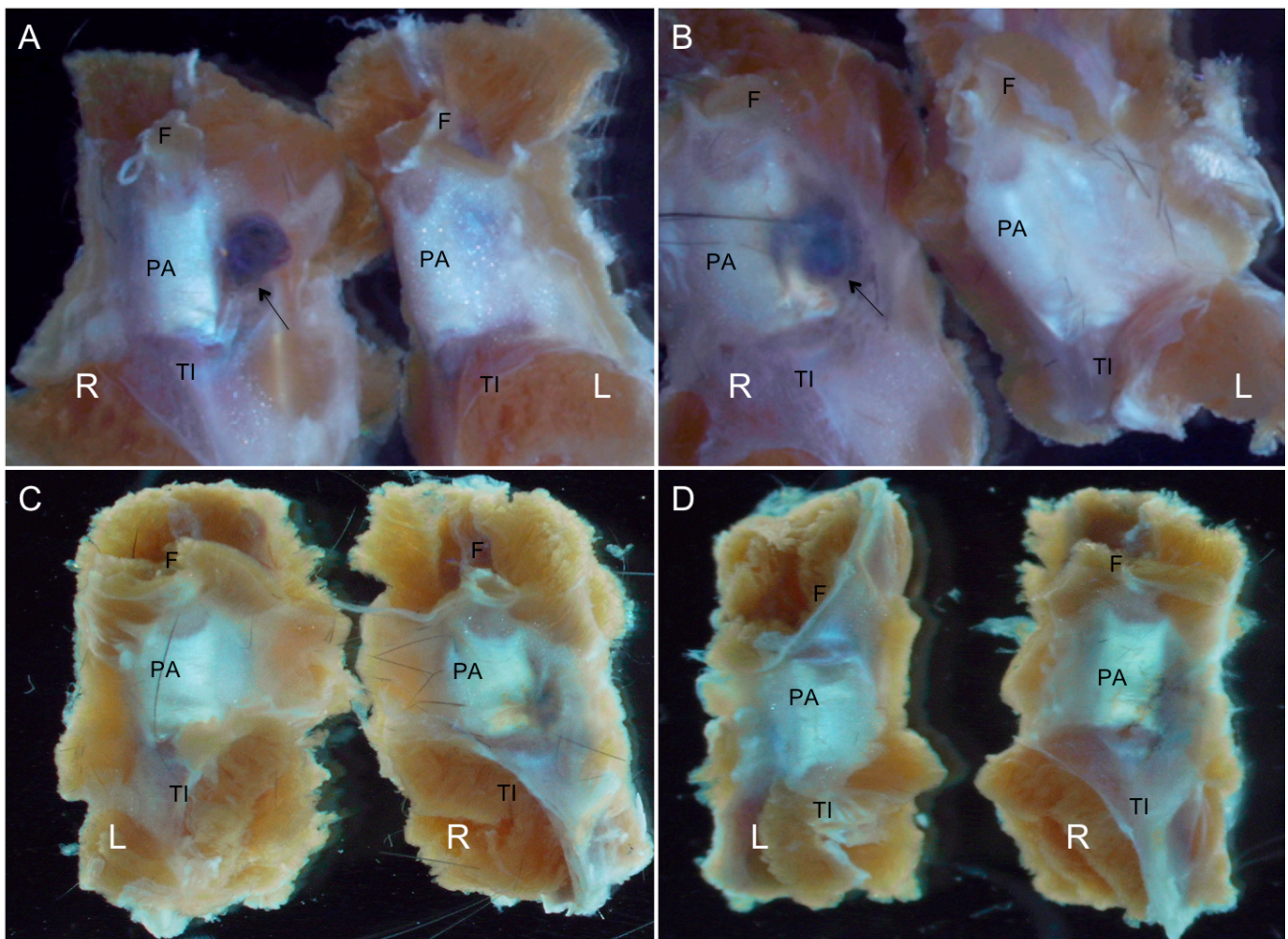


Figure 39: Wholemount operated + unoperated knees from line 12 (β AGG20 line) expressing the β AGG20 transgene (blue staining) only on the medial side of right operated knees 4 weeks post-surgery (black arrows). A and B) Paired operated and contralateral knees from mice sacrificed 4 weeks after DMM. C and D) Paired operated and contralateral knees from mice sacrificed 8 weeks after DMM. L and R represent the Left unoperated and Right operated contralateral knees of a mouse. F (femur), PA (Patella), TI (tibia).

It was apparent when analysing the levels of transgene expression that there was a clear difference in β -galactosidase expression between the different compartments of the operated and unoperated knees in the 4 and 8-week groups (Table 23).

Table 23: Average levels of expression of β -galactosidase in different compartments of the operated and unoperated knees of β AGG20 transgenic mice after overnight staining at room temperature. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

Knee compartment/4 weeks post-operatively	Left unoperated	Right operated	Knee compartment/8 weeks post-operatively	Left unoperated	Right Operated
Lateral articular cartilage	+	+++	Lateral articular cartilage	+	++
Medial articular cartilage	+	+++	Medial articular cartilage	+	++
Lateral tibial growth plate	++	++++	Lateral tibial growth plate	++	+++
Medial tibial growth plate	++	++++	Medial tibial growth plate	++	+++

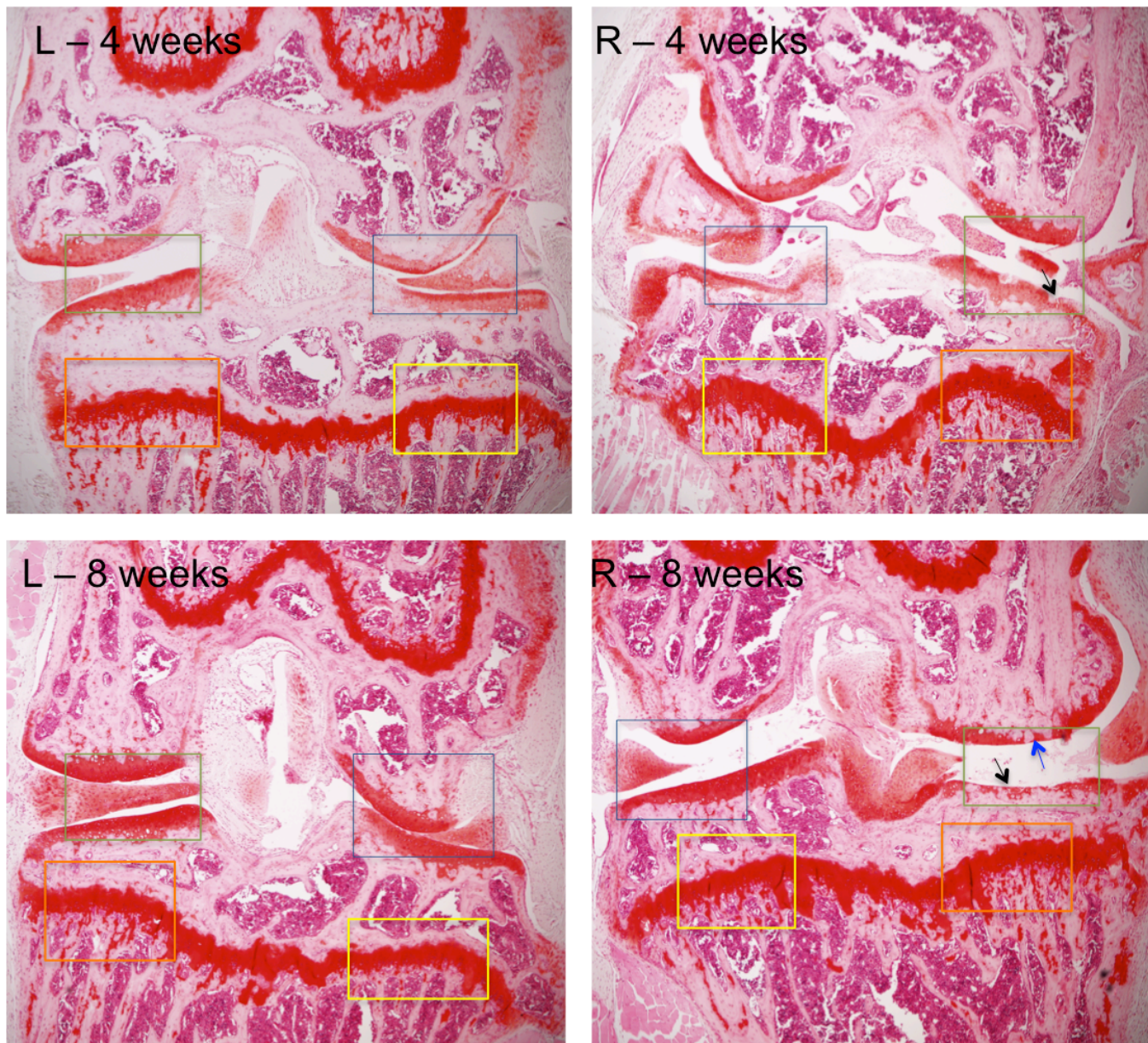
A

Figure 40 (A): Coronal sections of representative line 12 (β AGG20) operated (DMM) right knees (R) and contralateral unoperated left knees (L) 4 and 8 weeks after surgery that have been stained overnight with X-gal. Sections are of whole knees counterstained with safranin O for proteoglycans (bright red) and counterstained with 'Haematoxylin and Eosin' at 4x magnification. A consistent feature of the DMM model was loss of Safranin-O, fibrillation and cartilage loss on the medial tibial plateau (black arrow) and medial femoral condyle (blue arrow). Blue boxed regions represent the lateral articular cartilage of each of the knees (magnified in Figure 40B), green boxed regions represent the medial articular cartilage of each of the knees (magnified in Figure 40C), yellow boxed regions represent the lateral tibial growth plate of each of the knees (magnified in Figure 40D), orange boxed regions represent the medial tibial growth plate of each of the knees (magnified in Figure 40E).

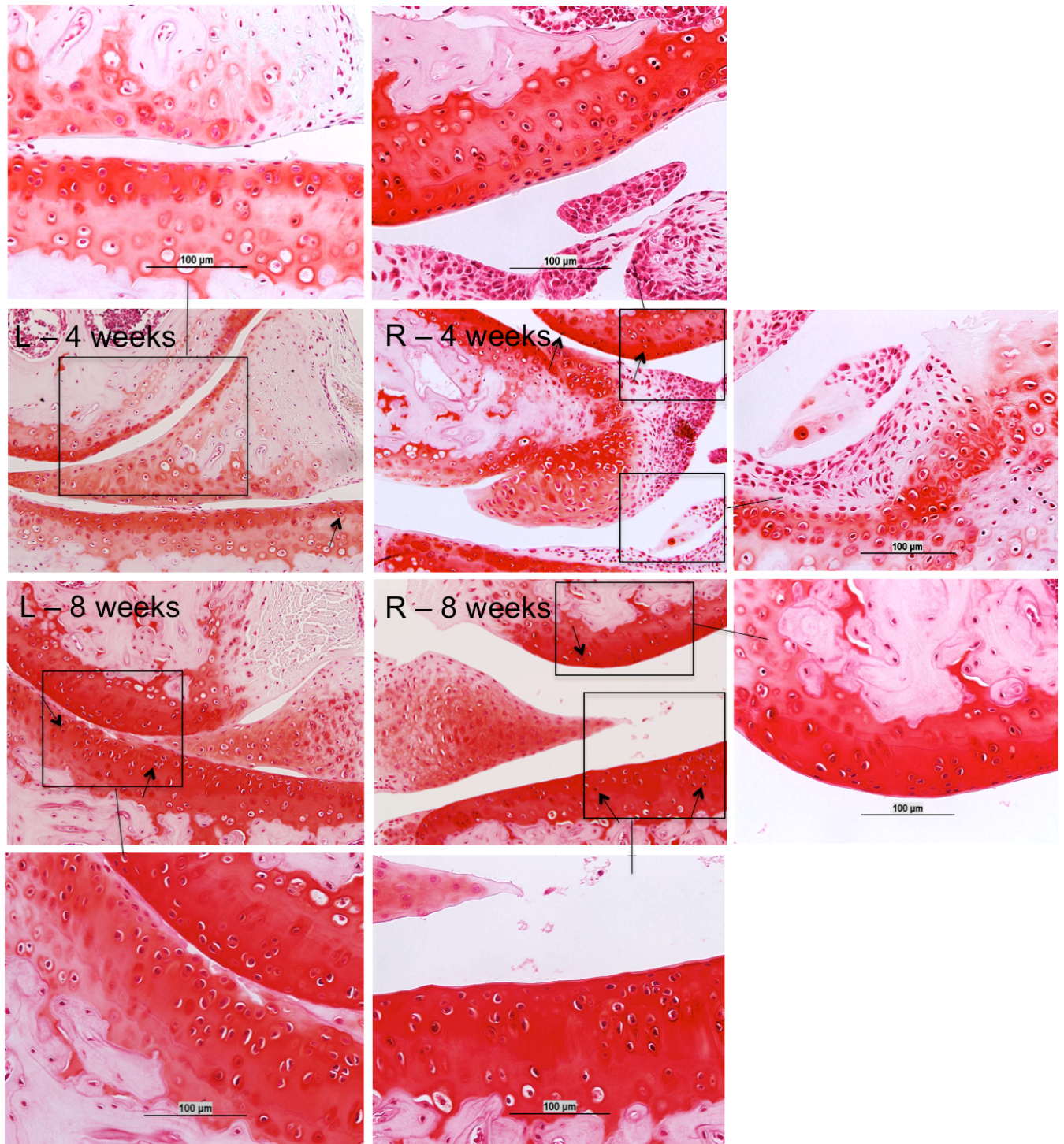
B

Figure 40 (B): High magnifications of coronal sections of the knees in figure 40A above showing the lateral side of articular cartilage. The sections are of representative line 12 (β AGG20) operated (DMM) right knees (R) and contralateral unoperated left knees (L) 4 and 8 weeks after surgery that have been stained overnight with X-gal. Black arrows point to blue stained cells that are positive for β -galactosidase expression. Sections are stained with safranin O for proteoglycans (bright red) and counterstained with 'Haematoxylin and Eosin' at 20x magnification. 40X magnifications of boxed regions are also shown.

C

Figure 40 (C): High magnifications of coronal sections of the knees in figure 40A above showing the medial side of articular cartilage. The sections are of representative line 12 (β AGG20) operated (DMM) right knees (R) and contralateral unoperated left knees (L) 4 and 8 weeks after surgery that have been stained overnight with X-gal. Black arrows point to blue stained cells that are positive for β -galactosidase expression. Sections are stained with safranin O for proteoglycans (bright red) and counterstained with 'Haematoxylin and Eosin' at 20x magnification. 40X magnifications of boxed regions are also shown.

Figure 40 (D) (next page): High magnifications of coronal sections of the knees in figure 40A above showing the lateral side of the tibial growth plate. The sections are of representative line 12 (β AGG20) operated (DMM) right knees (R) and contralateral unoperated left knees (L) 4 and 8 weeks after surgery that have been stained overnight with X-gal. Black arrows point to blue stained cells that are positive for β -galactosidase expression. Sections are stained with safranin O for proteoglycans (bright red) and counterstained with 'Haematoxylin and Eosin' at 20x magnification. 40X magnifications of boxed regions are also shown.

D

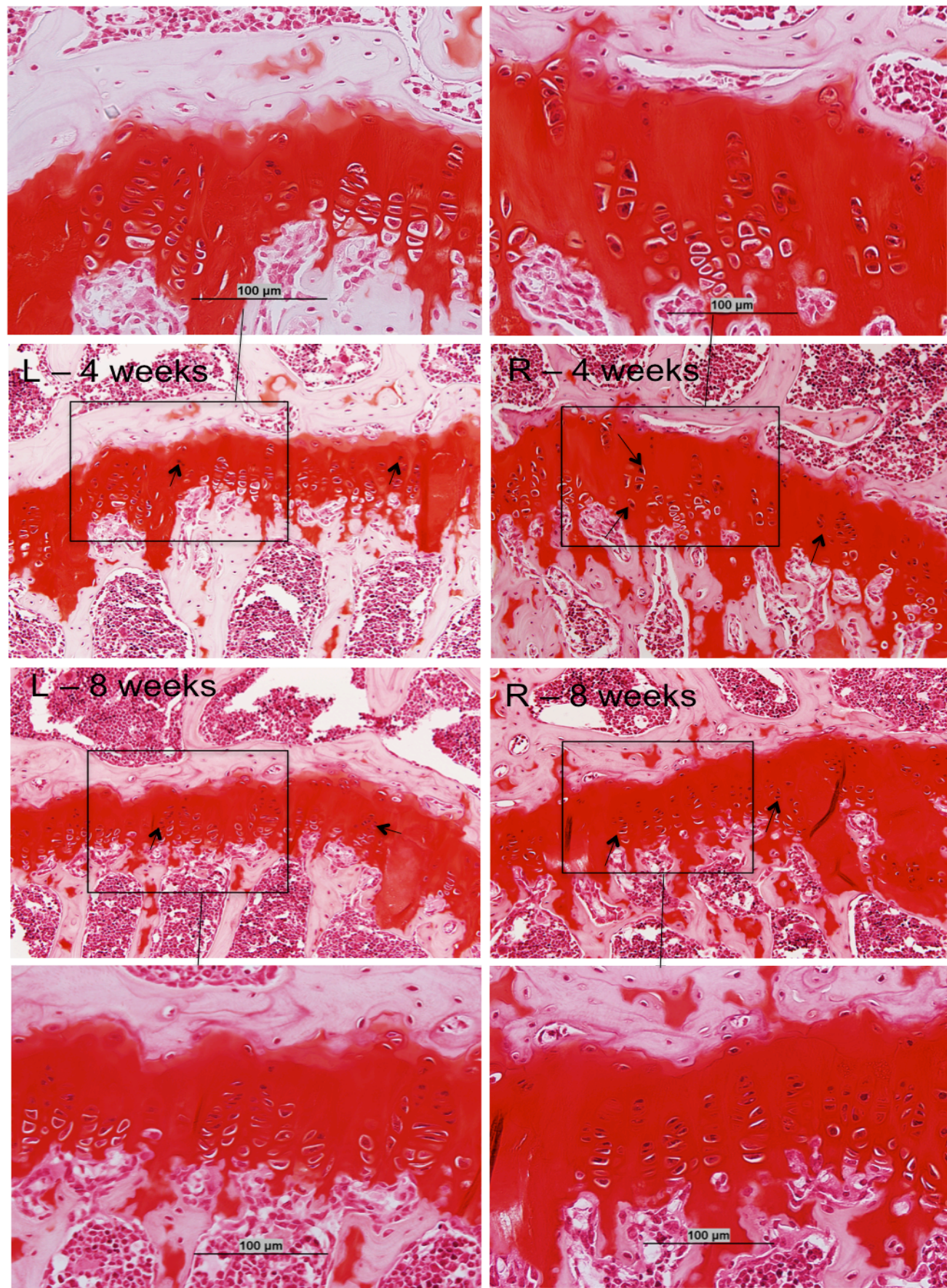
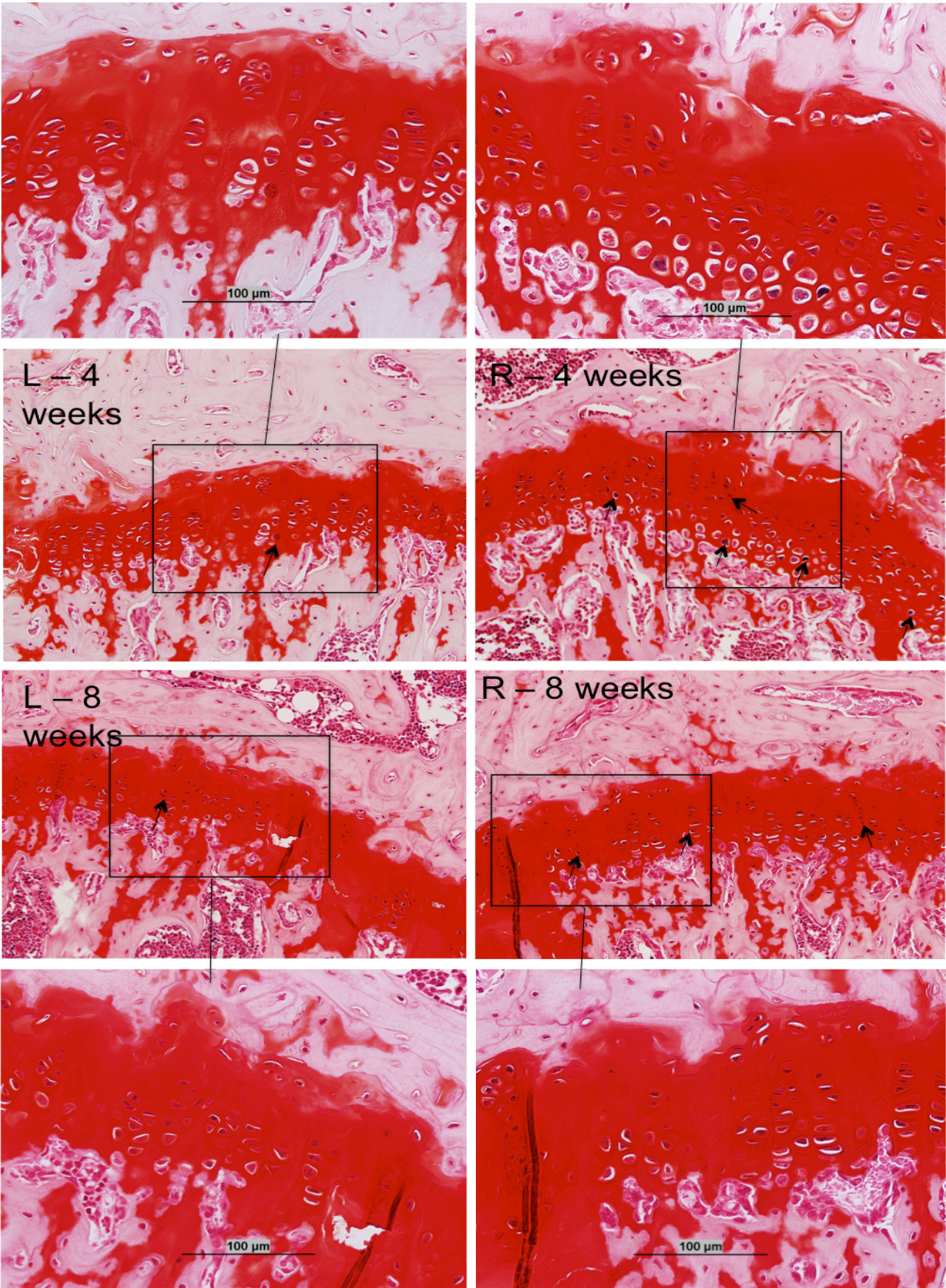


Figure 40 (E) (next page): High magnifications of coronal sections of the knees in figure 40A above showing the medial side of the tibial growth plate. The sections are of representative line 12 (β AGG20) operated (DMM) right knees (R) and contralateral unoperated left knees (L) 4 and 8 weeks after surgery that have been stained overnight with X-gal. Black arrows point to blue stained cells that are positive for β -galactosidase expression. Sections are stained with safranin O for proteoglycans (bright red) and counterstained with 'Haematoxylin and Eosin' at 20x magnification. 40X magnifications of boxed regions are also shown.

E



3.2.2 Axial Loading of line 12 (β AGG20 transgenics)

'Non-invasive axial loading of mice tibias' was performed on transgenic mice from line 12 based on the described method in Results sections 8.2 & 8.2.1. Knees were then analysed based on the described method in Results sections 8.2.1. Unlike the operated knees of the line 12 (β AGG20 transgenics) from the 4 week post-operative group, there was no blue X-gal staining in the right loaded and left unloaded knees in all 12 of the sacrificed animals that could be visualised macroscopically.

For histological analysis of reporter gene expression, all knees (loaded and unloaded) were analysed microscopically and loaded right knees with the highest score (indicating a more biologically aggressive disease) based on The OARSI Osteoarthritis Cartilage Histopathology scoring system were photographed alongside the contralateral unloaded left knees and are shown below in Figure 41 A-E. β -galactosidase antibody staining was also performed for comparison. Figure 41 A-E shows representative images corresponding to the sections with the maximum score (highest grade/stage) for the loaded as well as their contralateral control knees. Also shown are the β -galactosidase stained successive sections. The histological analysis of the knees revealed no signs of OA in the unloaded control knees (grade 0, stage 0) (Figure 41 A-C).

The right knees, which had undergone the non-invasive axial loading, had very mild signs of OA development in the lateral femoral articular cartilage (grade 1, stage 1).

It was apparent when analysing the levels of transgene expression that there was a clear difference in β -galactosidase expression between the different compartments of the loaded and unloaded knees (Table 24).

Table 24: Average levels of expression of β -galactosidase in different compartments of the loaded and unloaded knees of β AGG20 transgenic mice using overnight staining at room temperature with X-gal and β -galactosidase antibody staining. – no staining, + staining just detectable, ++ low level staining, +++high level staining, ++++very high level staining.

Knee compartment	Left unloaded	Right loaded	Knee compartment	Left unloaded	Right loaded
	X-gal staining			B-galactosidase antibody staining	
Lateral articular cartilage (Femoral/ Tibial)	-	-	Lateral articular cartilage (Femoral/ Tibial)	+++	+++
Medial articular cartilage (Femoral/ Tibial)	-	-	Medial articular cartilage (Femoral/ Tibial)	+++	+++
Lateral tibial growth plate	-	++	Lateral tibial growth plate	++	++
Medial tibial growth plate	-	+++	Medial tibial growth plate	+	+++

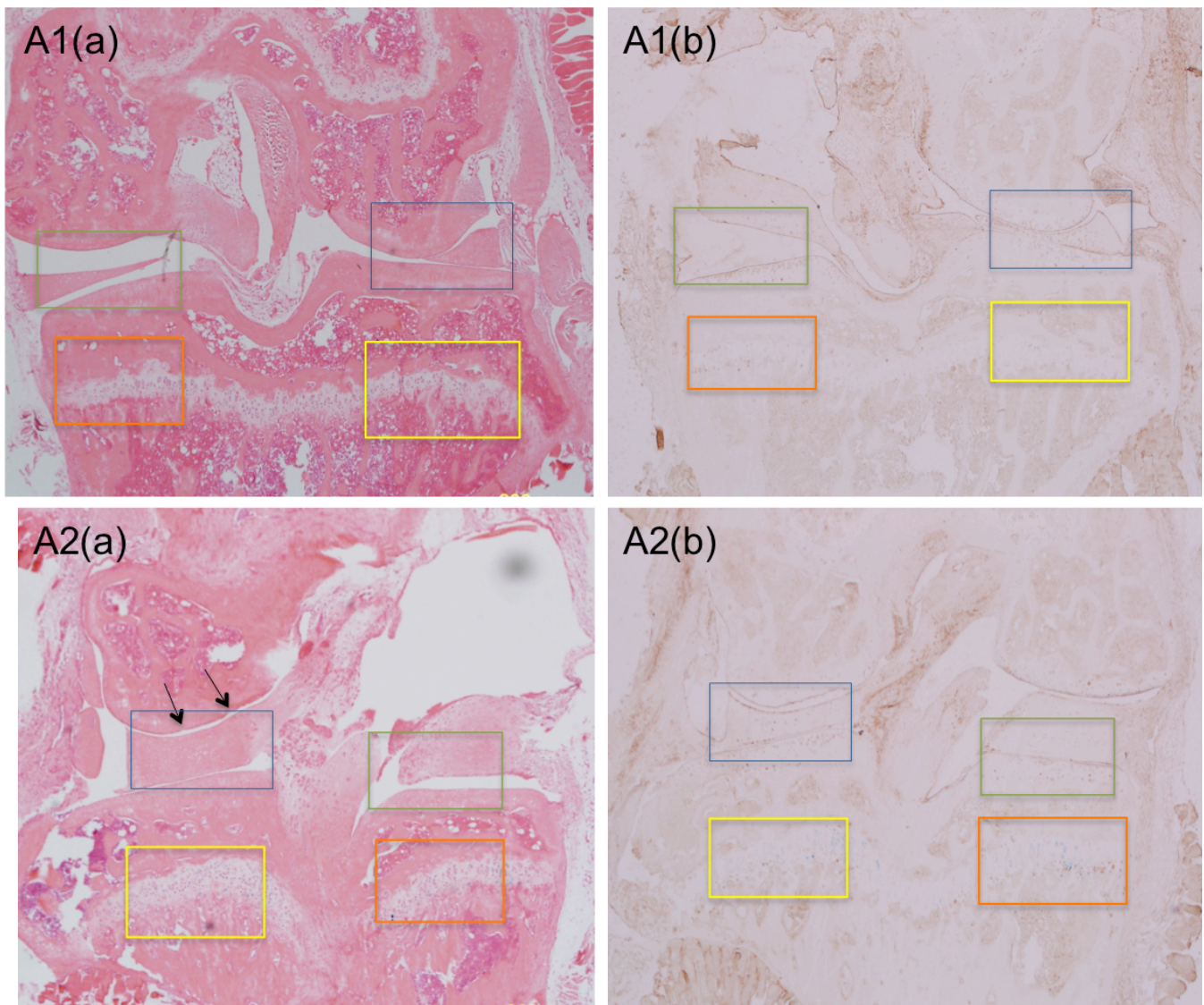
A

Figure 41 (A): Coronal sections of a representative line 12 (β AGG20) loaded right knee (A2) and contralateral unloaded left knee (A1) that have been stained overnight with X-gal and counterstained with eosin or the β -galactosidase antibody. A1+A2(a) 4x magnification of whole knee counterstained with eosin A1+A2(b) 4x magnification of whole knee stained with β -galactosidase antibody (brown staining). A consistent feature of the axial loading model was the loss of cartilage on the lateral femoral condyle (black arrows). Blue boxed regions represent the lateral articular cartilage of each of the knees (magnified in Figure 41B), green boxed regions represent the medial articular cartilage of each of the knees (magnified in Figure 41C), yellow boxed regions represent the lateral tibial growth plate of each of the knees (magnified in Figure 41D), Blue boxed regions represent the medial tibial growth plate of each of the knees (magnified in Figure 41E).

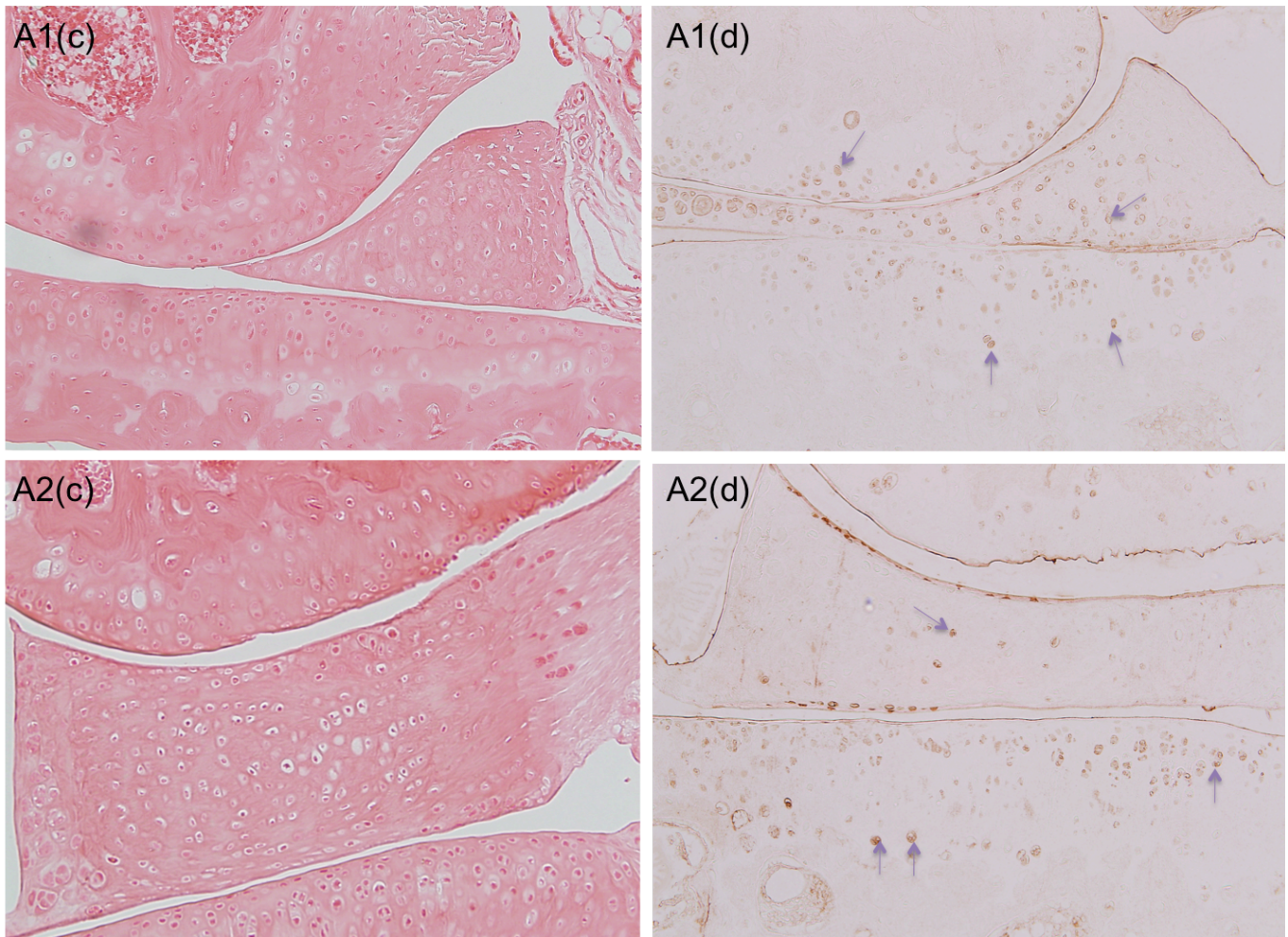
B

Figure 41 (B): High magnifications of coronal sections of the knees in figure 41A above showing the lateral side of articular cartilage. The sections are of a representative line 12 (βAGG20) loaded right knee (A2) and contralateral unloaded left knee (A1) that have been stained overnight with X-gal and counterstained with eosin or the β-galactosidase antibody. A1+A2(c) 20x magnification counterstained with eosin A1+A2(d) 20x magnification of the knee stained with β-galactosidase antibody (brown staining). Purple arrows point to brown cells positive for β-galactosidase antibody staining indicating expression of β-galactosidase.

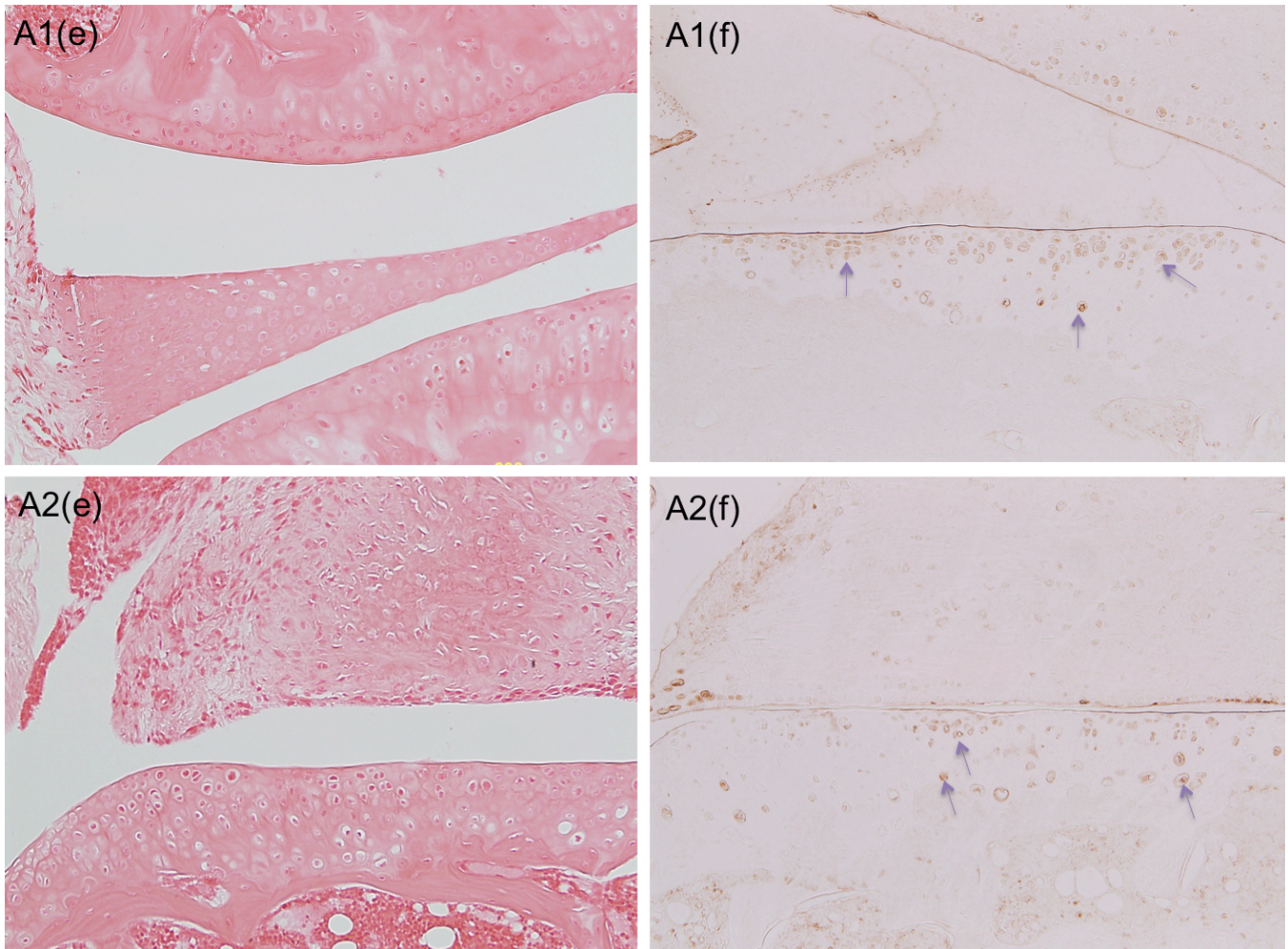
C

Figure 41 (C): High magnifications of coronal sections of the knees in figure 41A above showing the medial side of articular cartilage. The sections are of a representative line 12 (β AGG20) loaded right knee (A2) and contralateral unloaded left knee (A1) that have been stained overnight with X-gal and counterstained with eosin or the β -galactosidase antibody. A1+A2(e) 20x magnification counterstained with eosin A1+A2(f) 20x magnification stained with β -galactosidase antibody (brown staining). Purple arrows point to brown cells positive for β -galactosidase antibody staining indicating expression of β -galactosidase.

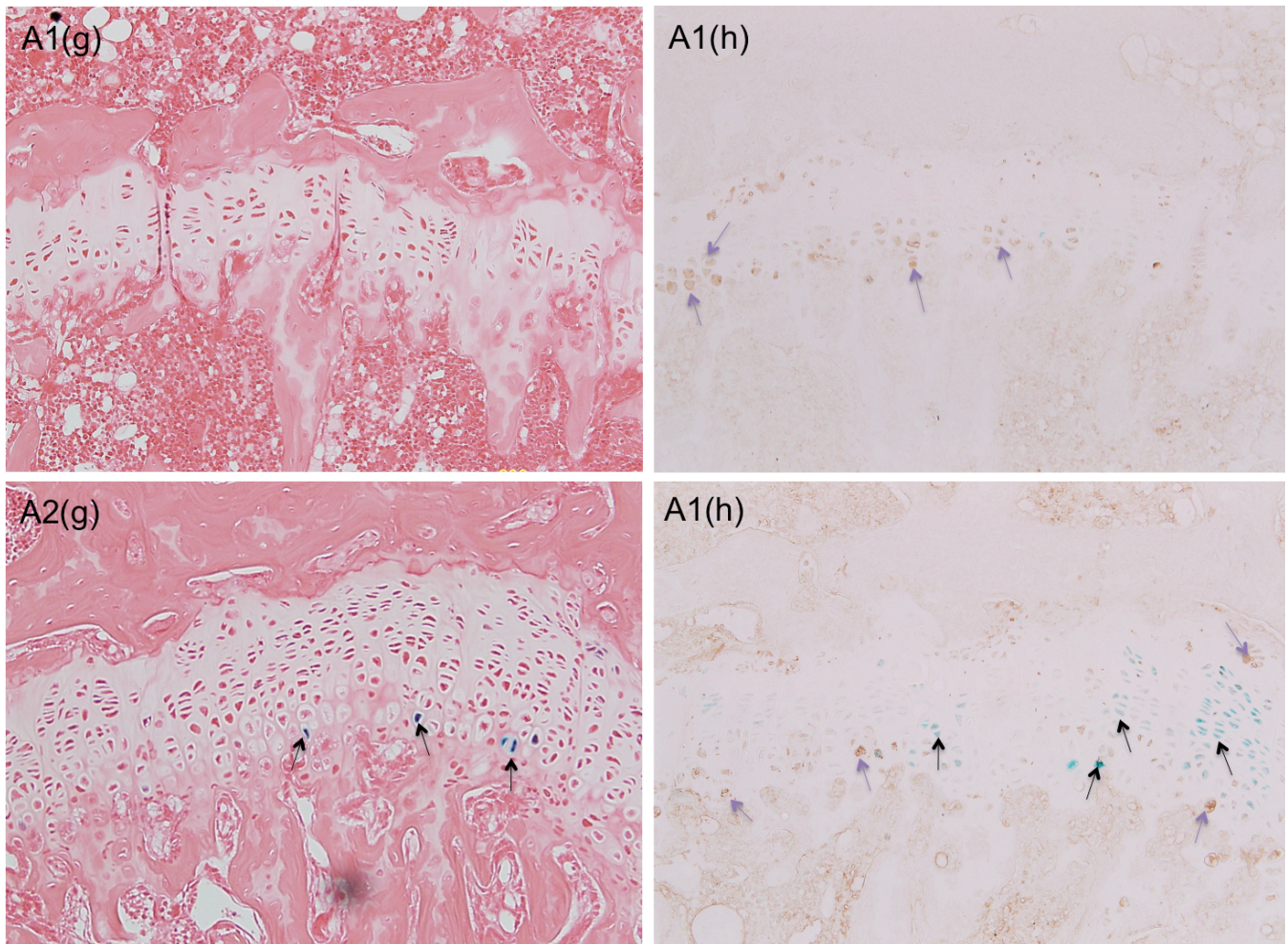
D

Figure 41 (D): High magnifications of coronal sections of the knees in figure 41A above showing the lateral side of the tibial growth plate. The sections are of a representative line 12 (βAGG20) loaded right knee (A2) and contralateral unloaded left knee (A1) that have been stained overnight with X-gal and counterstained with eosin or the β-galactosidase antibody. A1+A2(g) 20x magnification counterstained with eosin A1+A2(h) 20x magnification stained with β-galactosidase antibody (brown staining). Purple arrows point to brown cells positive for β-galactosidase antibody staining indicating expression of β-galactosidase. Black arrows point to blue cells positive for X-gal staining indicating expression of β-galactosidase.

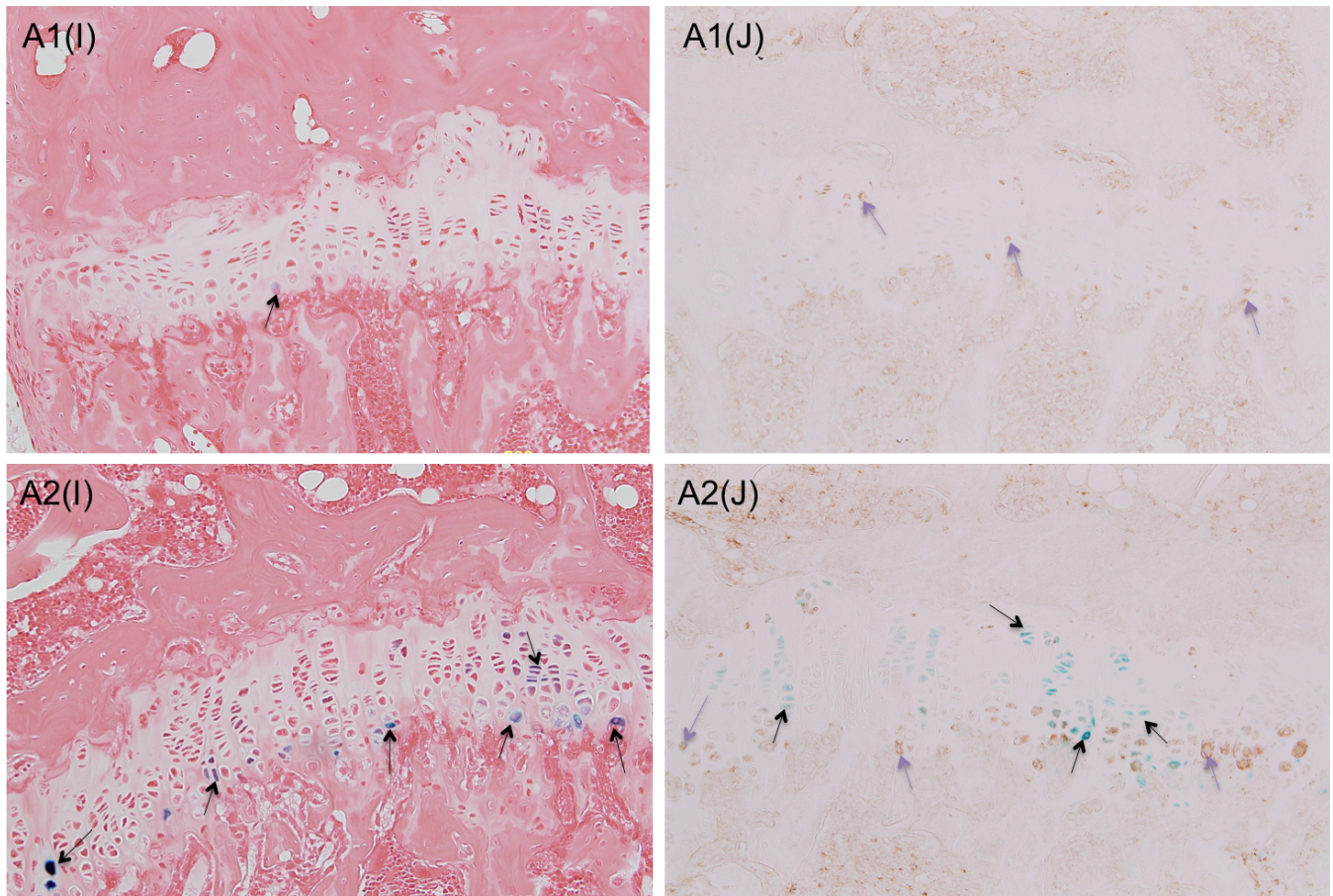
E

Figure 41 (E): High magnifications of coronal sections of the knees in figure 41A above showing the medial side of the tibial growth plate. The sections are of a representative line 12 (β AGG20) loaded right knee (A2) and contralateral unloaded left knee (A1) that have been stained overnight with X-gal and counterstained with eosin or the β -galactosidase antibody. A1+A2(I) 20x magnification counterstained with eosin A1+A2(J) 20x magnification stained with β -galactosidase antibody (brown staining). Purple arrows point to brown cells positive for β -galactosidase antibody staining indicating expression of β -galactosidase. Black arrows point to blue cells positive for X-gal staining indicating expression of β -galactosidase.

3.2.3 DMM of line 39 (Col14xA1Cre transgenics)

DMM was performed on 5 transgenic mice from line 39 and one wild-type littermate (control) based on the described method in Methods sections 8.1 & 8.1.2. Knees were then analysed based on the described method in Methods section 7.2.2.2. Figure 41A shows bioluminescent images of a representative col14xA1 mouse before surgery (week 0) and 1,2,3,4 and 8 weeks post-operatively. The luminescence emitted by the knees of each mouse was visualized as an orange pseudo-colour and so the luciferase/Col14xA1Cre promoter activity was measured by collecting pseudo-colour intensity values for a defined region of interest (ROI) occupying the whole knee (identified with corresponding X-ray images). This was done for all 5 mice and the mean pseudo-colour intensity values for the left and right knees for each week were calculated. A comparison of the mean pseudo-colour (reflecting luciferase/promoter activity) between the right operated knees (RK) and the contralateral left unoperated knees (LK) before surgery (0 weeks) and at 1,2,3,4 and 8 weeks post surgery was performed by subtracting the pseudo-colour intensity values of the left unoperated knees from the right operated knees (RK-LK) (Figure 41B). The mean pseudo-colour values (reflecting luciferase/promoter activity) of other cartilaginous tissues including the left ankles (LA), the left footpads (LF) and the left tibias (LTb) at 0 weeks (before surgery) and at 1,2,3,4 and 8 weeks were also collected alongside the left knee values and are depicted in the graph in Figure 41C. The same is shown for the right side in the graph in Figure 41D. The graphs in Figure 41C+D show that while the mean pseudo-colour values of the left and right ankles, footpad and tibias vary from week to week, this is most probably due to experimental limitations as opposed to differences in promoter activity since the pattern of variation from week to week is extremely similar for all these examined regions. However, for the right knees but more so for the left knees, the pattern of mean pseudo-colour value variation from week to week is different to that of all the other examined regions indicating that this is most likely due to variations in promoter activity as opposed to experimental limitations.

Before surgery and at 1 week post surgery, the pseudo-colour intensity/luciferase activity in the left unoperated knees was higher than the right operated knees (negative values in Figure 41B). At these timepoints, the right operated knees showed no activation or repression of the promoter since the mean pseudo-colour intensity values were similar to those of the right ankles, tibias and footpads. However, the pseudo-colour intensity values of the left unoperated knees were lower than the average pseudo-colour intensity values of the left ankles, footpads and tibias. Around 2-3 weeks post-surgery, the difference in mean pseudo-colour intensity values between the right operated knees and left unoperated knees (RK-LK) were positive (Figure 41B). This was due to a repression in promoter/luciferase activity in the left unoperated knees as opposed to an increase in promoter activity in the right operated knees. At 4 weeks, the promoter activity of the left unoperated knees increased and the difference in mean pseudo-colour activity between right and left knees (RK-LK) became negative again (Figure 41B). Up until this point, the differences in luciferase activity between the right operated knees and left unoperated knees was due mostly to variations in promoter activity in the left unoperated knees. At 8 weeks however, the luciferase/promoter activity in both the left and right knees was activated compared to all other regions examined and there was no difference in pseudo-colour intensity/luciferase activity between them. The wild-type littermate showed no orange pseudo-colour/luciferase activity at any time point.

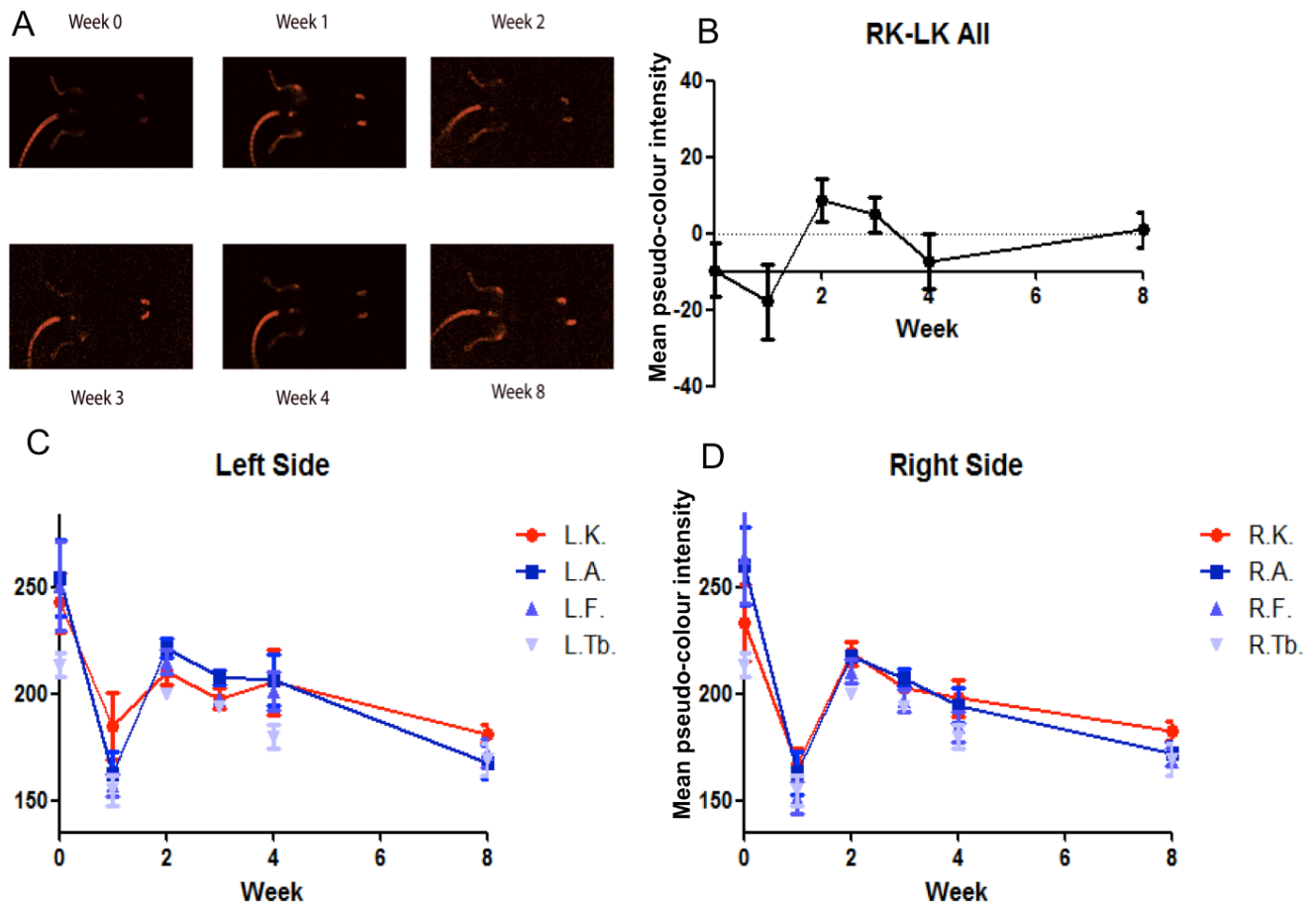


Figure 42: Results of bioluminescence imaging of luciferase activity in adult transgenic mice from the *Col14xA1Cre* transgenic line before and after DMM surgery A) Dorsal images from a representative adult transgenic mouse before surgery (week 0) and 1,2,3,4 and 8 weeks after DMM showing luminescence/luciferase activity in cartilaginous areas including the knees (orange pseudo-colour). B) Graph depicting the difference in mean measured pseudo-colour intensity (luciferase activity) $\pm$ s.d between DMM right operated knees (RK) and contralateral left unoperated knees (LK) at several timepoints. C) Graph depicting the mean measured pseudo-colour intensity (luciferase activity) $\pm$ s.d of the left knees (LK), left ankles (LA), left footpads (LF) and left Tibias (LTb) at several timepoints. D) Graph depicting the mean measured pseudo-colour intensity (luciferase activity) $\pm$ s.d of the right operated knees (RK), right ankles (RA), right footpads (RF) and right tibias (RTb) at several timepoints. Luciferase activity was measured for a set region of interest (ROI) on both the left and right knees, ankles, footpads and tibias by retrieving pseudo-colour intensity values (arbitrary units). 5 mice were used for weeks 0-4 and 3 mice were used for week 8.

3.3 Discussion

The pathophysiology of OA can be further elucidated by murine OA models as several of the biochemical features observed in murine OA models, such as matrix proteoglycan depletion by extracellular proteinases, are similar to changes found in human OA cartilage. In this chapter, I used two of these models to elucidate the response that the regulatory fragments in the 'βAGG20' and 'ColII4xA1Cre' transgenic lines 12 and 39 respectively have when subjected to injury. I did this since both regulatory regions in theory drive an 'aggrecan type' expression and aggrecan is one of the major components of cartilage whose degradation/dysregulation is indicative of cartilage extracellular matrix destruction.

3.3.1 Transgene regulation in βAGG20 transgenics after DMM

Previous studies on aggrecan gene expression levels from OA joints have been quite varied possibly due to the differences in stages and grades of OA at which point these levels were analysed. For our purposes, we used both the 'axial loading of mouse tibias' model and the 'DMM surgery' model to induce OA in mouse knees from line 12.

The DMM model was selected, as it is an established and reliable technique in the institute. It is well known that this particular model of cartilage degradation is characterized by initial aggrecan depletion of intact cartilage followed by fibrillation, clefting and subsequent loss of cartilage down to the tidemark. This depletion is due to the increased activity of several proteinases, in particular the aggrecanase ADAMTS5 (Glasson et al., 2005). Since the transgenic product driven by the βAGG20 fragment in line 12 is the β-galactosidase protein as opposed to the aggrecan protein and hence not degraded by aggrecanases, it is possible to hypothesise that the differing levels of β-galactosidase protein in the chondrocytes is a reflection of the varied regulation of the transgene at a transcriptional rather than post-

translational level. Unfortunately, it is not possible to assume that the level of β -galactosidase protein within a chondrocyte corresponds to the level of β -galactosidase mRNA within that chondrocyte at a particular timepoint. This is due to the relative high stability of the β -galactosidase protein. One group showed that β -galactosidase activity decays with a half-life of approximately 13 hours (Jacobsen and Willumsen, 1995) while another estimated a maximum half life of 24-48 hours (Smith et al., 1995).

DMM was performed on 8 week old line 12 mice and the operated as well as control contralateral knee analysed 4 and 8 weeks post-operatively. In the 4 week post-operative group, several points could be made: a) in the whole mount X-gal stained knees, blue staining was only found macroscopically in the medial injured side of the operated knee; b) the expression of the transgene was significantly higher in the medial and lateral tibial growth plates of the operated knee when compared to the same compartments within the contralateral non-operated knee; c) the expression of the transgene was higher in the medial and lateral articular cartilage of the operated knee when compared to the the entire articular cartilage of the contralateral non-operated knee, both of which had similar levels of expression ; d) the overall transgene expression in the articular cartilage and the tibial growth plates of the contralateral non-operated knee was higher than that in the naïve (pre-operative) knee (Figure 32). In the 8 week post-operative group however, the other points to be made are: a) in the whole mount X-gal stained knees, there was no blue staining to be found macroscopically in the operated or contralateral knee; b) the overall expression of the transgene decreased in the lateral and medial tibial growth plates of the operated knee compared to its expression in these compartments in the 4 week post-operative knee; c) the overall expression of the transgene decreased in the articular cartilage of the operated knee compared to its expression in the 4 week post-operative knee but was still higher than the expression in the contralateral knee; d) the level of transgene expression in all compartments of the contralateral knee remain at a

similar level to the level of expression within the corresponding compartments of the contralateral knee in the 4 week post-operative group.

The various studies that have focused on the analysis of gene expression in OA versus control articular cartilage somewhat support the results discussed above. The increase in transgene expression seen in the articular cartilage of the operated knee in comparison with the contralateral unoperated knee at **4 weeks** has also been demonstrated in a canine model of osteoarthritis. In this particular study by Matyas et al, northern blot analysis revealed that expression of aggrecan in canines was 2 times higher in the articular cartilage of the knees/OA joints produced by anterior cruciate ligament transaction compared to the unoperated contralateral knee joints. The operated knee joints were analysed 10 weeks after surgery, which in dogs, corresponds to early experimental OA. The group suggested that while both the aggrecan mRNA and type II collagen mRNA are elevated (8 fold for type II collagen), the disproportionate levels of synthesis of both mRNAs in this early phase of OA in this animal model is a metabolic characteristic of early OA that may contribute to its pathogenesis (Matyas et al., 1995). An additional gene expression analysis was performed by Annika Burleigh at the KIR (PhD student) in which the mRNA levels within the articular cartilage of surgically induced mouse knees (by DMM) were compared with that of naive, sham and contralateral non-operated knees. It was shown that aggrecan is down regulated initially post DMM surgery (6hrs-7days) however by two weeks it was upregulated 1.8 fold above naive and sham limbs. This rose to 12.84 fold at 4wks (personal communication).

The upregulation of *LacZ* expression in the articular cartilage of the operated line 12 knees was localised to the medial compartment, which corresponds to the site of injury (more lesions and signs of deterioration at the medial side of the joint). Also, it was clear when analysing operated joints from different mice that the upregulation of the transgene was much less

marked in operated joints that had lower scores. It is possible therefore that the increase in β AGG20 promoter activity is proportional to the amount of damage found within the articular cartilage at this particular timepoint.

The expression of the transgene was also significantly higher in the medial and lateral tibial growth plates of the operated knee when compared to the same compartments within the contralateral non-operated knee. This was true for the whole depth of the growth plate but the most intense blue staining was found in the deeper zones containing the hypertrophic chondrocytes. As with the expression of the transgene in articular cartilage, the level of *LacZ* expression in the tibial growth plate was also proportionate to the amount of damage found within the joint at this time point. These growth plate results challenge the notion that abnormalities in the growth plate are not related to nor have a role in the development of OA. This notion was proposed by Chambers et al who showed via *insitu* hybridization that both CBA and STR/ort mice (spontaneously develop OA) had similar patterns of expression of aggrecan in the growth plate (Chambers et al., 2002). This difference could be due to the fact that the STR/ort mice have different mechanisms by which they develop OA than mice that have undergone surgery or axial loading.

The decrease in the expression of the transgene in the articular cartilage of the operated knee at **8 weeks**, compared to its expression in the articular cartilage at 4 weeks post-operatively also corresponds to some recent studies performed by Brew et al (2010). These experiments focused on the gene expression in the knees of patients with OA that had undergone knee replacements (aged 60-84), which was analysed by quantifying the RNA extracted from the articular cartilage of these knees by PCR. The expression of aggrecan was reduced in OA cartilage in comparison with control tissue (4 fourfold lower) and interestingly the expression of Sox9, the well-known transcriptional regulator of aggrecan, was even more greatly reduced in

OA cartilage in comparison with control tissue (38-fold lower). Both genes were downregulated at the intact as well as the damaged sites within the articular cartilage. The gene expression profiles at both the damaged and intact sites of cartilage on an OA joint were similar to each other and significantly different from those of age-matched controls (Brew et al.).

It is important to note that, unlike the gene expression analysis performed by Matyas et al which focused on early experimental OA, the analysis performed by Brew et al (2010) provides an assessment of gene expression in cartilage late in the OA process at joint replacement. The differences in gene expression at both experimental OA timepoints may thus reflect a pattern of failed remodelling of the joint, which at the late stage shows abnormal matrix production (aggrecan), and potentially enhanced matrix turnover. The pattern of *LacZ* staining in the operated knee shows that its expression is elevated initially at 4 weeks, when the joint has mild-moderate OA (early experimental OA), and suppressed at the later 8-week timepoint, when the joint develops a more moderate phenotype. This moderate phenotype has not reached a stage that is late in the OA process but is obviously more developed than the mild-moderate OA present in the operated knees of the 4-week timepoint. This is probably why the expression of the transgene is reduced but has not reached below the control contralateral knee as is true for the Brew et al (2010) study. The expression of aggrecan may change during progression of OA from initial enhancement, to later suppression.

At the 8-week timepoint, there is also a decrease in the protein levels within the medial and lateral condensed tibial growth plate, in which it is more difficult to assign cells to different zones, when comparing it to the transgene expression in the tibial growth plate at 4 weeks post-operatively. This reflects the dysregulation of transgene expression that occurs within the articular cartilage between the two timepoints and further supports the idea that the growth plate may play a role in the development of OA.

Another important point to make is that the overall transgene expression in the articular cartilage and the tibial growth plates of the contralateral non-operated knee was slightly higher than that in the naïve knee. This was also shown by A Burleigh (KIR), who found that although no significant difference was found in the sham limb compared to naïve, some up regulation was also seen in the contralateral non-operated limb. Aggrecan was 8.59 fold above naïve in the contralateral limb 4 weeks post DMM (unpublished data). This is possibly due to the fact that the contralateral limb revealed very low levels of OA (grade 1, stage 1).

All operated knees from the 4-week post-operative group showed very similar X-gal staining patterns, as was also the case for the knees from the 8-week post-operative group. This was also true for the control contralateral knees of both groups.

3.3.2 Transgene regulation in β AGG20 transgenic after Axial Loading

The 'axial loading of mice tibia' model was selected as it involves simple and non-invasive *in vivo* repetitions of loads on mice tibial plateaus on alternate days for 2 weeks that eventually lead to the development of lesions within the articular cartilage of the loaded joint. This particular model is characterized by significant load magnitude-related increases in cortical bone formation. Also, Micro-CT scans show that loading significantly increases trabecular bone volume in 8 week old mice, but modifies trabecular organization with decreases in trabecular bone volume in 12 and 20 week old mice (De Souza et al., 2005). The axial loading was performed on 8-week-old mice and the loaded as well as control contralateral knee analysed after the final loading 2 weeks later. The loaded and contralateral knee sections were also stained with the β -galactosidase antibody.

At first glance, when looking at the eosin counterstained sections, it appears that, like the naïve knee, there is no β -galactosidase activity (no blue staining) within the articular

chondrocytes in either the lateral or medial compartments of the loaded and contralateral unloaded knee. However, staining with the β -galactosidase antibody actually revealed that a significant proportion of the articular chondrocytes down to the tidemark stained positive. *LacZ* reporter gene detection by antibody staining of the β -galactosidase protein is a much more sensitive way of detection than an enzymatic assay using the X-gal substrate. This is because the amount of β -galactosidase protein present within each cell must reach a certain threshold before it can efficiently process the substrate. However, antibody staining can detect minute quantities of the protein. Nevertheless, the medial and lateral articular cartilages in both loaded versus control knees revealed no visible quantitative difference in β -galactosidase antibody staining.

This model is an early model of OA and as such is better compared to the DMM model at the earlier timepoint of 4 weeks post-operatively. It is interesting that the more invasive DMM model, which caused a lot more damage to the articular cartilage, led to higher reporter gene expression in operated knees overall. It is possible to make this comparison as the β -galactosidase protein was detected in the DMM operated knees by the less sensitive detection method of an enzymatic assay. Even the articular chondrocytes of the contralateral knees, which also have some blue staining, seem to have a higher expression of *LacZ* when compared to the loaded knees and their controls. This is possibly due to the fact that they have developing OA as well that may be slightly more advanced than the loaded knee. This further supports the notion that more deteriorated articular cartilage in early experimental OA is associated with higher levels of reporter gene expression.

In the same way that the expression of the transgene was higher in the medial and lateral tibial growth plates of the DMM operated knee compared to the control non-operated knee, there was also an increase in transgene expression in the medial and lateral growth plates of the

loaded knee versus the control knee as indicated by the blue staining and β -galactosidase antibody staining. This was true for the whole depth of the growth plate but the darkest blue stained and antibody stained cells were found in the deep zone containing the hypertrophic cells. Although there was a clear difference in the expression of the transgene between the loaded growth plate and the contralateral unloaded growth plate, the overall expression in either knee was lower than the 4 and 8 week postoperative DMM operated knee and contralateral knee since fewer chondrocytes were stained. The activity of the β AGG20 promoter could be a function of the length of time that the knee was loaded and the extent of the loading itself. The DMM operated knees have the most developed signs of OA as the instability caused by severing the medial meniscus leads to increased loading on the joint. The contralateral knee to the DMM operated knee has increased loading, as the mouse must rely more on it for movement. Also the DMM operated and contralateral knees were analysed 4 and 8 weeks post-operatively. In contrast to this, the axial loaded knee was loaded a total of 6 times for 2 weeks only. The loading must be less invasive as is demonstrated by the extent of damage found within the articular cartilage. The transgene expression pattern in the contralateral unloaded knee bears the closest similarity to the naïve knee. The hypothesis that the response of the β AGG20 promoter, which in theory should mimic the activity of the aggrecan promoter, in the chondrocytes of the growth plate is related to the loading of the knees is further supported by the fact that STR/ort mice, which develop OA by a completely different mechanism that is not related to loading and more a result of genetics, had similar patterns of expression of aggrecan in the growth plate (Chambers et al., 2002). The increased expression of aggrecan in the growth plate may have a protective role as it is unusual that the axial loading model only leads to lesions in the lateral femoral articular cartilage and not the lateral tibial articular cartilage. It is possible that as OA develops, the growth plate becomes less protective due to suppression in aggrecan gene activity.

On a final note, the response that the β AGG20 transgene has to the induction of injury in transgenic mice could be due to particular sites found within the 3' end of *c/s* 5. The expression of aggrecan can be modulated by mechanical forces, cytokines, and other serum factors. There are two NF- κ B sites in the promoter that may confer responsiveness to cytokines. This is suggested by the overlapping NF- κ B and STAT binding sites at +362 - +375bp relative to the aggrecan TSS and the fact that NF- κ B is often activated by cytokines such as tumor necrosis factor. There are also four SSREs (one in the promoter and three in the 5'UTR within *c/s* 5), which are potential sites for regulating aggrecan gene expression by shear stresses.

All the loaded knees showed very similar X-gal staining as well as β -galactosidase antibody staining patterns. This was also true for the control contralateral unloaded knees.

3.3.3 DMM of line 39 (Col114xA1Cre transgenics)

DMM was performed on 5 8-week-old line 39 mice and the operated as well as control contralateral knees analysed by whole-mount bioluminescence imaging before surgery (week 0) and at 1,2,3,4 and 8 weeks post-operatively. The purpose of this study was to establish whether this promoter remains active and able to deliver Cre recombinase gene expression after the induction of OA and whether it is upregulated or downregulated within the injured cartilage of the right operated knee compared with the contralateral left unoperated knee. The luminescence emitted by the knees of each mouse was visualized as an orange pseudo-colour and so the luciferase/Col114xA1Cre promoter activity was measured by collecting pseudo-colour intensity values for a defined region of interest (ROI) occupying the whole knee. The comparison between the Col114xA1Cre promoter activity in the cartilages of the right and left knees at each timepoint was done by subtracting the mean pseudo-colour intensity values of the left unoperated knees from the right operated knees (Figure 41B). A value of 0 would

indicate no difference in promoter activity between both knees. A positive value would either indicate higher promoter activity in the right operated knees compared with the left unoperated knees or lower promoter activity in the left unoperated knees compared to the right operated knees. A negative value, on the other hand, would either indicate higher promoter activity in the left unoperated knees compared with the right operated knees or lower promoter activity in the right operated knee compared with the left unoperated knee. In order to verify the cause of the differences in mean pseudo-colour intensity values/promoter activity between the contralateral knees, the mean pseudo-colour values (reflecting luciferase/promoter activity) of the left and right ankles, footpads and tibias were also collected (Figure 41C+D). This was achievable since the variations in mean pseudo-colour values across these regions from week to week followed a similar pattern indicating that these variations were due to experimental limitations such as the luciferin substrate on the day, the time it takes for the mice to be anaesthetised and any slight changes in the weight of the mice.

Interestingly, it was the left unoperated knees as opposed to the right operated knees that showed variation in promoter activity until 4 weeks post surgery. The activity of the promoter increased slightly 1 week post surgery but then decreased 2-3 weeks post surgery. The initial increase was probably due to the mouse using its left unoperated knee more due to the injury in the right operated knee. By 8 weeks, however, the promoter activity in both the left unoperated knees and right operated knees had risen by a similar amount. This somewhat mimics the DMM study performed on the β AGG20 transgenics in that the activity of the promoter in both the right operated knees and the left unoperated knees was higher at 8 weeks compared with the naïve knee. However, in the case of the β AGG20 transgenics, there was a difference in promoter activity between the two contralateral knees (promoter activity was higher in the right operated knees) and unlike the Col11a4xA1Cre transgenics, the overall promoter activity in the operated knees was lower at 8 weeks compared to 4 weeks post

surgery. Using the *Col14xA1Cre* promoter would be potentially useful for delivery of Cre recombinase to the joint since its activity is not reduced in the early stages of experimental OA. Instead, it is upregulated by early-middle stages of OA development.

Conclusions

Using mouse transgenic experiments, an approximately 20kb sequence (between -20,054bp and +375bp relative to the aggrecan transcriptional start site including a deleted 92 base pair sequence between -62bp and +25bp) was demonstrated to have chondrocyte specific regulatory activities *in vivo*. The regulatory activity of this sequence begins at E13.5 (Figure 30.1D) where reporter gene expression can be found in the developing cartilages of the long bones of the hind and forelimbs. At E14.5, this sequence can drive reporter gene expression in the developing cartilage of the long bones of the fore and hind limbs, the ribs, the scapulae, the digits and some craniofacial cartilages (meckel's cartilage) but this does not extend to the vertebral cartilage (gene expression is expected at this stage of development) (Figure 17 and Figure 30.1A&B). At E16.5, the ~20kb sequence can drive gene expression in all the anatomical regions mentioned above but with a much higher intensity. Also, the gene expression at this stage of development extends to include more facial cartilages such as the nasal cartilage, the cartilage of the ear and the cartilage of the eye socket as well as the vertebral bodies and intervertebral disc. However, there is no expression in the chondrocranium, which is expected at this stage of development (Figure 19). Furthermore, the promoter activity of this ~20kb sequence is extremely weak in the adult transgenic line harbouring this DNA sequence and is limited to the intervertebral discs and growth plate of the tail and a few randomly scattered chondrocytes in other cartilaginous structures of the hip and the knee (Figure 30.2 and 31).

Progressive deletions to this ~20kb sequence have enabled us to hypothesise the roles of different potential *cis*-regulatory elements that were selected due to the assumption that regulatory elements lie within sequences that have high evolutionary identity across species. The minimal promoter/*cis* 5, which corresponds to the sequence from -452bp to +375bp

relative to the start site, seems to be inactive on its own *in vivo*. *Cis* 3 (-10.9 to -10.2kb), which corresponds to the mouse enhancer sequence A1 identified by Han and Lefebvre during the course of this study (Han and Lefebvre, 2008), has been shown to have 3 binding sites for L-sox5/sox6 dimers and one binding site for a sox9 dimer. These transcriptional activators may recruit different coactivators such as HATs to drive transcription of aggrecan. We hypothesise that *cis* 4 (-3.8 to -3.7kb) could be a silencer element with a binding site(s) for repressors that can recruit corepressors such as HDACs and HMTs in order to inhibit the transcription of aggrecan. We also hypothesise that *cis* 1 and/or 2 (-20 to -19.7kb and -14.5 to -14.4kb respectively), have binding sites for transcriptional activators that can potentially recruit tissue-specific cofactors that function alongside the positive cofactors recruited by *cis* 3 to counteract the transcriptional inhibition caused by *cis* 4. We also investigated the role of *cis* 6, a 3kb sequence present within the first intron of the gene (+5kb to +8kb). When fused downstream of the ~20kb sequence, it downregulated the activity of the overall promoter. It could possibly be a silencer element that inhibits the transcription of the gene. See Figure 43 for proposed roles of *cis* 1-6.

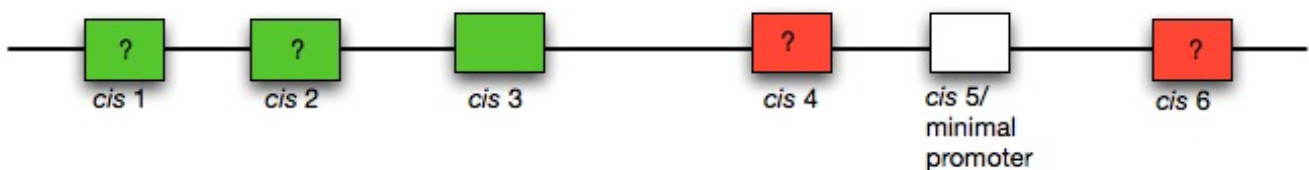


Figure 43: Proposed action of ‘*cis*-regulatory elements’ *cis* 1-6. Green boxes represent potential enhancer elements while the red boxes represent potential silencer elements. The white box represents the basal promoter. The ‘*cis*-regulatory elements’ with question marks are those that are still speculative.

The absent 92 base pair sequence within the ~20kb sequence described above contains an overlapping SP-1/AP-2 binding site that potentially recruits general positive cofactors to aid in the assembly of the preinitiation complex. It is possible that positive cofactors recruited by *cis* 1 and/or 2 may compensate for the deleted 92 base pair sequence. Nevertheless, the presence

of this 92 base pair sequence may be essential to drive full endogenous expression of aggrecan within humans. It is possible that all the positive cofactors that are potentially recruited via the transcription factors that bind to putative binding sites within *cis* 1 and/or 2, *cis* 3 and the minimal promoter (including the deleted 92 base pair sequence) are necessary to confer full endogenous aggrecan expression and counteract the negative regulatory effects of repressors and corepressors. Figure 44 shows a schematic diagram of the hypothesised regulation of the human aggrecan gene based on data presented in this thesis.

The full endogenous pattern of expression in embryos/adults harbouring the ~20kb sequence including the deleted 92 base pair sequence may require the presence of the missing 92 base pair sequence and/or the additional presence of other potential *cis*-regulatory sequences such as insulator elements. The major difference in the embryonic and adult expression pattern driven by this sequence is either due to the inactivation of the promoter in the adult transgenic mouse line due to epigenetic modifications or possibly a difference in regulation of the aggrecan gene developmentally and post-natally. It is likely that the missing 92 base pairs are crucial for adult-specific regulation of the gene.

I also conclude that the conditional/inducible cartilage-specific transgenic mouse that was created using the cartilage specific promoter Col14xA1 from Han and Lefebvre (2008) fused to the *CreER*^{T2} gene has proved to be functional, tissue specific (Cre recombinase only present in chondrocytes) and not leaky (no Cre recombinase activity prior to induction with tamoxifen) at post-natal ages P7 and P14.

Finally, the ~20kb promoter discussed in the transgenic line that was created is activated in knee articular chondrocytes as well as tibial growth plate chondrocytes with the induction of OA in the knee through different models of injury (DMM and axial loading of the mouse tibias). This

activation seems to be proportionate to the extent of tissue damage in these knees. The response of the ColII4xA1 promoter in chondrocytes of mouse knees to OA induction through DMM, however, is somewhat different. This promoter is less responsive to injury and its activity does not seem to be proportionate to the extent of tissue damage and cartilage loss in mouse knees. It is possible that, unlike the ColII4xA1 promoter containing only *cis* 3 and the mouse aggrecan minimal promoter, the ~20kb promoter reacts more to OA due to it containing other *cis*-regulatory elements such as *cis* 1 and/or 2 that could possibly be more responsive to cellular signalling cascades brought on by injury.

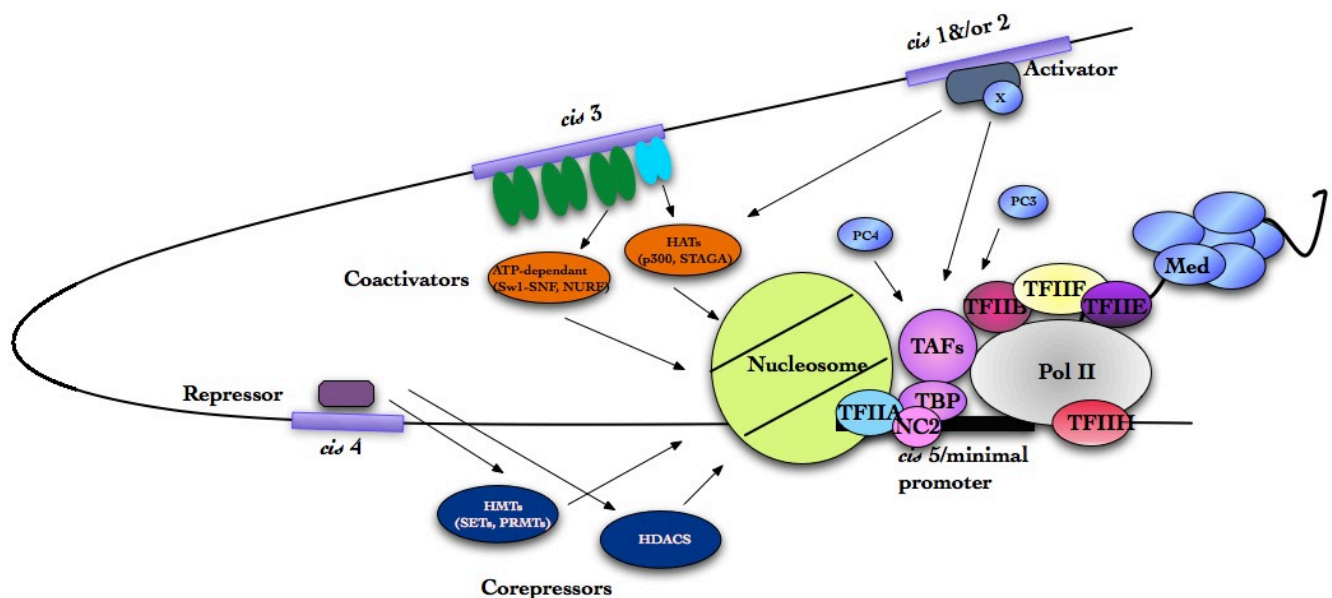


Figure 44: Schematic of hypothesised human aggrecan gene regulation. The ultimate purpose of DNA-binding activators is to mediate Pol II recruitment and function. In this case, Pol II along with the preinitiation complex is recruited and bound to *cis* 5/the minimal promoter. *Cis* 3 has 3 known binding sites for L-sox5/sox6 dimers (dark green) and one binding site for a sox9 dimer (turquoise). The effects of the L-sox6/sox5 and sox9 activators on Pol II are hypothetically indirect and may involve several interactions with diverse cofactors that modify chromatin structure and ease the recruitment of Pol II and general initiation factors (ATP-dependent factors, HATs shown in orange). *Cis* 4 may contain binding sites for repressor(s) that can recruit negative cofactors (HMTs and HDACs shown in indigo) that also modify chromatin structure to inhibit the initiation of transcription. *Cis* 1 and/or 2 may contain binding sites for activator(s) that can recruit cofactors that act at the level of DNA (shaded blue) such as cell specific factors (X). *Cis* 5/ minimal promoter may also recruit general positive cofactors such as PC3 and PC4 via positive activators. The absent 92 base pairs in the ~20kb sequence may have binding sites for such activators.

Future Work

Although the work presented here has demonstrated a role for the ~20kb sequence in human aggrecan gene regulation, the reporter gene expression driven by this promoter sequence did not exactly match the endogenous pattern of aggrecan gene expression embryonically and post-natally. Since a possibly crucial 92 base pair sequence present endogenously within the minimal promoter of the gene is absent in the investigated ~20kb sequence, it will be extremely important to investigate its activity in context of the entire 20kb sequence. To do this, transgenic embryos harbouring the ~20kb sequence including the 92 base pair sequence will be created as well as transgenic lines to monitor the expression driven by the complete sequence post-natally. Also, in order to establish whether the reduced post-natal expression driven by the ~20kb sequence lacking the 92 base pair sequence in transgenic line 12 is due to the absence of the 92 base pairs as opposed to a positional artefact caused by the location of the promoter within the genome, more transgenic lines containing this sequence will be created and compared to the transgenic lines containing the ~20kb sequence that includes the 92 base pair sequence. If the results of these experiments show a role for the 92 base pair sequence, it will be important to establish the *trans*-acting factors that bind to this sequence (potentially SP-1/AP-2) via EMSA studies (common affinity electrophoresis technique used to study protein-DNA interactions) and verify whether these transacting factors interact with any of the sox proteins that bind to the enhancer sequence *cis* 3. This could be done using ChIP analysis (method that gives a picture of the protein-DNA interactions that occur inside the nucleus of living cells).

Also, the hypothesized activities of *cis* 1,2 and 4 will be verified further by creating more combinations of these different potential *cis*-regulatory elements alongside *cis* 3 and the

minimal promoter and investigating their effect on overall reporter gene activity in transgenic embryos.

Furthermore, in order to investigate the efficiency of line 39 containing Col114xA1Cre in a quantitative manner, the line will be crossed to *Rosa26* β -galactosidase reporter mice as opposed to the global fluorescence reporter mouse used. The visualisation of the membrane targetted mT and mG polypeptides in the cartilage of knees requires them to be frozen and eventually cut without decalcification. This can lead to the fragmentation of sections. However, the visualisation of β -galactosidase in knees does not require them to be frozen and they can be decalcified, paraffin embedded and cut into high quality knee sections. The Cre recombinase will eventually be activated in the progeny by injection with tamoxifen at 1,3 and 6 months of age. The activity of the promoter at these different ages will be monitored as well as the efficiency of recombination (can be measured by counting X-gal positive stained cells).

If the results of the above experiments show significant chondrocyte-specific conversion into X-gal positive stained cells in the cartilages of mice at all the ages inspected with no leakiness, line 39 will be crossed with a number of different floxed transgenic mice. It can be mated to mice in which a particular gene of interest is floxed (loss of function) or in which a stop sequence inhibiting the transcription of a gene is floxed (gain of function). For example, it can be mated to a transgenic mouse in which Timp3 or FGF-2 are floxed. DMM will eventually be performed on the knees of the F1 progeny. Tamoxifen can be given to these mice directly before or even after the onset of OA.

References

- Adelman, K., Wei, W., Ardehali, M. B., Werner, J., Zhu, B., Reinberg, D., and Lis, J. T., 2006, *Drosophila* Paf1 modulates chromatin structure at actively transcribed genes, *Mol Cell Biol* **26**(1):250-60.
- Agrawal, A., Guttapalli, A., Narayan, S., Albert, T. J., Shapiro, I. M., and Risbud, M. V., 2007, Normoxic stabilization of HIF-1 α drives glycolytic metabolism and regulates aggrecan gene expression in nucleus pulposus cells of the rat intervertebral disk, *Am J Physiol Cell Physiol* **293**(2):C621-31.
- Aigner, T., Soder, S., Gebhard, P. M., McAlinden, A., and Haag, J., 2007, Mechanisms of disease: role of chondrocytes in the pathogenesis of osteoarthritis--structure, chaos and senescence, *Nat Clin Pract Rheumatol* **3**(7):391-9.
- Akizuki, S., Mow, V. C., Muller, F., Pita, J. C., and Howell, D. S., 1987, Tensile properties of human knee joint cartilage. II. Correlations between weight bearing and tissue pathology and the kinetics of swelling, *J Orthop Res* **5**(2):173-86.
- Alley, E. W., Murphy, W. J., and Russell, S. W., 1995, A classical enhancer element responsive to both lipopolysaccharide and interferon-gamma augments induction of the iNOS gene in mouse macrophages, *Gene* **158**(2):247-51.
- Alvarez, J., Costales, L., Serra, R., Balbin, M., and Lopez, J. M., 2005, Expression patterns of matrix metalloproteinases and vascular endothelial growth factor during epiphyseal ossification, *J Bone Miner Res* **20**(6):1011-21.
- Arciniegas, E., Neves, C. Y., Candelle, D., and Parada, D., 2004, Differential versican isoforms and aggrecan expression in the chicken embryo aorta, *Anat Rec A Discov Mol Cell Evol Biol* **279**(1):592-600.
- Arikawa-Hirasawa, E., Watanabe, H., Takami, H., Hassell, J. R., and Yamada, Y., 1999, Perlecan is essential for cartilage and cephalic development, *Nat Genet* **23**(3):354-8.
- Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., and Teichmann, S. A., 2004, Structure and evolution of transcriptional regulatory networks, *Curr Opin Struct Biol* **14**(3):283-91.
- Baldwin, C. T., Reginato, A. M., and Prockop, D. J., 1989, A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain, *J Biol Chem* **264**(27):15747-50.
- Barry, F. P., Neame, P. J., Sasse, J., and Pearson, D., 1994, Length variation in the keratan sulfate domain of mammalian aggrecan, *Matrix Biol* **14**(4):323-8.

- Barry, F. P., Rosenberg, L. C., Gaw, J. U., Koob, T. J., and Neame, P. J., 1995, N- and O-linked keratan sulfate on the hyaluronan binding region of aggrecan from mature and immature bovine cartilage, *J Biol Chem* **270**(35):20516-24.
- Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., and Reinberg, D., 2003, FACT facilitates transcription-dependent nucleosome alteration, *Science* **301**(5636):1090-3.
- Bender, M. A., Roach, J. N., Halow, J., Close, J., Alami, R., Bouhassira, E. E., Groudine, M., and Fiering, S. N., 2001, Targeted deletion of 5'HS1 and 5'HS4 of the beta-globin locus control region reveals additive activity of the DNaseI hypersensitive sites, *Blood* **98**(7):2022-7.
- Berger, S. L., 2002, Histone modifications in transcriptional regulation, *Curr Opin Genet Dev* **12**(2):142-8.
- Bhosale, A. M., and Richardson, J. B., 2008, Articular cartilage: structure, injuries and review of management, *Br Med Bull* **87**:77-95.
- Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R., and de Crombrughe, B., 1999, Sox9 is required for cartilage formation, *Nat Genet* **22**(1):85-9.
- Bi, W., Huang, W., Whitworth, D. J., Deng, J. M., Zhang, Z., Behringer, R. R., and de Crombrughe, B., 2001, Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization, *Proc Natl Acad Sci U S A* **98**(12):6698-703.
- Billinghurst, R. C., Dahlberg, L., Ionescu, M., Reiner, A., Bourne, R., Rorabeck, C., Mitchell, P., Hambor, J., Diekmann, O., Tschesche, H., Chen, J., Van Wart, H., and Poole, A. R., 1997, Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage, *J Clin Invest* **99**(7):1534-45.
- Bird, A., 2002, DNA methylation patterns and epigenetic memory, *Genes Dev* **16**(1):6-21.
- Bird, A. P., 1986, CpG-rich islands and the function of DNA methylation, *Nature* **321**(6067):209-13.
- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetric, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E.

- A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., et al., 2007, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature* **447**(7146):799-816.
- Blumer, M. J., Longato, S., Richter, E., Perez, M. T., Konakci, K. Z., and Fritsch, H., 2005, The role of cartilage canals in endochondral and perichondral bone formation: are there similarities between these two processes?, *J Anat* **206**(4):359-72.
- Bortvin, A., and Winston, F., 1996, Evidence that Spt6p controls chromatin structure by a direct interaction with histones, *Science* **272**(5267):1473-6.
- Brew, C. J., Clegg, P. D., Boot-Handford, R. P., Andrew, J. G., and Hardingham, T., Gene expression in human chondrocytes in late osteoarthritis is changed in both fibrillated and intact cartilage without evidence of generalised chondrocyte hypertrophy, *Ann Rheum Dis* **69**(1):234-40.
- Brown, G. M., Huckerby, T. N., Bayliss, M. T., and Nieduszyński, I. A., 1998, Human aggrecan keratan sulfate undergoes structural changes during adolescent development, *J Biol Chem* **273**(41):26408-14.
- Buckwalter, J. A., and Mankin, H. J., 1998, Articular cartilage: tissue design and chondrocyte-matrix interactions, *Instr Course Lect* **47**:477-86.
- Buckwalter, J. A., Pita, J. C., Muller, F. J., and Nessler, J., 1994, Structural differences between two populations of articular cartilage proteoglycan aggregates, *J Orthop Res* **12**(1):144-8.
- Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A., 1989, Five intermediate complexes in transcription initiation by RNA polymerase II, *Cell* **56**(4):549-61.
- Burke, T. W., and Kadonaga, J. T., 1997, The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila, *Genes Dev* **11**(22):3020-31.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., Forrest, A. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A., and Hayashizaki, Y., 2006, Genome-wide analysis of mammalian promoter architecture and evolution, *Nat Genet* **38**(6):626-35.
- Chakalova, L., Carter, D., Debrand, E., Goyenechea, B., Horton, A., Miles, J., Osborne, C., and Fraser, P., 2005, Developmental regulation of the beta-globin gene locus, *Prog Mol Subcell Biol* **38**:183-206.

- Chamberland, A., Wang, E., Jones, A. R., Collins-Racie, L. A., LaVallie, E. R., Huang, Y., Liu, L., Morris, E. A., Flannery, C. R., and Yang, Z., 2009, Identification of a novel HtrA1-susceptible cleavage site in human aggrecan: evidence for the involvement of HtrA1 in aggrecan proteolysis in vivo, *J Biol Chem* **284**(40):27352-9.
- Chambers, M. G., Kuffner, T., Cowan, S. K., Cheah, K. S., and Mason, R. M., 2002, Expression of collagen and aggrecan genes in normal and osteoarthritic murine knee joints, *Osteoarthritis Cartilage* **10**(1):51-61.
- Chan, E. T., Quon, G. T., Chua, G., Babak, T., Trocheset, M., Zirngibl, R. A., Aubin, J., Ratcliffe, M. J., Wilde, A., Brudno, M., Morris, Q. D., and Hughes, T. R., 2009, Conservation of core gene expression in vertebrate tissues, *J Biol* **8**(3):33.
- Chapman, R. D., Heidemann, M., Albert, T. K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D., 2007, Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7, *Science* **318**(5857):1780-2.
- Chen, B. S., and Hampsey, M., 2002, Transcription activation: unveiling the essential nature of TFIID, *Curr Biol* **12**(18):R620-2.
- Clapier, C. R., and Cairns, B. R., 2009, The biology of chromatin remodeling complexes, *Annu Rev Biochem* **78**:273-304.
- Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L., and Myers, R. M., 2006, Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome, *Genome Res* **16**(1):1-10.
- Corden, J. L., Cadena, D. L., Ahearn, J. M., Jr., and Dahmus, M. E., 1985, A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II, *Proc Natl Acad Sci U S A* **82**(23):7934-8.
- Cramer, P., 2004, RNA polymerase II structure: from core to functional complexes, *Curr Opin Genet Dev* **14**(2):218-26.
- Cramer, P., Bushnell, D. A., and Kornberg, R. D., 2001, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution, *Science* **292**(5523):1863-76.
- Cserjesi, P., Brown, D., Ligon, K. L., Lyons, G. E., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., and Olson, E. N., 1995, Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis, *Development* **121**(4):1099-110.
- De Souza, R. L., Matsuura, M., Eckstein, F., Rawlinson, S. C., Lanyon, L. E., and Pitsillides, A. A., 2005, Non-invasive axial loading of mouse tibiae increases cortical bone formation and modifies trabecular organization: a new model to study cortical and cancellous compartments in a single loaded element, *Bone* **37**(6):810-8.
- de Wet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R., and Subramani, S., 1987, Firefly luciferase gene: structure and expression in mammalian cells, *Mol Cell Biol* **7**(2):725-37.

- Deng, W., and Roberts, S. G., 2005, A core promoter element downstream of the TATA box that is recognized by TFIIB, *Genes Dev* **19**(20):2418-23.
- Doege, K., Hall, L. B., McKinnon, W., Chen, L., Stephens, D. T., and Garrison, K., 2002, A remote upstream element regulates tissue-specific expression of the rat aggrecan gene, *J Biol Chem* **277**(16):13989-97.
- Doege, K., Sasaki, M., and Yamada, Y., 1990, Rat and human cartilage proteoglycan (aggrecan) gene structure, *Biochem Soc Trans* **18**(2):200-2.
- Doege, K. J., Coulter, S. N., Meek, L. M., Maslen, K., and Wood, J. G., 1997, A human-specific polymorphism in the coding region of the aggrecan gene. Variable number of tandem repeats produce a range of core protein sizes in the general population, *J Biol Chem* **272**(21):13974-9.
- Doege, K. J., Garrison, K., Coulter, S. N., and Yamada, Y., 1994, The structure of the rat aggrecan gene and preliminary characterization of its promoter, *J Biol Chem* **269**(46):29232-40.
- Doege, K. J., Sasaki, M., Kimura, T., and Yamada, Y., 1991, Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms, *J Biol Chem* **266**(2):894-902.
- Domowicz, M. S., Sanders, T. A., Ragsdale, C. W., and Schwartz, N. B., 2008, Aggrecan is expressed by embryonic brain glia and regulates astrocyte development, *Dev Biol* **315**(1):114-24.
- Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L., and Karsenty, G., 1997, Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation, *Cell* **89**(5):747-54.
- Dudhia, J., 2005, Aggrecan, aging and assembly in articular cartilage, *Cell Mol Life Sci* **62**(19-20):2241-56.
- Dvir, A., 2002, Promoter escape by RNA polymerase II, *Biochim Biophys Acta* **1577**(2):208-223.
- Dynan, W. S., and Tjian, R., 1983, Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II, *Cell* **32**(3):669-80.
- Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P., 1996, Ligand-activated site-specific recombination in mice, *Proc Natl Acad Sci U S A* **93**(20):10887-90.
- Felsenfeld, G., and Groudine, M., 2003, Controlling the double helix, *Nature* **421**(6921):448-53.

- Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., and McCallion, A. S., 2006, Conservation of RET regulatory function from human to zebrafish without sequence similarity, *Science* **312**(5771):276-9.
- Flanagan, P. M., Kelleher, R. J., 3rd, Sayre, M. H., Tschochner, H., and Kornberg, R. D., 1991, A mediator required for activation of RNA polymerase II transcription in vitro, *Nature* **350**(6317):436-8.
- Fondell, J. D., Ge, H., and Roeder, R. G., 1996, Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex, *Proc Natl Acad Sci U S A* **93**(16):8329-33.
- Forriol, F., and Shapiro, F., 2005, Bone development: interaction of molecular components and biophysical forces, *Clin Orthop Relat Res* (432):14-33.
- Fosang, A. J., and Hardingham, T. E., 1989, Isolation of the N-terminal globular protein domains from cartilage proteoglycans. Identification of G2 domain and its lack of interaction with hyaluronate and link protein, *Biochem J* **261**(3):801-9.
- Fuda, N. J., Ardehali, M. B., and Lis, J. T., 2009, Defining mechanisms that regulate RNA polymerase II transcription in vivo, *Nature* **461**(7261):186-92.
- Fukada, K., Shibata, S., Suzuki, S., Ohya, K., and Kuroda, T., 1999, In situ hybridisation study of type I, II, X collagens and aggrecan mRNAs in the developing condylar cartilage of fetal mouse mandible, *J Anat* **195** (Pt 3):321-9.
- Fulop, C., Walcz, E., Valyon, M., and Glant, T. T., 1993, Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module, *J Biol Chem* **268**(23):17377-83.
- Funderburgh, J. L., 2000, Keratan sulfate: structure, biosynthesis, and function, *Glycobiology* **10**(10):951-8.
- Georges, A. B., Benayoun, B. A., Caburet, S., and Veitia, R. A., Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from?, *FASEB J* **24**(2):346-56.
- Gerber, H. P., Vu, T. H., Ryan, A. M., Kowalski, J., Werb, Z., and Ferrara, N., 1999, VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation, *Nat Med* **5**(6):623-8.
- Glasson, S. S., Askew, R., Sheppard, B., Carito, B., Blanchet, T., Ma, H. L., Flannery, C. R., Peluso, D., Kanki, K., Yang, Z., Majumdar, M. K., and Morris, E. A., 2005, Deletion of active ADAMTS5 prevents cartilage degradation in a murine model of osteoarthritis, *Nature* **434**(7033):644-8.

- Glasson, S. S., Blanchet, T. J., and Morris, E. A., 2007, The surgical destabilization of the medial meniscus (DMM) model of osteoarthritis in the 129/SvEv mouse, *Osteoarthritis Cartilage* **15**(9):1061-9.
- Gleghorn, L., Ramesar, R., Beighton, P., and Wallis, G., 2005, A mutation in the variable repeat region of the aggrecan gene (AGC1) causes a form of spondyloepiphyseal dysplasia associated with severe, premature osteoarthritis, *Am J Hum Genet* **77**(3):484-90.
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D., 2001, Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution, *Science* **292**(5523):1876-82.
- Goldman, S. R., Ebright, R. H., and Nickels, B. E., 2009, Direct detection of abortive RNA transcripts in vivo, *Science* **324**(5929):927-8.
- Goldring, M. B., and Goldring, S. R., 2007, Osteoarthritis, *J Cell Physiol* **213**(3):626-34.
- Gondor, A., and Ohlsson, R., 2009, Chromosome crosstalk in three dimensions, *Nature* **461**(7261):212-7.
- Guilak, F., Alexopoulos, L. G., Upton, M. L., Youn, I., Choi, J. B., Cao, L., Setton, L. A., and Haider, M. A., 2006, The pericellular matrix as a transducer of biomechanical and biochemical signals in articular cartilage, *Ann N Y Acad Sci* **1068**:498-512.
- Hall, B. K., and Miyake, T., 1992, The membranous skeleton: the role of cell condensations in vertebrate skeletogenesis, *Anat Embryol (Berl)* **186**(2):107-24.
- Hall, B. K., and Miyake, T., 1995, Divide, accumulate, differentiate: cell condensation in skeletal development revisited, *Int J Dev Biol* **39**(6):881-93.
- Hall, B. K., and Miyake, T., 2000, All for one and one for all: condensations and the initiation of skeletal development, *Bioessays* **22**(2):138-47.
- Hampsey, M., 1998, Molecular genetics of the RNA polymerase II general transcriptional machinery, *Microbiol Mol Biol Rev* **62**(2):465-503.
- Han, M., and Grunstein, M., 1988, Nucleosome loss activates yeast downstream promoters in vivo, *Cell* **55**(6):1137-45.
- Han, Y., and Lefebvre, V., 2008, L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer, *Mol Cell Biol* **28**(16):4999-5013.
- Hardingham, T. E., and Fosang, A. J., 1992, Proteoglycans: many forms and many functions, *FASEB J* **6**(3):861-70.

- Hatori, M., Klatte, K. J., Teixeira, C. C., and Shapiro, I. M., 1995, End labeling studies of fragmented DNA in the avian growth plate: evidence of apoptosis in terminally differentiated chondrocytes, *J Bone Miner Res* **10**(12):1960-8.
- Heinegard, D., 2009, Proteoglycans and more--from molecules to biology, *Int J Exp Pathol* **90**(6):575-86.
- Heintzman, N. D., and Ren, B., 2009, Finding distal regulatory elements in the human genome, *Curr Opin Genet Dev* **19**(6):541-9.
- Henry, S. P., Jang, C. W., Deng, J. M., Zhang, Z., Behringer, R. R., and de Crombrughe, B., 2009, Generation of aggrecan-CreERT2 knockin mice for inducible Cre activity in adult cartilage, *Genesis* **47**(12):805-14.
- Ho, C. K., and Shuman, S., 1999, Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme, *Mol Cell* **3**(3):405-11.
- Holmes, M. W., Bayliss, M. T., and Muir, H., 1988, Hyaluronic acid in human articular cartilage. Age-related changes in content and size, *Biochem J* **250**(2):435-41.
- Homandberg, G. A., 1999, Potential regulation of cartilage metabolism in osteoarthritis by fibronectin fragments, *Front Biosci* **4**:D713-30.
- Horkay, F., Basser, P. J., Hecht, A. M., and Geissler, E., 2008, Insensitivity to salt of assembly of a rigid biopolymer aggrecan, *Phys Rev Lett* **101**(6):068301.
- Horton, W. E., Jr., Lethbridge-Cejku, M., Hochberg, M. C., Balakir, R., Precht, P., Plato, C. C., Tobin, J. D., Meek, L., and Doege, K., 1998, An association between an aggrecan polymorphic allele and bilateral hand osteoarthritis in elderly white men: data from the Baltimore Longitudinal Study of Aging (BLSA), *Osteoarthritis Cartilage* **6**(4):245-51.
- Hu, J. C., and Athanasiou, K. A., 2006, Chondrocytes from different zones exhibit characteristic differences in high density culture, *Connect Tissue Res* **47**(3):133-40.
- Huertas, P., and Aguilera, A., 2003, Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination, *Mol Cell* **12**(3):711-21.
- Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P., and Metzger, D., 1999, Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases, *Nucleic Acids Res* **27**(22):4324-7.
- Jacobsen, K. D., and Willumsen, B. M., 1995, Kinetics of expression of inducible beta-galactosidase in murine fibroblasts: high initial rate compared to steady-state expression, *J Mol Biol* **252**(3):289-95.

- James, A. W., Xu, Y., Lee, J. K., Wang, R., and Longaker, M. T., 2009, Differential effects of TGF-beta1 and TGF-beta3 on chondrogenesis in posterofrontal cranial suture-derived mesenchymal cells in vitro, *Plast Reconstr Surg* **123**(1):31-43.
- Just, W., Klett, C., Vetter, U., and Vogel, W., 1993, Assignment of the human aggrecan gene AGC1 to 15q25-->q26.2 by in situ hybridization, *Hum Genet* **92**(5):516-8.
- Juven-Gershon, T., Cheng, S., and Kadonaga, J. T., 2006a, Rational design of a super core promoter that enhances gene expression, *Nat Methods* **3**(11):917-22.
- Juven-Gershon, T., Hsu, J. Y., and Kadonaga, J. T., 2006b, Perspectives on the RNA polymerase II core promoter, *Biochem Soc Trans* **34**(Pt 6):1047-50.
- Juven-Gershon, T., and Kadonaga, J. T., Regulation of gene expression via the core promoter and the basal transcriptional machinery, *Dev Biol* **339**(2):225-9.
- Kaiser, K., and Meisterernst, M., 1996, The human general co-factors, *Trends Biochem Sci* **21**(9):342-5.
- Karsenty, G., 2008, Transcriptional control of skeletogenesis, *Annu Rev Genomics Hum Genet* **9**:183-96.
- Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B., 2005, A high-resolution map of active promoters in the human genome, *Nature* **436**(7052):876-80.
- Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H., and Kornberg, R. D., 1994, A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II, *Cell* **77**(4):599-608.
- Kimata, K., Barrach, H. J., Brown, K. S., and Pennypacker, J. P., 1981, Absence of proteoglycan core protein in cartilage from the cmd/cmd (cartilage matrix deficiency) mouse, *J Biol Chem* **256**(13):6961-8.
- Kingston, R. E., and Narlikar, G. J., 1999, ATP-dependent remodeling and acetylation as regulators of chromatin fluidity, *Genes Dev* **13**(18):2339-52.
- Kist, R., Schrewe, H., Balling, R., and Scherer, G., 2002, Conditional inactivation of Sox9: a mouse model for campomelic dysplasia, *Genesis* **32**(2):121-3.
- Klose, R. J., and Bird, A. P., 2006, Genomic DNA methylation: the mark and its mediators, *Trends Biochem Sci* **31**(2):89-97.
- Knudson, W., Casey, B., Nishida, Y., Eger, W., Kuettner, K. E., and Knudson, C. B., 2000, Hyaluronan oligosaccharides perturb cartilage matrix homeostasis and induce chondrocytic chondrolysis, *Arthritis Rheum* **43**(5):1165-74.

- Komarnitsky, P., Cho, E. J., and Buratowski, S., 2000, Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription, *Genes Dev* **14**(19):2452-60.
- Korenberg, J. R., Chen, X. N., Doege, K., Grover, J., and Roughley, P. J., 1993, Assignment of the human aggrecan gene (AGC1) to 15q26 using fluorescence in situ hybridization analysis, *Genomics* **16**(2):546-8.
- Krasnokutsky, S., Samuels, J., and Abramson, S. B., 2007, Osteoarthritis in 2007, *Bull NYU Hosp Jt Dis* **65**(3):222-8.
- Krebsbach, P. H., Nakata, K., Bernier, S. M., Hatano, O., Miyashita, T., Rhodes, C. S., and Yamada, Y., 1996, Identification of a minimum enhancer sequence for the type II collagen gene reveals several core sequence motifs in common with the link protein gene, *J Biol Chem* **271**(8):4298-303.
- Krull, N. B., and Gressner, A. M., 1992, Differential expression of keratan sulphate proteoglycans fibromodulin, lumican and aggrecan in normal and fibrotic rat liver, *FEBS Lett* **312**(1):47-52.
- Kuras, L., Kosa, P., Mencia, M., and Struhl, K., 2000, TAF-Containing and TAF-independent forms of transcriptionally active TBP in vivo, *Science* **288**(5469):1244-8.
- Kuras, L., and Struhl, K., 1999, Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme, *Nature* **399**(6736):609-13.
- Kutach, A. K., and Kadonaga, J. T., 2000, The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters, *Mol Cell Biol* **20**(13):4754-64.
- Lachner, M., and Jenuwein, T., 2002, The many faces of histone lysine methylation, *Curr Opin Cell Biol* **14**(3):286-98.
- Lakso, M., Sauer, B., Mosinger, B., Jr., Lee, E. J., Manning, R. W., Yu, S. H., Mulder, K. L., and Westphal, H., 1992, Targeted oncogene activation by site-specific recombination in transgenic mice, *Proc Natl Acad Sci U S A* **89**(14):6232-6.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chisoe, S. L., Wendl, M. C., Delehaunty, K.

- D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al., 2001, Initial sequencing and analysis of the human genome, *Nature* **409**(6822):860-921.
- Le Maitre, C. L., Pockert, A., Buttle, D. J., Freemont, A. J., and Hoyland, J. A., 2007, Matrix synthesis and degradation in human intervertebral disc degeneration, *Biochem Soc Trans* **35**(Pt 4):652-5.
- Lee, T. I., and Young, R. A., 2000, Transcription of eukaryotic protein-coding genes, *Annu Rev Genet* **34**:77-137.
- Leone, D. P., Genoud, S., Atanasoski, S., Grausenburger, R., Berger, P., Metzger, D., Macklin, W. B., Chambon, P., and Suter, U., 2003, Tamoxifen-inducible glia-specific Cre mice for somatic mutagenesis in oligodendrocytes and Schwann cells, *Mol Cell Neurosci* **22**(4):430-40.
- Leuther, K. K., Bushnell, D. A., and Kornberg, R. D., 1996, Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation, *Cell* **85**(5):773-9.
- Li, H., and Schwartz, N. B., 1995, Gene structure of chick cartilage chondroitin sulfate proteoglycan (aggrecan) core protein, *J Mol Evol* **41**(6):878-85.
- Li, Q., Peterson, K. R., Fang, X., and Stamatoyannopoulos, G., 2002, Locus control regions, *Blood* **100**(9):3077-86.
- Li, Q., Zhang, M., Duan, Z., and Stamatoyannopoulos, G., 1999a, Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region, *Genomics* **61**(2):183-93.
- Li, X. Y., Virbasius, A., Zhu, X., and Green, M. R., 1999b, Enhancement of TBP binding by activators and general transcription factors, *Nature* **399**(6736):605-9.
- Lim, C. Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J. T., 2004, The MTE, a new core promoter element for transcription by RNA polymerase II, *Genes Dev* **18**(13):1606-17.
- Lis, J., 1998, Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation, *Cold Spring Harb Symp Quant Biol* **63**:347-56.
- Liu, P., Jenkins, N. A., and Copeland, N. G., 2003, A highly efficient recombineering-based method for generating conditional knockout mutations, *Genome Res* **13**(3):476-84.
- Loeser, R. F., 2008, Molecular mechanisms of cartilage destruction in osteoarthritis, *J Musculoskelet Neuronal Interact* **8**(4):303-6.

- Loeser, R. F., and Shakoar, N., 2003, Aging or osteoarthritis: which is the problem?, *Rheum Dis Clin North Am* **29**(4):653-73.
- Long, F., Zhang, X. M., Karp, S., Yang, Y., and McMahon, A. P., 2001, Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation, *Development* **128**(24):5099-108.
- Lorch, Y., LaPointe, J. W., and Kornberg, R. D., 1987, Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones, *Cell* **49**(2):203-10.
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J., 1997, Crystal structure of the nucleosome core particle at 2.8 Å resolution, *Nature* **389**(6648):251-60.
- Luo, W., Kuwada, T. S., Chandrasekaran, L., Zheng, J., and Tanzer, M. L., 1996, Divergent secretory behavior of the opposite ends of aggrecan, *J Biol Chem* **271**(28):16447-50.
- Malik, S., and Roeder, R. G., 2000, Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells, *Trends Biochem Sci* **25**(6):277-83.
- Manicourt, D. H., Pita, J. C., Thonar, E. J., and Howell, D. S., 1991, Proteoglycans nondissociatively extracted from different zones of canine normal articular cartilage: variations in the sedimentation profile of aggregates with degree of physiological stress, *Connect Tissue Res* **26**(4):231-46.
- Maroudas, A., Bayliss, M. T., Uchitel-Kaushansky, N., Schneiderman, R., and Gilav, E., 1998, Aggrecan turnover in human articular cartilage: use of aspartic acid racemization as a marker of molecular age, *Arch Biochem Biophys* **350**(1):61-71.
- Martel-Pelletier, J., Boileau, C., Pelletier, J. P., and Roughley, P. J., 2008, Cartilage in normal and osteoarthritis conditions, *Best Pract Res Clin Rheumatol* **22**(2):351-84.
- Maston, G. A., Evans, S. K., and Green, M. R., 2006, Transcriptional regulatory elements in the human genome, *Annu Rev Genomics Hum Genet* **7**:29-59.
- Matthews, R. T., Kelly, G. M., Zerillo, C. A., Gray, G., Tiemeyer, M., and Hockfield, S., 2002, Aggrecan glycoforms contribute to the molecular heterogeneity of perineuronal nets, *J Neurosci* **22**(17):7536-47.
- Matyas, J. R., Adams, M. E., Huang, D., and Sandell, L. J., 1995, Discoordinate gene expression of aggrecan and type II collagen in experimental osteoarthritis, *Arthritis Rheum* **38**(3):420-5.
- Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P., 2005, A structural perspective of CTD function, *Genes Dev* **19**(12):1401-15.
- Meisterernst, M., Roy, A. L., Lieu, H. M., and Roeder, R. G., 1991, Activation of class II gene transcription by regulatory factors is potentiated by a novel activity, *Cell* **66**(5):981-93.

- Melrose, J., Smith, S., Cake, M., Read, R., and Whitelock, J., 2005, Comparative spatial and temporal localisation of perlecan, aggrecan and type I, II and IV collagen in the ovine meniscus: an ageing study, *Histochem Cell Biol* **124**(3-4):225-35.
- Menendez, D., Inga, A., and Resnick, M. A., 2009, The expanding universe of p53 targets, *Nat Rev Cancer* **9**(10):724-37.
- Merika, M., and Thanos, D., 2001, Enhanceosomes, *Curr Opin Genet Dev* **11**(2):205-8.
- Middleton, J., Manthey, A., and Tyler, J., 1996, Insulin-like growth factor (IGF) receptor, IGF-I, interleukin-1 beta (IL-1 beta), and IL-6 mRNA expression in osteoarthritic and normal human cartilage, *J Histochem Cytochem* **44**(2):133-41.
- Middleton, J. F., and Tyler, J. A., 1992, Upregulation of insulin-like growth factor I gene expression in the lesions of osteoarthritic human articular cartilage, *Ann Rheum Dis* **51**(4):440-7.
- Miyaki, S., Nakasa, T., Otsuki, S., Grogan, S. P., Higashiyama, R., Inoue, A., Kato, Y., Sato, T., Lotz, M. K., and Asahara, H., 2009, MicroRNA-140 is expressed in differentiated human articular chondrocytes and modulates interleukin-1 responses, *Arthritis Rheum* **60**(9):2723-30.
- Morgelin, M., Paulsson, M., Hardingham, T. E., Heinegard, D., and Engel, J., 1988, Cartilage proteoglycans. Assembly with hyaluronate and link protein as studied by electron microscopy, *Biochem J* **253**(1):175-85.
- Murphy, C. L., Thoms, B. L., Vaghjiani, R. J., and Lafont, J. E., 2009, Hypoxia. HIF-mediated articular chondrocyte function: prospects for cartilage repair, *Arthritis Res Ther* **11**(1):213.
- Murphy, J. M., Heinegard, R., McIntosh, A., Sterchi, D., and Barry, F. P., 1999, Distribution of cartilage molecules in the developing mouse joint, *Matrix Biol* **18**(5):487-97.
- Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L., and Luo, L., 2007, A global double-fluorescent Cre reporter mouse, *Genesis* **45**(9):593-605.
- Myer, V. E., and Young, R. A., 1998, RNA polymerase II holoenzymes and subcomplexes, *J Biol Chem* **273**(43):27757-60.
- Narlikar, L., and Hartemink, A. J., 2006, Sequence features of DNA binding sites reveal structural class of associated transcription factor, *Bioinformatics* **22**(2):157-63.
- Naumann, A., Dennis, J. E., Awadallah, A., Carrino, D. A., Mansour, J. M., Kastenbauer, E., and Caplan, A. I., 2002, Immunochemical and mechanical characterization of cartilage subtypes in rabbit, *J Histochem Cytochem* **50**(8):1049-58.
- Neame, P. J., and Barry, F. P., 1994, The link proteins, *EXS* **70**:53-72.

- Neidlinger-Wilke, C., Liedert, A., Wuerztz, K., Buser, Z., Rinkler, C., Kafer, W., Ignatius, A., Claes, L., Roberts, S., and Johnson, W. E., 2009, Mechanical stimulation alters pleiotrophin and aggrecan expression by human intervertebral disc cells and influences their capacity to stimulate endothelial migration, *Spine (Phila Pa 1976)* **34**(7):663-9.
- Nilsson, B., De Luca, S., Lohmander, S., and Hascall, V. C., 1982, Structures of N-linked and O-linked oligosaccharides on proteoglycan monomer isolated from the Swarm rat chondrosarcoma, *J Biol Chem* **257**(18):10920-7.
- Niwa, H., Yamamura, K., and Miyazaki, J., 1991, Efficient selection for high-expression transfectants with a novel eukaryotic vector, *Gene* **108**(2):193-9.
- Nobrega, M. A., Ovcharenko, I., Afzal, V., and Rubin, E. M., 2003, Scanning human gene deserts for long-range enhancers, *Science* **302**(5644):413.
- Oberlender, S. A., and Tuan, R. S., 1994, Expression and functional involvement of N-cadherin in embryonic limb chondrogenesis, *Development* **120**(1):177-87.
- Ogbourne, S., and Antalis, T. M., 1998, Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes, *Biochem J* **331** (Pt 1):1-14.
- Oh, C. D., Maity, S. N., Lu, J. F., Zhang, J., Liang, S., Coustry, F., de Crombrughe, B., and Yasuda, H., Identification of SOX9 interaction sites in the genome of chondrocytes, *PLoS One* **5**(4):e10113.
- Olsen, B. R., Reginato, A. M., and Wang, W., 2000, Bone development, *Annu Rev Cell Dev Biol* **16**:191-220.
- Orban, P. C., Chui, D., and Marth, J. D., 1992, Tissue- and site-specific DNA recombination in transgenic mice, *Proc Natl Acad Sci U S A* **89**(15):6861-5.
- Pacifici, M., Koyama, E., and Iwamoto, M., 2005, Mechanisms of synovial joint and articular cartilage formation: recent advances, but many lingering mysteries, *Birth Defects Res C Embryo Today* **75**(3):237-48.
- Palmer, G. D., Chao Ph, P. H., Raia, F., Mauck, R. L., Valhmu, W. B., and Hung, C. T., 2001, Time-dependent aggrecan gene expression of articular chondrocytes in response to hyperosmotic loading, *Osteoarthritis Cartilage* **9**(8):761-70.
- Parker, C. S., and Topol, J., 1984, A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene, *Cell* **37**(1):273-83.
- Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J. A., and Yamamoto, K. R., 1981, Purified glucocorticoid receptors bind selectively in vitro to a cloned DNA fragment whose transcription is regulated by glucocorticoids in vivo, *Proc Natl Acad Sci U S A* **78**(11):6628-32.

- Pechak, D. G., Kujawa, M. J., and Caplan, A. I., 1986, Morphological and histochemical events during first bone formation in embryonic chick limbs, *Bone* **7**(6):441-58.
- Pedram, M., Sprung, C. N., Gao, Q., Lo, A. W., Reynolds, G. E., and Murnane, J. P., 2006, Telomere position effect and silencing of transgenes near telomeres in the mouse, *Mol Cell Biol* **26**(5):1865-78.
- Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., and Rubin, E. M., 2006, In vivo enhancer analysis of human conserved non-coding sequences, *Nature* **444**(7118):499-502.
- Peterlin, B. M., and Price, D. H., 2006, Controlling the elongation phase of transcription with P-TEFb, *Mol Cell* **23**(3):297-305.
- Phatnani, H. P., and Greenleaf, A. L., 2006, Phosphorylation and functions of the RNA polymerase II CTD, *Genes Dev* **20**(21):2922-36.
- Pirok, E. W., 3rd, Li, H., Mensch, J. R., Jr., Henry, J., and Schwartz, N. B., 1997, Structural and functional analysis of the chick chondroitin sulfate proteoglycan (aggrecan) promoter and enhancer region, *J Biol Chem* **272**(17):11566-74.
- Pitsillides, A. A., and Ashhurst, D. E., 2008, A critical evaluation of specific aspects of joint development, *Dev Dyn* **237**(9):2284-94.
- Place, R. F., Li, L. C., Pookot, D., Noonan, E. J., and Dahiya, R., 2008, MicroRNA-373 induces expression of genes with complementary promoter sequences, *Proc Natl Acad Sci U S A* **105**(5):1608-13.
- Pokholok, D. K., Hannett, N. M., and Young, R. A., 2002, Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo, *Mol Cell* **9**(4):799-809.
- Poole, A. R., Kojima, T., Yasuda, T., Mwale, F., Kobayashi, M., and Lavery, S., 2001, Composition and structure of articular cartilage: a template for tissue repair, *Clin Orthop Relat Res* (391 Suppl):S26-33.
- Pratta, M. A., Tortorella, M. D., and Arner, E. C., 2000, Age-related changes in aggrecan glycosylation affect cleavage by aggrecanase, *J Biol Chem* **275**(50):39096-102.
- Pritzker, K. P., Gay, S., Jimenez, S. A., Ostergaard, K., Pelletier, J. P., Revell, P. A., Salter, D., and van den Berg, W. B., 2006, Osteoarthritis cartilage histopathology: grading and staging, *Osteoarthritis Cartilage* **14**(1):13-29.
- Ptashne, M., and Gann, A., 1997, Transcriptional activation by recruitment, *Nature* **386**(6625):569-77.

- Radons, J., Bosserhoff, A. K., Grassel, S., Falk, W., and Schubert, T. E., 2006, p38MAPK mediates IL-1-induced down-regulation of aggrecan gene expression in human chondrocytes, *Int J Mol Med* **17**(4):661-8.
- Ralphs, J. R., and Benjamin, M., 1994, The joint capsule: structure, composition, ageing and disease, *J Anat* **184** (Pt 3):503-9.
- Rasmussen, E. B., and Lis, J. T., 1993, In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes, *Proc Natl Acad Sci U S A* **90**(17):7923-7.
- Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M., and Felsenfeld, G., 2002, Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities, *Proc Natl Acad Sci U S A* **99**(10):6883-8.
- Rittenhouse, E., Dunn, L. C., Cookingham, J., Calo, C., Spiegelman, M., Dooher, G. B., and Bennett, D., 1978, Cartilage matrix deficiency (cmd): a new autosomal recessive lethal mutation in the mouse, *J Embryol Exp Morphol* **43**:71-84.
- Rivas, R., and Shapiro, F., 2002, Structural stages in the development of the long bones and epiphyses: a study in the New Zealand white rabbit, *J Bone Joint Surg Am* **84-A**(1):85-100.
- Roeder, R. G., 1996, The role of general initiation factors in transcription by RNA polymerase II, *Trends Biochem Sci* **21**(9):327-35.
- Roeder, R. G., 1998, Role of general and gene-specific cofactors in the regulation of eukaryotic transcription, *Cold Spring Harb Symp Quant Biol* **63**:201-18.
- Roeder, R. G., 2003, Lasker Basic Medical Research Award. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen, *Nat Med* **9**(10):1239-44.
- Roeder, R. G., and Rutter, W. J., 1969, Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms, *Nature* **224**(5216):234-7.
- Rosenfeld, M. G., Lunyak, V. V., and Glass, C. K., 2006, Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response, *Genes Dev* **20**(11):1405-28.
- Roughley, P., Martens, D., Rantakokko, J., Alini, M., Mwale, F., and Antoniou, J., 2006, The involvement of aggrecan polymorphism in degeneration of human intervertebral disc and articular cartilage, *Eur Cell Mater* **11**:1-7; discussion 7.
- Roughley, P. J., 2006, The structure and function of cartilage proteoglycans, *Eur Cell Mater* **12**:92-101.
- Roughley, P. J., and White, R. J., 1980, Age-related changes in the structure of the proteoglycan subunits from human articular cartilage, *J Biol Chem* **255**(1):217-24.

- Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T., 2005, Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb, *Development* **132**(4):797-803.
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D. A., 2007, Mammalian RNA polymerase II core promoters: insights from genome-wide studies, *Nat Rev Genet* **8**(6):424-36.
- Sandelin, A., and Wasserman, W. W., 2005, Prediction of nuclear hormone receptor response elements, *Mol Endocrinol* **19**(3):595-606.
- Sandell, L. J., 1994, In situ expression of collagen and proteoglycan genes in notochord and during skeletal development and growth, *Microsc Res Tech* **28**(6):470-82.
- Sandy, J. D., 2006, A contentious issue finds some clarity: on the independent and complementary roles of aggrecanase activity and MMP activity in human joint aggrecanolysis, *Osteoarthritis Cartilage* **14**(2):95-100.
- Santer, V., White, R. J., and Roughley, P. J., 1982, O-Linked oligosaccharides of human articular cartilage proteoglycan, *Biochim Biophys Acta* **716**(3):277-82.
- Savagner, P., Miyashita, T., and Yamada, Y., 1990, Two silencers regulate the tissue-specific expression of the collagen II gene, *J Biol Chem* **265**(12):6669-74.
- Sawadogo, M., and Roeder, R. G., 1985, Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region, *Cell* **43**(1):165-75.
- Sawadogo, M., Van Dyke, M. W., Gregor, P. D., and Roeder, R. G., 1988, Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei, *J Biol Chem* **263**(24):11985-93.
- Schaefer, L., and Iozzo, R. V., 2008, Biological functions of the small leucine-rich proteoglycans: from genetics to signal transduction, *J Biol Chem* **283**(31):21305-9.
- Schinagl, R. M., Gurskis, D., Chen, A. C., and Sah, R. L., 1997, Depth-dependent confined compression modulus of full-thickness bovine articular cartilage, *J Orthop Res* **15**(4):499-506.
- Schwartz, N. B., and Domowicz, M., 2004, Proteoglycans in brain development, *Glycoconj J* **21**(6):329-41.
- Sekiya, I., Tsuji, K., Koopman, P., Watanabe, H., Yamada, Y., Shinomiya, K., Nifuji, A., and Noda, M., 2000, SOX9 enhances aggrecan gene promoter/enhancer activity and is up-regulated by retinoic acid in a cartilage-derived cell line, TC6, *J Biol Chem* **275**(15):10738-44.

- Shepard, J. B., Krug, H. A., LaFoon, B. A., Hoffman, S., and Capehart, A. A., 2007, Versican expression during synovial joint morphogenesis, *Int J Biol Sci* **3**(6):380-4.
- Silbert, J. E., and Sugumaran, G., 2002, Biosynthesis of chondroitin/dermatan sulfate, *IUBMB Life* **54**(4):177-86.
- Sims, R. J., 3rd, Belotserkovskaya, R., and Reinberg, D., 2004, Elongation by RNA polymerase II: the short and long of it, *Genes Dev* **18**(20):2437-68.
- Smale, S. T., and Baltimore, D., 1989, The "initiator" as a transcription control element, *Cell* **57**(1):103-13.
- Smith, R. L., Geller, A. I., Escudero, K. W., and Wilcox, C. L., 1995, Long-term expression in sensory neurons in tissue culture from herpes simplex virus type 1 (HSV-1) promoters in an HSV-1-derived vector, *J Virol* **69**(8):4593-9.
- Somesh, B. P., Reid, J., Liu, W. F., Sogaard, T. M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J. Q., 2005, Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest, *Cell* **121**(6):913-23.
- Soriano, P., 1999, Generalized lacZ expression with the ROSA26 Cre reporter strain, *Nat Genet* **21**(1):70-1.
- Sosic, D., Brand-Saberi, B., Schmidt, C., Christ, B., and Olson, E. N., 1997, Regulation of paraxis expression and somite formation by ectoderm- and neural tube-derived signals, *Dev Biol* **185**(2):229-43.
- Stormo, G. D., 2000, DNA binding sites: representation and discovery, *Bioinformatics* **16**(1):16-23.
- Strahl, B. D., and Allis, C. D., 2000, The language of covalent histone modifications, *Nature* **403**(6765):41-5.
- Svejstrup, J. Q., Li, Y., Fellows, J., Gnatt, A., Bjorklund, S., and Kornberg, R. D., 1997, Evidence for a mediator cycle at the initiation of transcription, *Proc Natl Acad Sci U S A* **94**(12):6075-8.
- Sztrolovics, R., Alini, M., Roughley, P. J., and Mort, J. S., 1997, Aggrecan degradation in human intervertebral disc and articular cartilage, *Biochem J* **326** (Pt 1):235-41.
- Sztrolovics, R., Grover, J., Cs-Szabo, G., Shi, S. L., Zhang, Y., Mort, J. S., and Roughley, P. J., 2002, The characterization of versican and its message in human articular cartilage and intervertebral disc, *J Orthop Res* **20**(2):257-66.
- Talusan, P., Bedri, S., Yang, S., Kattapuram, T., Silva, N., Roughley, P. J., and Stone, J. R., 2005, Analysis of intimal proteoglycans in atherosclerosis-prone and atherosclerosis-resistant human arteries by mass spectrometry, *Mol Cell Proteomics* **4**(9):1350-7.

- Tanaka, S., Antoniv, T. T., Liu, K., Wang, L., Wells, D. J., Ramirez, F., and Bou-Gharios, G., 2004, Cooperativity between far upstream enhancer and proximal promoter elements of the human $\{\alpha\}2(I)$ collagen (COL1A2) gene instructs tissue specificity in transgenic mice, *J Biol Chem* **279**(53):56024-31.
- Tanimoto, K., Liu, Q., Bungert, J., and Engel, J. D., 1999, Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice, *Nature* **398**(6725):344-8.
- Tchetina, E. V., Kobayashi, M., Yasuda, T., Meijers, T., Pidoux, I., and Poole, A. R., 2007, Chondrocyte hypertrophy can be induced by a cryptic sequence of type II collagen and is accompanied by the induction of MMP-13 and collagenase activity: implications for development and arthritis, *Matrix Biol* **26**(4):247-58.
- Thomas, M. C., and Chiang, C. M., 2006, The general transcription machinery and general cofactors, *Crit Rev Biochem Mol Biol* **41**(3):105-78.
- Thoms, B. L., and Murphy, C. L., Inhibition of hypoxia-inducible factor-targeting prolyl hydroxylase domain-containing protein 2 (PHD2) enhances matrix synthesis by human chondrocytes, *J Biol Chem* **285**(27):20472-80.
- Tortorella, M. D., Pratta, M., Liu, R. Q., Austin, J., Ross, O. H., Abbaszade, I., Burn, T., and Arner, E., 2000, Sites of aggrecan cleavage by recombinant human aggrecanase-1 (ADAMTS-4), *J Biol Chem* **275**(24):18566-73.
- Valenzuela, L., and Kamakaka, R. T., 2006, Chromatin insulators, *Annu Rev Genet* **40**:107-38.
- Valhmu, W. B., Palmer, G. D., Dobson, J., Fischer, S. G., and Ratcliffe, A., 1998a, Regulatory activities of the 5'- and 3'-untranslated regions and promoter of the human aggrecan gene, *J Biol Chem* **273**(11):6196-202.
- Valhmu, W. B., Palmer, G. D., Rivers, P. A., Ebara, S., Cheng, J. F., Fischer, S., and Ratcliffe, A., 1995, Structure of the human aggrecan gene: exon-intron organization and association with the protein domains, *Biochem J* **309** (Pt 2):535-42.
- Valhmu, W. B., Stazzone, E. J., Bachrach, N. M., Saed-Nejad, F., Fischer, S. G., Mow, V. C., and Ratcliffe, A., 1998b, Load-controlled compression of articular cartilage induces a transient stimulation of aggrecan gene expression, *Arch Biochem Biophys* **353**(1):29-36.
- van der Kraan, P. M., Buma, P., van Kuppevelt, T., and van den Berg, W. B., 2002, Interaction of chondrocytes, extracellular matrix and growth factors: relevance for articular cartilage tissue engineering, *Osteoarthritis Cartilage* **10**(8):631-7.
- Vaquero, A., Loyola, A., and Reinberg, D., 2003, The constantly changing face of chromatin, *Sci Aging Knowledge Environ* **2003**(14):RE4.
- Venn, M. F., 1979, Chemical composition of human femoral and head cartilage: influence of topographical position and fibrillation, *Ann Rheum Dis* **38**(1):57-62.

- Visel, A., Akiyama, J. A., Shoukry, M., Afzal, V., Rubin, E. M., and Pennacchio, L. A., 2009, Functional autonomy of distant-acting human enhancers, *Genomics* **93**(6):509-13.
- Voros, G., Sandy, J. D., Collen, D., and Lijnen, H. R., 2006, Expression of aggrecan(ases) during murine preadipocyte differentiation and adipose tissue development, *Biochim Biophys Acta* **1760**(12):1837-44.
- Wagner, E. F., and Karsenty, G., 2001, Genetic control of skeletal development, *Curr Opin Genet Dev* **11**(5):527-32.
- Wasserman, W. W., and Sandelin, A., 2004, Applied bioinformatics for the identification of regulatory elements, *Nat Rev Genet* **5**(4):276-87.
- Watanabe, H., Cheung, S. C., Itano, N., Kimata, K., and Yamada, Y., 1997a, Identification of hyaluronan-binding domains of aggrecan, *J Biol Chem* **272**(44):28057-65.
- Watanabe, H., Gao, L., Sugiyama, S., Doege, K., Kimata, K., and Yamada, Y., 1995, Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure and promoter sequence, *Biochem J* **308** (Pt 2):433-40.
- Watanabe, H., Kimata, K., Line, S., Strong, D., Gao, L. Y., Kozak, C. A., and Yamada, Y., 1994, Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene, *Nat Genet* **7**(2):154-7.
- Watanabe, H., Nakata, K., Kimata, K., Nakanishi, I., and Yamada, Y., 1997b, Dwarfism and age-associated spinal degeneration of heterozygote cmd mice defective in aggrecan, *Proc Natl Acad Sci U S A* **94**(13):6943-7.
- Watanabe, H., and Yamada, Y., 2002, Chondrodysplasia of gene knockout mice for aggrecan and link protein, *Glycoconj J* **19**(4-5):269-73.
- Watanabe, H., Yamada, Y., and Kimata, K., 1998, Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function, *J Biochem* **124**(4):687-93.
- Weinmann, R., and Roeder, R. G., 1974, Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes, *Proc Natl Acad Sci U S A* **71**(5):1790-4.
- West, A. G., Gaszner, M., and Felsenfeld, G., 2002, Insulators: many functions, many mechanisms, *Genes Dev* **16**(3):271-88.
- Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W., Stott, D., and Alleman, R. K., 1991, Multiple spatially specific enhancers are required to reconstruct the pattern of Hox-2.6 gene expression, *Genes Dev* **5**(11):2048-59.
- Williams, A., Harker, N., Ktistaki, E., Veiga-Fernandes, H., Roderick, K., Tolaini, M., Norton, T., Williams, K., and Kioussis, D., 2008, Position effect variegation and imprinting of transgenes in lymphocytes, *Nucleic Acids Res* **36**(7):2320-9.

- Workman, J. L., and Roeder, R. G., 1987, Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II, *Cell* **51**(4):613-22.
- Wright, E., Hargrave, M. R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A., and Koopman, P., 1995, The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos, *Nat Genet* **9**(1):15-20.
- Xiao, L., Kim, M., and DeJong, J., 2006, Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice, *Gene Expr Patterns* **6**(4):409-19.
- Yan, Z., Cui, K., Murray, D. M., Ling, C., Xue, Y., Gerstein, A., Parsons, R., Zhao, K., and Wang, W., 2005, PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes, *Genes Dev* **19**(14):1662-7.
- Yang, X., and Karsenty, G., 2002, Transcription factors in bone: developmental and pathological aspects, *Trends Mol Med* **8**(7):340-5.
- Yasumoto, T., Bird, J. L., Sugimoto, K., Mason, R. M., and Bayliss, M. T., 2003, The G1 domain of aggrecan released from porcine articular cartilage forms stable complexes with hyaluronan/link protein, *Rheumatology (Oxford)* **42**(2):336-42.
- Young, M. F., 2005, Mouse models of osteoarthritis provide new research tools, *Trends Pharmacol Sci* **26**(7):333-5.
- Zanin, M. K., Bundy, J., Ernst, H., Wessels, A., Conway, S. J., and Hoffman, S., 1999, Distinct spatial and temporal distributions of aggrecan and versican in the embryonic chick heart, *Anat Rec* **256**(4):366-80.
- Zawel, L., Kumar, K. P., and Reinberg, D., 1995, Recycling of the general transcription factors during RNA polymerase II transcription, *Genes Dev* **9**(12):1479-90.
- Zhang, Z., and Gilmour, D. S., 2006, Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript, *Mol Cell* **21**(1):65-74.
- Zheng, J., Luo, W., and Tanzer, M. L., 1998, Aggrecan synthesis and secretion. A paradigm for molecular and cellular coordination of multiglobular protein folding and intracellular trafficking, *J Biol Chem* **273**(21):12999-3006.
- Zhou, G., Garofalo, S., Mukhopadhyay, K., Lefebvre, V., Smith, C. N., Eberspaecher, H., and de Crombrughe, B., 1995, A 182 bp fragment of the mouse pro alpha 1(II) collagen gene is sufficient to direct chondrocyte expression in transgenic mice, *J Cell Sci* **108** (Pt 12):3677-84.